The Good Polity

Normative Analysis of the State

Edited by
Alan Hamlin and
Philip Pettit

7
Political Obligation

Russell Hardin

Theories of political obligation address what is perhaps the most bothersome issue in political philosophy: the justification of certain classes of coercion of individuals by government. In recent times, the notion of political obligation has been subjected to severe and telling attack. The principal focus of the attack has been on the element of voluntarism or consent on which obligation is often supposed to be based. If there is no consent, it is supposed that there can be no obligation. This is often little more than a definitional move. By definition, one can have an obligation only if one voluntarily assumes or consents to it.1 As has long been argued, it is implausible to assert of many people that they have voluntarily assumed the obligation of political obedience. It follows that they cannot be politically obligated.

The argument against the factuality of consent was well put more than two centuries ago by Hume (1748). Yet, the notion of political obligation has continued to motivate many writers who seemingly accept the brunt of Hume's argument. On the definitional criticism, one might suppose these writers have simply missed the point in practice even while accepting it in theory. Alternatively, one may suppose that the definitional criticism misses a fundamentally important point. The function of a notion of political obligation is to put the onus for governmental coercion on those who may be coerced. Hume's dismissal of consent theory merely means that this function is not served if obligation depends on consent. One may well suppose that it can be based on something other than consent. Most political theories—perhaps all, except extreme anarchism, among those that are taken seriously by contemporary political philosophers—include justifications for coercion in certain circumstances.2 Unless these theories are flawed at base, as genuine consent
theories effectively are, they can present us with serious candidate theories of obligation.

The attack on the notion of political obligation may largely have been directed against Lockean contract thinking in political philosophy. It is an attack on a particular kind of justification for governmental coercion of individuals in certain contexts in which the individuals may be supposed to be obedient. If obedience cannot be grounded, coercion cannot be grounded. Because the general import of the criticism of Lockean contract thinking was well established long ago, the contemporary criticisms of political obligation would be of little more than marginal interest if their purpose were merely to refine the arguments of Hume and others. In the most compelling contemporary political theories, consent plays a role, if any, only at a rationalist level of derivation or justification, not at the level of practice or of actual citizens. Are the critics merely kicking a dead horse or tidying up our vocabulary? Evidently not. Some of the contemporary critics of political obligation seem to have a further purpose, which is to show that government and law can be given no privileged position in the binding determination of right conduct by individuals. For the Lockean contract tradition, this readily flows from the absence of real consent. For theories not based on consent, however, their argument seems inconclusive. These theories may also fail to justify coercion, but they do not fail merely by definition.

Among theories under wide contemporary discussion, there are at least three general classes of political theory that are to some degree compelling but that are not subject to the standard criticism of political obligation. These are utilitarian, mutual-advantage and fairness theories. Within each of these, one can ground a principle of obligation in something other than consent. The role of obligation in fairness theories, especially that (or those) of Rawls, has been widely discussed, and I shall only briefly address it here. Rawls, who is one of the definitive critics of political obligation, asserts (1971, pp. 113–14) that the notion of obligation does not apply to citizens because citizens cannot plausibly be supposed generally to have consented. However, he also speaks of the need for coercion and continues to speak of ‘duties and obligations’ (pp. 240–1, 333–91), so that one might suppose a coherent Rawlsian theory of political obligation could be constructed.

The extent to which a political theory can motivate a principle of obligation or of justified coercion is a measure of its credibility. A theory that cannot motivate a principle of obligation to abide by its results cannot finally be compelling. I wish here to focus on utilitarian and mutual-advantage theories, which have in common that they derive institutional arrangements and, hence, political obligation from interest or welfare. I wish to investigate the coherence of a principle of political obligation that may be derived from each of these theories. It should become clear that, while a utilitarian theory straightforwardly implies a justification for coercion in its own terms, a mutual-advantage theory faces inherent difficulties in justifying coercion to abide by its tenets. I shall briefly outline each of these classes of theory before constructing the form of their defence of political obligation – which is to say, the form of their justification of coercion by government.

On a utilitarian theory an individual may be obligated or may have a duty to obey a government if that government produces more good on the whole than any alternative. The good that an alternative might produce must include the (likely) costs of its coming to replace the extant government. It is generally supposed that act utilitarianism reduces this conclusion to the following very weak one: I should obey this government at this moment if that act of obedience has better effects than non-obedience. This is a weak conclusion because it is no more than the usual act-utilitarian principle, so that the law has no privileged role in my obligation to perform one or another action. This characterization, however, misses the point of utilitarian political theory, which is inherently about problems that individuals cannot expect to handle without institutions, often powerful institutions. An institutional utilitarian can justify the creation of institutions that will be coercive if these will bring about better results than non-coercive institutions would. If this contingent condition is met, we can sensibly speak of a utilitarian theory of political obligation that is based in considerations of the general welfare.

