

THE POSSIBILITY OF RATIONAL POLITICS¹

JON ELSTER
University of Chicago

I. Introduction

We know, more or less, what rationality means when applied to individual choices and decisions. It is less clear what it could mean when applied to policy choices. In this article I discuss three questions. First, can one assign a meaning to the idea of rational political choice? Secondly, assuming that this assignment is feasible, what is the scope for rational political decisions? Thirdly, if this scope turns out to be limited (as it will), could there be an alternative guide to political action?

Section (II) states, very briefly, the theory of rational choice at the individual level. I shall assume that political rationality is defined by an extension from individual choices to policy choices which are, in some sense, made *by and for "society"*. I exclude, that is, conceptions of politics corresponding to Napoleon's "*Tout pour le peuple, rien par le peuple*". Within the framework of social choice theory, this amounts to imposing the non-dictatorship condition.

Section (III) discusses whether the notion of political rationality is at all meaningful, by considering various objections which derive from the principle of methodological individualism. Individuals act, on the basis of their desires and beliefs. Collectivities, as such, do not act; nor do they have desires and beliefs. Hence problems arise over aggregating preferences, centralizing information and implementing choices. Section (IV) considers the problem of politi-

¹ I thank King K. Tsao for invaluable research assistance.

cal irrationality, by looking at political analogues to various forms of individual irrationality that may be briefly described as weakness of will, excess of will and adaptive preference formation. Politics, to be rational, must devise ways of avoiding or coping with these tendencies. Section (V) asks whether political rationality always yield unambiguous prescriptions and suggests a largely negative answer, by arguing that *uncertainty* is a major obstacle to rational policy choices. The argument is not simply that it is difficult to anticipate the outcomes of policy decisions. More fundamentally, it is that incremental planning, social engineering and similar tactics are inadequate responses to this difficulty. In Section (VI) I conclude by arguing that *justice* rather than *rationality* must guide the fundamental policy choices. Rationality has a role to play, but it is subsidiary rather than primary.

II. *Individual rational choice*

The following exposition of the theory of rational choice will be highly schematic, intended only to provide a background for the discussions of political rationality in later sections.² Rational choice, on my conception, can be defined in two steps. First, we require that the desires and beliefs which lead to the action (in the way specified in the second step) be internally consistent. For the present purposes, the most important consistency requirement is that the agent's preferences be transitive: if an option x is perceived to be at least as good as another option y , and y is judged to be at

² For a more elaborate presentation of the conception of rational choice underlying this exposition, I refer to some of my other writings on the topic: *Ulysses and the Sirens*, rev. ed., Cambridge University Press 1984; *Sour Grapes*, Cambridge University Press 1983; "The nature and scope of rational-choice explanation", in E. LePore and B.P. McLaughlin (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell 1985, pp. 60-72; Introduction to J. Elster (ed.), *Rational Choice*, Oxford: Blackwell 1986, pp. 1-33; and "When rationality fails", forthcoming in K. Cook and M. Levi (eds.), *Limits to Rationality*.

least as good as z , then x must be at least as good as z .³ Secondly, we stipulate the following relation between action, beliefs, desires and evidence. An action is rational when (i) it can be justified as the best way of carrying out the agent's desires, given his beliefs, (ii) those beliefs can be justified by the evidence available to him and (iii) the amount of evidence collected by the agent can be justified in terms of his desires and the constraints on available information.

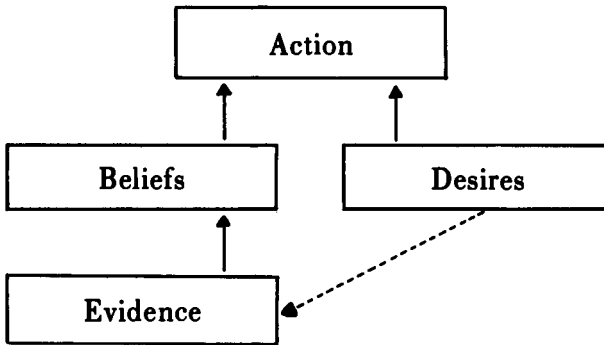


Fig. 1

In other words, showing that an action is rational amounts to offering a sequence of arguments in which the desire of the agent is taken as given but everything else has to be justified—ultimately in terms of that desire. *Explaining* an action as rational requires, in addition, a demonstration that the action was carried out *because* it was rational, and not by some accidental cause that just happened to produce a rational action.