Now consider mutual-advantage theories of politics. These do not generally require consent. Rather, they stipulate that one arrangement, A, is better than another, B, if in A everyone is better off than in B. Hence, we move from B to A because it is to our mutual advantage to do so. Of course, such a theory would be of no interest if there were never any possibility of universal improvement. Commitment to mutual-advantage theories requires belief that some governmental arrangement could be mutually advantageous as compared to some alternative, especially a status quo of weak or no government. Again, if this contingent condition is met, we can sensibly speak of a mutual-advantage theory of political obligation that is based in considerations of mutual benefit.

One might characterize mutual-advantage theories as a blend of elements of utilitarian and contract theories. They are, like utilitarianism, concerned with enhancing welfare. And, assuming people would voluntarily choose to be better off, they are, like contractarianism, concerned with universal voluntarism, of a kind, however, that can only be theoretically assumed rather than actually, tacitly or implicitly revealed. The principal difference between mutual-advantage and utilitarian theories of politics is that utilitarian theories readily allow trade-offs between individual welfare, whereas mutual-advantage theories do not. Rawls's theory is partly a mutual-advantage theory, although its more important element is concern for fairness in distribution (see Barry forthcoming). David Gauthier's theory of 'morals by agreement' (Gauthier 1986) is a full-fledged mutual-advantage theory. Many contemporary contractarian accounts have a strong concern for mutual
advantage and, indeed, given the implausibility of any consensual basis, they may generally be more accurately characterized as mutual-advantage than as contractarian theories.

I The role of institutions in welfarist theories

How can a utilitarian or mutual-advantage theory ground obligation? Essentially through recognition of the necessity for institutions designed to achieve the ends the theory prescribes. At the political level, utilitarian and mutual-advantage theories do not focus on individual actions or choices. Rather they focus on general means for achieving relevantly determined outcomes for all concerned. Because individuals suffer from such liabilities as limited cognitive capacity, limited information, poor theories of cause and effect, and perverse incentives from interaction with others, they can generally expect to achieve better outcomes according to either the utilitarian or the mutual-advantage criterion, if there are institutions to help get them to those outcomes.

Rawls argues, in ‘Two Concepts of Rules’ (1955), that utilitarianism is not subject to the traditional, even trite, complaint that it must violate justice in the following sense: if it were evidently true that the execution of an innocent after a contrived trial would prevent more deaths than it would entail, then a utilitarian would evidently have to recommend the unjust execution. Rawls objects to this argument that, in essence, it is set blandly somewhere out in the ether; it is not grounded in an institution that could actually carry out its recommendation. An actual institution of criminal justice is apt to work best in a utilitarian sense if it is designed to punish only those who are shown to be guilty. If so, it is such an institution that utilitarianism recommends, not the ethereal abuse of innocents to motivate others.

Suppose that we design utilitarian institutions. It is plausible that their workings would require actions of particular kinds by those who fill offices in them. These officers might decide, however, to act from direct, act-utilitarian calculations in violation of their offices. If such actions disabled the institutions, they would be wrong on a utilitarian account. Hence, the officers of our utilitarian institutions should follow the rules of the institutions, not the apparent dictates of act utilitarianism. This is a utilitarian conclusion. (See, further, Hardin 1988, esp. section 21.)

As already noted, the reason we need institutions, such as those for justice, for maintaining public works, and for collecting taxes to pay for the costs of government and its actions, is that institutions may be organized to do the job better than it could be done by the spontaneous efforts of individuals, even well-meaning utilitarian individuals. This is true in many ways. Consider two. First, institutions can help to co-ordinate actions for mutual and general benefit. Those who doubt their safety in the airspace over Chicago these days might ponder what it would be like without a coercive agency to regulate movements in that airspace. Thomas Nagel characterizes such coercions as unproblematic in that we can all be presumed to consent to the achievement of the relevant benefit and, hence, to the coercion that might be necessary to compel us in various moments, just as we must all consent to having traffic controls. As Nagel (1987, p. 224) says, such coercion ‘is not really forcing people to do what they don’t want to do, but rather enabling them to do what they want to do by forcing them to do it’. There may be complications of fact to muddy our account of any such coercion, but in principle this is not likely to be a contested view.

Second, one of the great values of government is to reap benefits from a division of labour in organizing ourselves beneficially. This consideration may cut much deeper into our concern with coercion. For example, a government specialist may decide on behalf of almost all of us that a certain drug or food or technology is unsafe for us and may ban its commercial development or sale. Most of us may consider this a benefit because we would be incompetent to judge for ourselves without massive wasted time on unwanted research, so that the government specialist either saves us that effort or protects us from the relevant harm. The government agent is not unlike our auto mechanics or dentists in specializing in ways that we do not. But the agent differs from these in acting, for better or for worse, without the incentives of market competition, so that no individual in the society may be able to seek alternative authoritative direction in using the putatively harmful drug or whatever. Because of this consideration, there may be real disagreement over government regulatory coercion, although there might not be.

Now note what we have done when we have put an institution in place to regulate our conduct for our own collective benefit, somehow defined. We have grounded any notion of individual political obligation in that collective benefit. The concept of political obligation generally is invoked to address conflicts between individual interests and the larger interests or will of a polity. In those political theories in which the larger interests of the polity are supposed to be grounded in individual interests, the notion of political obligation is inherently subject to a fallacy in reasoning about collective interests. If it is, then for these theories we may find that the notion of political obligation is incoherent for contingent reasons just as consent theories are incoherent for the contingent reason that too few of us can seriously be thought to have consented to any government. I wish to address this fallacy and to determine whether it must undermine any principle of political obligation for utilitarian and mutual-advantage theories.

The fallacy that threatens these interest- or welfare-based theories may be illustrated for certain theories based in mutual advantage. These theories are often couched in contractarian language as though individuals have somehow
contracted to the rule of their government and as though their consent creates an obligation. The presumption that people have consented is often itself apparently read off directly from the nature of the benefits that would flow to them from their political order. It is in this move that a fallacy potentially vitiates any claim of obligation. The fallacy is inherent in the assumption that you and I must consent to the rule of government because we must consent to what would be in our mutual interest. Unfortunately, what would be in our mutual or collective interest in our political organization is likely to be too ill defined to justify any particular government.

The form of the contractarian argument is that we consent to what is in our interest; the fallacy is that 'what is in our interest' is not uniquely defined, as I shall spell out below.

One way around problems inherent in supposing that we consent to government is to speak of a duty rather than an obligation to obey the law and to support a government. Hence, we might argue that there is no obligation for citizens per se, that there is mostly only a natural duty to abide by the laws of the state if the state or its laws satisfy some criterion. For example, Rawls (1971, pp. 113-14) argues that there is a natural duty of citizens to abide by the laws of a state that is just. A utilitarian can argue in similar fashion that, although they may have no political obligations based in consent, individuals have a duty to abide by laws that conduct to the general benefit. Unfortunately our fallacy can affict this argument as well: what conduces to the general benefit may not be well defined. This conclusion generally applies to utilitarian theories in which the possibility of precise addition of benefits and costs across individuals is ruled out. If our assessment of overall distributions is based on limited interpersonal comparisons, especially on only ordinal comparisons, of welfare, we cannot determine which of many optimal states is best and we are left with a fundamental conflict of evaluations that may undercut the logic of adopting one state rather than another.

For interest- or welfare-based theories, I wish to consider how certain claims for political obligation may make strategic sense, and then to consider complications that prevent them from being generally compelling. The discussion of political obligation often implicitly assumes that the structure of the problems we wish to regulate is that of certain 'easy' cases. What we wish to regulate, however, is, in principle, a much broader and more complex panoply of problems of varied structures. It is in its generalization from certain easy cases to more complex and yet commonplace problems that the notion of political obligation may lose its supposed mooring in individual interests.

It will be useful to break the discussion into parts by separately considering the kinds of political relationship we might wish to regulate with claims of obligation or duty. We may classify these political relationships in three general strategic categories of social interaction: those of essentially pure conflict, those of pure co-ordination, and those that involve a mixture of these two. In the mixed category, I shall especially be concerned with the centrally important class of interactions that we may call co-operation or exchange. I shall also be concerned with two very difficult classes of interaction that may be pervasive in political order: classes that we may call unequal co-ordination and biased co-operation. These classes of interaction will be defined later as variants of the modal classes defined immediately below.

Pure conflict interactions are represented by game 1, in which the pay-offs are strictly ordinal. The Row player can choose her strategy 1 or her strategy 1. She ranks the outcomes, considered from her own interests alone, in the order represented by the first pay-off in each of the two cells, with the pay-off of 1 as her best outcome and 2 as her second-best outcome. Column ranks the outcomes in the reverse order. Hence, the interaction is one of pure conflict, because what makes Column better off makes Row worse off. We can, of course, vastly expand the matrix of choices for Row and Column while keeping their interaction one of pure conflict. If Row ranks the outcomes of a large matrix from 1 to 30, Column must rank them in reverse order, from 30 to 1, if the interaction is pure conflict.