These are the bare bones of rational-choice theory. I shall now proceed to flesh it out, and to point to some of the

³ We may note, for later reference, that this implies that if the agent is indifferent between x and y and between y and z , he must be indifferent between x and z , since indifference between two options can be defined by saying that each of them is at least as good as the other.

problems that it may encounter when applied to actual decisions. A first problem to be discussed, if not resolved, is raised by the assumption that desires are given and not themselves subject to rational justification, over and above the formal requirement of logical consistency. The intuitive, pre-theoretical constraints of the notion of rationality suggest that sometimes we would indeed want to say that desires can be irrational in a substantive sense, but it is not easy to say exactly what that sense is.

A first suggestion would be to say that rational desires are desires having which makes one happy. On reflection, this proposal is inadequate as a general characterization. It would, for example, support the official Soviet characterization of political dissenters as irrational, on the grounds that they are manifestly made unhappy by their unrealizable desire for political freedom.⁴ In other cases, the proposal is more attractive. One might well argue that having a very high rate of time discounting or a strong preference for risk-taking is irrational, because it tends to make one's life as a whole an unhappy one. These attitudes are irrational because they are self-destructive: an individual who at any given time acts according to these preferences undermines his opportunity of thus acting at later times.

Another suggestion could be to look at the causes of the desires rather than at their effects. One might equate rational desires with autonomous desires, *i.e.* desires with the right kind of causal history.⁵ The notion of autonomy is notoriously complex. In a recent discussion, Joel Feinberg distinguishes between four main senses of the notion and finds, within the variety which is most relevant here ("autonomy as condition"), twelve further

⁴ See S. Bloch and P. Reddaway, *Russia's Political Hospitals*, London: Futura Books 1978, p. 255.

⁵ If one could demonstrate that myopia, risk-seeking and other selfdestructive phenomena are necessarily non-autonomous attitudes, the two definitions of preference-rationality would coincide. I cannot see, however, how this demonstration might be carried out.

subvarieties. Of these, the idea of autonomy as authenticity seems the most promising one. "To the degree to which a person is autonomous he is not merely the mouthpiece of other persons or forces. Rather his tastes, opinions, ideals, goals, values, and preferences are all authentically *his*."⁶ Unfortunately, this view is more effective as a negative theory of heteronomous or inauthentic desires than as a positive theory of autonomy. We may all agree that people are heteronomous and perhaps even irrational when their desires are shaped by a tendency to conform and adjust to other people and to circumstances, but this falls far short of a positive characterization of autonomy.⁷ We can, nevertheless, retain the idea of *improvements* in rationality, when desires with an objectionable causal history are eliminated.

A person may be subject to myopia or excessive risk-seeking (or excessive risk-aversion), know it, deplore it, and take steps to contain the damage. While not fully rational, he deploys "imperfect rationality."⁸ He can do so by precommitting himself to the behaviour which corresponds to his considered judgment,⁹ or by constructing unbreakable rules for himself.¹⁰ This involves a two-step approach to choice, roughly similar to the distinction between constitutional law and other legislation. (This analogy is further explored in (IV) below.) It is not as if the person is choosing his preferences. Rather, he is protecting them against interference from his other propensities, which for this purpose are treated as blind causal forces external to his will.

Rationality as described concerns instrumental efficiency. Some ends, however, cannot be attained by

⁶ J. Feinberg, *Harm to Self*, Oxford University Press 1986, p. 32.

⁷ On this point see also my *Sour Grapes*, Ch. 1.3.

⁸ See my *Ulysses and the Sirens*, Ch. II.

⁹ T.C. Schelling, *Choice and Consequence*, Cambridge, Mass.: Harvard University Press 1984, Chs. 3 and 4.