<table>
<thead>
<tr>
<th>Game 1 Pure conflict</th>
<th></th>
<th>Column</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Row</strong></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>2, 1</td>
</tr>
</tbody>
</table>

A two-person interaction is pure co-ordination if, as in game 2, both players rank the outcomes in identical order. A pure co-ordination game may pose no problems of choice if both players are well informed of each other’s interests and if there are not two or more outcomes tied for first choice in the two players’ rankings. If there is a tie for first choice, there may be a genuine problem of co-ordinating our choices of strategies to ensure that we end up with one of the preferred outcomes. If we can communicate before choosing, we should have no problem co-ordinating. Alternatively, even if we cannot communicate easily but our interaction is iterated many times, then we may generally expect to achieve successful co-ordination by trial and error. Once we succeed, we simply continue with the successful strategy choice as a convention (see Lewis 1969).

<table>
<thead>
<tr>
<th>Game 2 Pure co-ordination</th>
<th>Column</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td><strong>Row</strong></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>II</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Game 3  Prisoner’s dilemma or exchange

<table>
<thead>
<tr>
<th></th>
<th>Yield x</th>
<th>Column</th>
<th>Keep x</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row</td>
<td>Yield y</td>
<td>2, 2</td>
<td>4, 1</td>
</tr>
<tr>
<td></td>
<td>Keep y</td>
<td>1, 4</td>
<td>3, 3</td>
</tr>
</tbody>
</table>

While these two pure types of interaction are not without great interest, much of what is interesting in social life seems to involve interactions that combine elements of co-ordination with elements of conflict, as in what are infelicitously called mixed-motive games. The variety of such games is large even for two persons, and it increases geometrically as the number of persons involved increases. On the evidence of how much has been written about it and of how many experiments have been run on it, the prisoner’s dilemma, as represented in game 3, is by far the most interesting of all such mixed-motive games. It is partly a game of co-ordination because both players prefer the upper-left outcome to the lower-right outcome. It is partly a game of conflict because the Column player prefers the upper-right to the lower-left outcome and the Row player has the opposite preference. The chief reason why the prisoner’s dilemma has received such great attention is arguably that it seems to fit a vast array of actual interactions that interest us. It fits them because they are exchange interactions and the prisoner’s dilemma is the game-theoretical model of exchange (Hardin 1982).

All of these strategic structures can be generalized in some meaningful way to interactions among large numbers of people, as in a legal system or a social order. Any one of these strategic categories may seem to play a larger role in the grounding of some theories of politics than in the grounding of others. For example, contractarian theories typically seem to be especially concerned with mixed interactions of exchange and co-operation to the neglect of pure conflict and co-ordination interactions. Indeed, in an extraordinary move, Gauthier (1986) goes so far as to say that, if morals are the result of (hypothetical) agreement among rational persons, they can only govern interactions with the structure of interests represented in the prisoner’s dilemma. Given the extensive role of government in essentially conflict and co-ordination interactions and in mixed interactions other than the prisoner’s dilemma, we must want a theory of political obligation to accommodate such interactions or to justify their exclusion. Without a sophisticated argument to say why resolution of the much broader class of interactions cannot be seen to yield mutual advantage, as resolution of prisoner’s dilemma interactions generally does, we want more in our political theory than Gauthier allows in his moral theory.

Suppose we wish to achieve regulation of a particular form of social interaction. How we should best regulate it, either on a consent theory or on a utilitarian theory, may depend on which category it fits into — conflict, co-ordination or mixed motive. Political obligation and legitimacy typically are at issue in any justification of the use of coercion to motivate compliance or to correct disobedience. How we regulate an interaction and how we achieve coercion may be strongly related in that the latter may imply the former. But they may be partly unrelated in that regulation may largely be achieved without coercion. Moreover, the possibility of coercion may even be parasitic on or determined by the success of non-coercive regulation of other interactions. Our subjective sense of the legitimacy of a particular coercive move may depend on which of these relationships holds for the relevant interaction and the coercive device used.

II The hardest case: pure conflict

Pure conflict interactions do not generalize to larger numbers of players for the simple reason that three or more players cannot all have exactly opposing preferences over a set of outcomes, as two players can. They might therefore seem to be irrelevant for social analysis. Against this conclusion, note that we may often simplify large-number interactions into effectively two-party interactions if we may sensibly suppose that the large numbers are actually assignable to two conflicting groups. This is essentially the move Rawls makes in his difference principle: he evaluates or ranks outcomes according to how they benefit the worst-off class in the relevant society, as though all others were in another, more or less opposing, class. The worst-off class should be made to benefit at the expense of the residual class of all others if that is possible without making some other group the new worst-off class and making them even worse off than the original worst-off class was.

Such pure conflict interactions are inherently problematic for any theory of political obligation or duty. They can be dealt with under utilitarian theories if sufficiently strong interpersonal comparisons of welfare can be assumed to justify a claim that transfers from one class to another yield net benefits. Any such assumption is rejected by many, including contemporary economists who object to interpersonal comparisons of welfare, although modest comparisons are commonly assumed by political theorists, even if often only implicitly.

Pure conflict interactions may be dealt with on traditional contractarian theories only if we can presume that all would agree to bear the relevant costs. Prima facie, this sounds like a dubious presumption. On my reading, this troubling issue is not addressed in an articulate way by anyone in the contractarian tradition before Rawls. Rawls deals with it by supposing that, ex ante, we choose a set of institutions that would maximize the worst-off state we
might find ourselves in. This move involves a normative imposition on our choices. One might also read some of Rawls’s argument as supposing that we trick up the problem of such a pure conflict in distributive results as a problem of the mutual advantage to be got from our prior joint organization of society (see Hardin 1988, ch. 4). If we succeed in the latter move, there is then no problem of pure conflict but only one of mixed-motive or exchange. That is to say, we do not simply face the task of distributing what we have produced; rather, we co-operate for mutual advantage to create a particular political order. A full definition of that order includes what we produce as well as how we distribute it. Ideally, the problem we now face is an easy case of simple cooperation, as discussed below. Hence, we might suppose it is possible to obtain genuine agreement ex ante to the resolution of certain classes of (ex post) pure conflict interactions so that we could then assert a political obligation to abide by the relevant resolution from a claim of consent or mutual advantage.

III The easy cases, 1: pure co-ordination

Suppose you and I are in an interaction of pure co-ordination, as represented in game 2. We are both indifferent between the two outcomes with pay-offs (1, 1) but we both clearly prefer either of those outcomes to the other outcomes. If it is clear to me that you have already chosen your strategy I or that you probably will, then I prefer to choose my strategy I. If you have chosen or are likely to choose your strategy II, then I also prefer to choose my strategy II. Now suppose there are many of us involved in an interaction that is a straightforward generalization for many persons of game 2. Also suppose we choose strategies more or less independently in the sense that we do not directly communicate and we do not receive direct signals from some coordinating authority. If it appears to me that virtually everyone is choosing strategy I, then I shall wish to choose strategy I.

If we generalise this interaction to a large number of actors in a relevant way, I shall seem to have even less choice in the determination of which outcome we co-ordinate on than I should in a two-person case. If most others have chosen or seem likely to choose strategy I, then I should do so as well. If we iterate the interaction several times, we can generally expect all of us to reach the same solution even without extensive communication or co-ordination by an active agent. It would be pointless for me to try to get everyone to co-ordinate on strategy II if they are virtually all co-ordinated on strategy I. It would be pointless, first, merely because I should gain nothing from the change. But it would also be pointless because it would be difficult for me to give the others any incentive to switch strategy choices. In a two-person co-ordination, the fact that I choose strategy II would be incentive for you to choose it as well.

In one important sense it would be odd to speak of political obligation to go along with a large-scale co-ordination. The sense is this: that it would be odd in general to speak of my obligation to do what is in my interest. The obligation seems otiose. What moral force there is in the obligation comes from the fact that it is in my interest. Why would it be immoral and hence contrary to duty to violate the co-ordination? Because it would violate my interest.

There may still be some sense, however, to the claim that I am obligated to go along with a co-ordination: namely, that my not going along might harm others. For example, if I choose to drive on the right in Australia or on the left in North America, I run the risk of harming someone else in addition to harming myself. One may suppose that I have a natural duty not to harm others where there is no off-setting beneficial reason, and that this duty comes into play once certain co-ordinations are achieved or once certain conventions are established. Hence, I may have a duty to go along with an established co-ordination; and for relevantly political co-ordinations, such as those to achieve beneficial order, I may have a specifically political duty to go along. In general, however, it would be odd to have to invoke such a duty, because failure to observe it would be directly contrary to anyone’s own interest.

Here the duty is not conceptually otiose because the duty or obligation derives from the interest of others and is not merely the circular moralization of one’s own interest. But still the duty need not affect anyone’s motivation. Suppose that the point of speaking of political obligation is largely to justify the use of coercion to motivate compliance with some aspect of political order. The point of coercion typically is to give people different incentives, to make some action be in their interest rather than contrary to their interest. Using coercion to get someone to conform to a co-ordination that is already in her interest merely makes it more in her interest. If coercion affects behaviour in such a case, we must wonder whether it really is such a case.