¹⁰ G. Ainslie, "Beyond microeconomics", in J. Elster (ed.), *The Multiple Self*, Cambridge: University Press 1986, pp. 133-75.

instrumental means. They may *come about* as the by-product of action, but cannot be *brought about* as the deliberate outcome of action. Belief and forgetfulness are typical examples of such states that are essentially by-products.¹¹ Often, it may be instrumentally useful to hold a belief which, in fact, is not true. Married couples, for example, are probably better off if they underestimate the probability that they will divorce. Yet one cannot simply decide to adopt a belief which one also believes not to be true. Similarly, trying to forget something may just engrave it more deeply in one's memory; trying to be spontaneous or trying to fall asleep may be similarly self-defeating. There is, in fact, a particular kind of irrationality —call it hyperrationality— which consists in the misplaced application of precepts of rational choice. The phrase "excess of will" has been coined to characterize this tendency.¹²

I have sketched an account of what it means to be rational, and given some examples of ways in which people deviate from rationality. It remains to discuss cases in which the concept of rationality is undefined or, more weakly, not uniquely defined. I shall not here touch upon the mostly trivial cases in which there are several optimal options. Instead, I shall focus on the problems which arise when there is no optimal option. As will be clear from Fig. 1, the existence of a rational action demand (i) that there exists an action which is optimal in the light of the beliefs and desires of the agent, (ii) that there exists a belief which is optimal in the light of the evidence available to him and (iii) that there exists an amount of evidence which it is optimal for him to collect in the light of his desires. If one or more of these conditions fail, the concept of rationality is to some extent indeterminate.

¹¹ See my *Sour Grapes*, Ch. II.

¹² L. Farber, *Lying, Despair, Jealousy, Envy, Sex, Suicide, Drugs and the Good Life*, New York: Basic Books 1976.

Indeterminacy of action, given desires and determinate beliefs, arises when the agent is unable to rank the consequences of the options facing him or, more technically, when his preference ordering is incomplete. There may then, for example, be two options such that each of them is preferred to all remaining options, but neither is preferred to the other, nor is the agent indifferent between them. If the choice will affect the welfare of other people in a way that matters to the agent, his inability of comparing their utilities may block comparison of the options. Also, if the choice is expected to have important consequences in the distant future, his inability to envisage now how they will matter to him then may have the same result. There may be, in other words, problems of intrapersonal as well as of interpersonal comparisons of utility which block the comparison of options between which the agent has to choose. Note that I am not saying that having incomplete preferences is irrational, only that when preferences are incomplete, the prescriptive and predictive power of rationality is to that extent blunted.

Indeterminacy of belief, given a determinate amount of evidence, arises in two main ways. First, there is uncertainty about the outcomes because of ignorance of singular facts or of causal connections. Again, this is especially likely to occur with consequences that ramify into the remote future. Next, belief indeterminacy can arise out of strategic interaction between rational individuals. If each of them has to form an expectation about what the others will do before he can make his own decision, the interaction structure may be such that there is no mutually supporting systems of beliefs.

Indeterminacy of how much evidence to collect, given the desires, is a feature of very many choice situations. To see why, one must keep in mind that rationality is a subjective notion through and through. In an objective sense, there may well exist an optimal amount of time, money or effort which the agent should spend on col-

lecting information. Usually he will not, however, be able to determine what that amount is. To do so requires more information which he is unlikely to have, except in highly stereotyped and repetitive choice situations. Typically, collection of information has associated with it known direct costs, uncertain opportunity costs and uncertain benefits. To assess the uncertain elements is itself a costly and uncertain operation. Often one will simply have to act, without any illusion that the decision is in any sense optimal.

In a sense, all forms of indeterminacy can be stated as uncertainty: uncertainty about values, about states of nature or laws of nature, about the behaviour of others, or about the expected costs and benefits of information. The pervasive uncertainty in human affairs is the fundamental reason why rationality may be a weak guide to action. We would not expect this to be any less true of political decisions than of individual ones. Section (V) elaborates on this point.