IV The easy cases, 2: simple co-operation

Simple co-operation has the structure of the prisoner’s dilemma, game 3, which is the structure of ordinary exchange. You have what I want; I have what you want; we would both prefer to swap than to keep what we have. We may generalize this in an appropriate way from two to $n$ persons, in which case it becomes a problem of collective action. Once our co-operation involves many more than two persons, however, it may begin to be true that there are no relevant incentives to co-operate. In the two-person case, I have incentive to co-operate if that is the only way for me to get what you have. In the $n$-person case, I may still benefit from the co-operative actions of everyone else even if I do not act co-operatively. If the cost to me of my own co-operation is less than the additional benefit to me from my own co-operation, I have an incentive not to co-operate or an incentive to free-ride.
In a multi-person co-ordination, the incentive for me to choose one strategy rather than another typically increases as the number of others in the co-ordination increases. In a multi-person co-operation, the incentive to me to choose the co-operative strategy typically decreases as the number of others in the co-operation increases. My own interest and the interests of others are likely to be in clear conflict in such a co-operative interaction. Hence, the potential force of a notion of obligation or duty to co-operate may be much greater than that of an obligation or duty to co-ordinate. Moreover, the role of coercion seems clear in this context. We might suppose that, if we put to a vote the question whether all co-operate or none co-operate, there would be a unanimous preference for having all co-operate. Hence, we might think it justified to coerce anyone benefiting from the co-operation of others to co-operate as well. This is a position that Hart (1955, p. 183) and Rawls (1971, p. 112) have supported and that, of course, others have disputed (see Hardin 1985, esp. pp. 412-16). I think the issue is one that depends in part on the problem raised below (under ‘Biased co-operation’): whether I should be coerced to co-operate to our mutual advantage may turn on whether I can reasonably claim to think there would be a better alternative. As a rule, it would be odd to suppose that there would be no alternative that any particular one of us would prefer. The rare exception to this possibility would presumably be an extraordinarily important instance of mutually beneficial action in circumstances that made for a virtually unique recommendation for individual action toward the collective benefit.

V Unequal co-ordination

Suppose we are involved in a co-ordination problem analogous to that in game 4. I prefer the outcome (1, 2); you prefer the outcome (2, 1). Both of us would strongly prefer either of these to the other outcomes. Suppose further that there is no way for either of us to compensate the other or to bind ourselves to compensate the other for acquiescing in our preferred outcome. Hence our successful co-ordination must put one of us at a disadvantage relative to the other in that one of us must receive only a second preference while the other receives a first preference. We are in an unequal co-ordination.

<table>
<thead>
<tr>
<th>Game 4 Unequal co-ordination</th>
<th>Column</th>
<th>II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>3, 3</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>2, 1</td>
</tr>
<tr>
<td>Row</td>
<td>I</td>
<td>1, 2</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>4, 4</td>
</tr>
</tbody>
</table>

Now let us generalize game 4 from two to \( n \) persons in the following way. Some of us, \( m \) persons, prefer one outcome over all others, and the remainder of us, \( n - m \) persons, prefer a different outcome over all others, and all of us prefer either of these two outcomes over all others. As in the \( n \)-person simple co-ordination game discussed above, we may now find ourselves individually unable to affect the outcome that is selected. We may simply find ourselves individually best off if we go along with the nearly unanimous co-ordination that has been achieved.

For some of us, this co-ordination will yield the best of all outcomes; for others it will yield only a second-best outcome. The difference may be substantial. Consider two examples that are commonplace in our lives. In order to achieve productive co-ordination of our efforts, the best arrangement for us may be to have a strong organizational hierarchy or to have a clear division of labour even without organizational hierarchy. Either the hierarchy or the division of labour may imply great inequality of benefits. In the hierarchical organization of our production it may not matter very much for our productivity who is on the top and who is on the bottom. But it may matter very much to any individual whether she is on the top or on the bottom. Similarly, division of labour may be very efficient for total production, but the division may put some in activities or occupations that are far less desirable than others. Even in the mini-society of a family, efficient production may imply unequal roles.

In the general class of interactions represented by game 4, the acquiescence of those who are at a disadvantage in the unequal co-ordination is merely a matter of necessity or interest. Suppose it ceased to be in the interest of some of them to acquiesce – for example, because they could co-ordinate among themselves on forcing an alternative co-ordination on the larger society. There might then be no obvious argument that it would be wrong for them to cease to acquiesce. Any claim that we are obligated to go along with a generally beneficial co-ordination is ill defined for this interaction, because more than one ‘generally beneficial’ co-ordination is possible. Hence, a claim of political obligation to maintain a generally beneficial co-ordination cannot stop someone with a revolutionary intent to re-ordinate. There might be claims for political obligation that would have force here – for example, one might urge against the revolutionary that revolution cannot succeed or that it will be woefully harmful to virtually everyone – but such claims would not derive directly from the co-ordination structure of the interaction.

The more a sense of political obligation is needed to make things work in such an unequal context, the less it is likely to be justifiable. Why? Because the more it is needed to override individual interest the less it can be justified on the ground of its serving individual interests. Hence, the notion of political obligation becomes not merely otiose but spurious in the context of unequal co-ordination.
VI Biased co-operation

As usual, let us address the problem of biased co-operation by first considering a two-person case in order to make the issues clearer. Suppose you and I each have two goods. You have \( x \) and \( y \); I have \( u \) and \( v \). Your two goods are near-substitutes for each other, mine are near-substitutes for each other, but either of yours would be a good complement for either of mine in the following senses. Two goods are substitutes if they can each perform roughly the same function for us, as, say, pasta and potatoes might substitute for each other as starches in a meal. Two goods are complements if each of them adds to the value or enjoyment of the other, as, say, pasta and sauce complement each other.

Suppose further that we have slightly different preferences over various possible outcomes involving trades between us. In general, however, we would both rather make a trade of one of our goods for one of the goods of the other. I would most prefer one trade, for example, \( v \) for \( x \); you would most prefer a different trade, for example \( y \) for \( u \). But neither of us would prefer to have both of the other’s goods to both of our own. My preferences over all possible pairs are in the order \( u, v, x, y \) (my original holdings) and \( y, v, u, x \) (your original holdings). Your preferences are identical except for reversing the final two pairs in the list. Our problem, then, is that we could engage in several exchanges of one of my goods for one of yours, but we would not wish to trade both for both. That is, we face a choice among several (four) mutually exclusive bilateral exchanges, any one of which would be desirable to both of us if there were no alternatives. But, if you get your most desired trade, I do not get my most desired trade.

Whatever trade we now make may be unequal in the sense that, out of all possible outcomes, you will rank the result of our trade higher in the list than I will or vice versa. For example, suppose our trade leaves you with \( y \) and \( v \), and me with \( x \) and \( u \). I now have my most preferred pairing, while you have your fourth preferred pairing. If you get your most preferred pairing, \( u \) and \( x \), I may get only my fourth preference, and so forth. If the differences in our absolute valuations over our rankings are large, we may find our resulting trade quite conflictual, even though it benefits both of us as compared to the status quo. An argument merely from mutual advantage cannot settle which of various trades we should make.

Such problems may not generally be grievous in actual trades between individuals, because, outside families, we do not generally engage in such commodity-for-commodity trades. Rather, we trade goods (or time and effort) for money and money for goods without any need to match what I trade away and what I receive in trade with the counterparts of any particular other person. In political decisions over collective provisions of group goods, however, we may necessarily have to choose between a good that one group prefers and an alternative, largely substitute good that another group prefers. For example, we may give subsidies to those who seek private education or we may directly subsidize public education. To the extent that we trade off one of these kinds of programme for the other, we differentially benefit one group over another. Yet it may genuinely be believed by all that more education of anyone tends to benefit the whole society. That is to say, if one of these policies were impossible, we might all opt for the alternative. But, if both are possible, if they offer differential benefits, and if the two are largely mutually exclusive or are largely substitutes for each other, then we face a conflictual choice. If we make one collective provision, the other may cease to be a beneficial prospect because its cost will now outweigh the incremental additional benefit it might bring beyond what has already been brought by the first collective provision.

The problem here is accurately described as one of biased co-operation rather than merely as an instance of the unequal co-ordination discussed above. Note one important difference. In a case of unequal co-ordination, the result of the co-ordination is essentially self-enforcing against anyone who wishes to change the result to a more favorable outcome. In a problem of biased co-operation, we may all still prefer to be free riders on the collective provision, which may be achieved only through sanctioning all of us to generate relevant contributions, such as tax payments. There might be little or no hope of successful, voluntary collective provision of either of two (or more) outcomes that, although each is unanimously preferred to some status quo of no provision at all, are conflictingly ranked by various groups.