III. *Preferences, information, and implementation*

On one conception of politics, it is like individual choice writ large. First, political *preferences* —goals and priorities— are defined, by the democratic political process. Next, government agencies gather *information* about factual matters and about means-ends relationships, to form an opinion about which policies will best realize those goals. Finally, other agencies *implement* these optimal policies. On this conception, political choice corresponds to the model represented by Fig. 1. Rational political choice can be represented in terms of the desires, beliefs and evidence of a supra-individual actor, "society". Few if any have believed that this view of political choice is literally true, but many have implicitly or explicitly assumed that we may proceed *as if* it were true, for explanatory or practical purposes. While paying lipservice to methodological individualism, they have

assumed that little harm is done by treating the polity as a unitary actor, with stable values, beliefs and a capacity to carry out its decisions. This assumption has been strongest in the study of international relations and in the theory of economic planning. For obvious reasons, it has been much less pronounced in the study of the domestic politics of pluralist democracies. Yet even here the temptation to use the convenient actor language can be overwhelming. It may be useful to explain exactly why it is so treacherous.

A very general reason why polities are unlike individuals can be summarized in one word: opportunism. It is much easier for an individual to deceive others than to deceive himself.¹³ When individuals engage in self-serving deception or opportunism, there is no guarantee that the aggregate outcome of their behaviour will correspond to the unitary-actor model of political rationality. Let me explain what this means in the three dimensions of choice which concern us here: preferences, information, and implementation.

As noted above, I assume that the method for aggregating individual preferences is constrained to be non-dictatorial. In addition we would want it to be invulnerable to opportunism: it should never offer an incentive to the individual to express preferences other than his true ones. He should never be able, by misrepresenting his preferences, to bring about an outcome which he prefers (according to his true preferences) over the one that would be brought about if he expressed his true preferences. If the input to the aggregation process are preferences which may be insincere in this sense, it would be absurd to say that the output represents "society's values". The outcome of the aggregation process might be a social preference for x over y , although all individuals prefer y to x . It turns out, unfortunately, to be very difficult to design aggregation mechanisms which are

¹³ See, however, Chs. 1-5 in J. Elster (ed.), *The Multiple Self*.

“strategy-proof in this sense”. The only voting procedure which is both strategy-proof and non-dictatorial is “random voting”, in which the probability of an option being chosen is equal to the proportion of individuals who rank it as their first choice.¹⁴ While this procedure may have some merits (besides that of being strategy-proof),¹⁵ it has other, obvious drawbacks that rule it out as a desirable scheme.¹⁶ Although strategy-proof mechanisms for revealing preferences can be devised for special cases,¹⁷ one cannot in general assume that people can be induced to be honest out of self-interest.

This problem of “incentive compatibility” extends to that of gathering information about factual matters. When economic agents are asked to provide information which is easily available to them, but would be available only at some cost (if at all) to others, one must assume that they will ask themselves whether it is in their interest to do so. Soviet-type economies, for example, are well known for the perverse incentives they create against truthfull reporting. Sometimes the fear of being punished as the bearer of bad tidings creates an incentive to present things as better than they really are. At other times self-interest leads one to present the situation as worse than it actually is, as when a manager underreports production to avoid an increase in his quota. These should not be dismissed as pathological phenomena, since essentially similar problems can be expected to arise in any system that depends on collecting information from decentralized sources. Again, while the problem may be overcome in special cases, there is no general recipe for making people tell the truth.

¹⁴ A. Gibbard, “Manipulation of voting schemes: A general result”, *Econometrica* 41 (1973), 587-601.

¹⁵ For a rare defense of random voting see Note, “Choosing representatives by lottery voting”, *Yale Law Journal* 93 (1984), 1283-1308.

¹⁶ For details, see my forthcoming Tanner Lectures (Oxford University) on “Taming chance: Randomization in individual and social decisions”.

¹⁷ See for example P.C. Ordeshook, *Game Theory, and Political*, Cambridge University Press 1986, Ch. 5.6.

Finally, incentive problems arise at the level of implementing the social decision. For the individual, there is usually no distance at all between making a decision and carrying it out, barring cases of weakness of will. The unity of the individual ensures that decisions, once made, are also carried out. The lack of unity of the polity makes this a much more problematic assumption. The "agents" who are charged with implementing the decisions made by the "principal" cannot be trusted to disregard their self-interest, nor can the principal always monitor their activities.¹⁸ There is by now a large public-choice literature on how budget-maximizing bureaucracies can distort the political will.¹⁹ The view that bureaucracies are a simple transmission machinery, analogous to the nerves and muscles that allow the hand to carry out the decisions made in the head, is flawed.