It is here, of course, that the problem of political legitimacy, obligation or duty enters. Am I obligated to contribute to our collective provision if I strongly prefer a different collective provision? One cannot simply answer this query with a claim that I am better off with the provision we have than with none. One must be able to say why I should be bound by a provision that is for me, and perhaps a very large group of others, inferior to what I could achieve with an alternative provision that similar-ly would make everyone better off than they would be with no provision. This question need not be bothersome if we seldom face such conflictual collective actions or provisions. Unfortunately, in a complex society with varied tastes and needs, we can arguably expect such conflicts over virtually all significant collective provisions. Hence, the static comparison between a particular collective provision and the status quo cannot settle the question of political legitimacy for that provision. The easy case of simple co-operation discussed above cannot generally fit the problem of political obligation or duty. Hence, mutual-advantage theories cannot motivate a principle of political obligation because the notion of ‘mutual advantage’ is too ambiguously defined in general.
VII Conclusion

For the last two classes of problems, unequal co-ordination and biased co-operation, it is inherent in political order that we often must choose one or the other mutually exclusive provision or regime in such hard cases. While the one or the other choice or even both choices might readily pass the test of utilitarianism, it is unclear how either can be said to pass the test of mutual advantage, which is inherently ill defined in the face of such choices. Mutual-advantage theories may be a contemporary effort to resuscitate some core part of contractarianism in moral and political theory. (This seems clearly to be true for the theories of Rawls and Gauthier.) Hence, they are framed ex ante, as though the problem is to compare possible regimes to some status quo of very limited or no government. In this move they are grievously afflicted by our fallacy. Alternatively, they could be useful in an effort to argue that it is in our mutual advantage to abide by an extant regime. In such an argument, the additional costs of making a serious change may typically mean that it would not be to our mutual advantage to move to any very different alternative regime. This, however, would be a very dispiriting use of mutual-advantage theory. Of course, in so far as mutual-advantage arguments are used in ideal theory, as they are by Rawls and Gauthier, this alternative use would be of little or no interest.

There are various ways in which we might deal with the problems of biased co-operation and unequal co-ordination in political life. Rather than try to judge the merits of a particular, perhaps fundamentally important, result of such interactions, we might judge the fit of that result with prior notions of how to resolve them. For example, Brennan and Buchanan (1985) wish to avoid conflicts of choices over actual goods and policies by reaching agreement on the structure of the institutions for making such choices. Alternatively, we might suppose that some particular institutional arrangement, such as majoritarian democracy, for making such choices is special or morally privileged (see, for example, B. Barry 1979). A major objection to first selecting an institutional arrangement and then letting it make the conflictual choices is that we may already face the conflictual choices at the level of choosing among institutional arrangements. As Brian Barry (1979) notes, in other words, of majoritarian democracy, if the majority population is green and the minority is orange and if their interests differ systematically, we may wonder how we could consensually agree on majoritarianism even ex ante from behind a more or less opaque veil of ignorance. Buchanan often notes that, in constitutional and social choice more generally, we start from here and now. Most of us will agree with this as a pragmatic judgement, although many may disagree with it as a judgement of ideal moral or political theory. In either case, we will have to recognize that, practically, Buchanan’s dictum implies that the choice of institutional arrangements does not cut the knot of conflicts over specific interests.

We might suppose that we could still find a society in which interests on political issues were so congruent that its members would not face such conflicts as those in interactions of unequal co-ordination or of biased co-operation. In such a society it would be hard to give much force to a notion of political obligation just because there would be no conflict of interest over political decisions. In less simple societies, however, ‘collective interest’ may be too conflicted a notion for political obligation to be grounded in it.

Notes

This essay was written while I was a Visiting Fellow in the Research School of the Social Sciences of the Australian National University. The idyllic conditions of intellectual life at ANU, with the glorious, not to say improbable, background of Australian flora and fauna, were a splendid stimulus to writing. I wish to thank Philip Pettit, Frank Jackson and Jane Marceau for their sponsorship of my time at ANU. I also wish to thank Philip Pettit, Reynolds B. Schultz and participants in the faculty seminar of the Center for Ethics, Rationality and Society at the University of Chicago for extended comments on an earlier draft of the paper.

1 Hart (1955), pp. 175–91; Rawls (1971), p. 113; Simmons (1979), pp. 14–16. Simmons gives useful citations to much of the extant literature on political obligation.

2 Simmons seems to be a fairly extreme anarchist. See Simmons (1987), esp. n. 9.


4 Simmons (1979, ch. 5) argues that a fairness theory cannot yield a principle of political obligation, but his argument turns on defining obligation as based in consent. See also Klosko (1987a, 1987b).

5 Hume (1739–40, III. ii. 9; 1978 edn, p. 553) argues that the obligation to obedience to government ceases when ‘interest ceases, in any great degree, and in a considerable number of instances’, to be served by government.