I do not want, however, to rest my case on the danger of opportunism. While there is always a risk of self-serving behaviour, the extent to which it is actually present varies widely. Much of the public choice literature, with its assumption of universally opportunistic behaviour, simply seems out of touch with real world of bureaucracies, in which there is often a great deal of public-spiritedness and sense of duty. If people always engaged in opportunistic behaviour when they could get away with it, civilization as we know it would not exist. Moreover, one of the tasks of politics is surely to shape social conditions and institutions so that people behave honestly, because they believe that the basic structure of their society is just.²⁰ We must ask, therefore, whether a just society, with an effective norm of honesty, would

¹⁸ A seminal article on the principal agent problem is M.C. Jensen and W.H. Meckling, "Theory of the firm: managerial behavior, agency costs and ownership structure", *Journal of Financial Economics* 3 (1976), 305-360.

¹⁹ For a survey see D.C. Mueller, *Public Choice*, Cambridge University Press 1979, Ch. 8.

²⁰ This is a main theme in John Rawls, *A Theory of Justice*, Cambridge, Mass.: Harvard University Press 1971.

be a good approximation to the unitary-actor model of rational politics.

The short answer is that while it would surely be a better approximation than a society in which opportunism was rampant, very serious difficulties would remain. Although the implementation problem would disappear, problems of aggregating preferences and centralizing information would remain. Even when preferences are honestly expressed. Arrow's impossibility theorem and later results along similar lines tell us that it is not in general —*i.e.* for any combination of individual preferences possible to generate a social preference ranking which is (i) complete, (ii) transitive and (iii) respects individual preferences. The third requirement can be further specified as follows: (iii.a) the social ranking should be non-dictatorial; (iii.b) it should prefer x to y when all individuals prefer x to y ; and (iii.c) when ranking x and y it should be sensitive only to how individuals rank x and y and not to how they rank other options.

The implications of Arrow's result for the present purposes are serious. True, the unitary-actor conception does not require a complete social ranking of all options. If we are willing to accept that rational individuals can have incomplete preferences, the extension of individual rationality to the political domain might also retain this feature. The problem, however, is to specify a partial ordering of some substantive interest. The rule of unanimity, as in the writings of Buchanan, Tullock and Brennan, yields an extremely incomplete ordering, even when applied, as they intend it to be, at the level of constitutional legislation.²¹ True, one can get around this by the formal trick of saying that in the absence of unanimous preference for x over y , society is indifferent between them. This (complete) social preference order lacks, however, the desired property of transitivity:

²¹ For the most recent statement of this approach, see C. Brennan and J.M. Buchanan, *The Reason of Rules*, Cambridge University Press 1986.

society may now be indifferent between x and y and between y and z , yet prefer x over z .²² And in any case, of course, it seems extremely undesirable to use an aggregation mechanism that would yield social "indifference" in the vast majority of pairwise choices. Nor do I know of any other incomplete ordering (or a complete extension of one) which would be more attractive.

One might question the requirement that the aggregation mechanism be defined for all constellations of individual preferences. One might argue, that is, that if individual preferences are sufficiently similar or fall into some particular pattern, paradoxes of intransitivity or cyclicity need not arise. Yet the conditions needed to avoid these paradoxes turn out to be quite stringent. It is very easy for paradoxes to occur.²³ One might argue that if politics is conceived as a process that aims not only at aggregating preferences, but also at transforming them through rational discussion, preferences are likely to be "laundered" of their selfish elements.²⁴ I believe there is something to be said for this view, but it does hardly amount to a proof that paradoxes will not arise. It is a shallow conception of politics which believes that self-interest is the only source of conflict. We know that concern for the environment, for those who are conceived but still unborn, or for those not yet conceived, can give rise to political struggle at least as intense as conflict over material issues. One might also argue, perhaps, that there is a feedback mechanism from social outcomes to individual preferences that shape the latter

²² Imagine a three-person society and three options, x , y and z . Writing P for strict individual preference, we may stipulate that their preferences are as follows: xP_1yP_1z , yP_2xP_2z , xP_3zP_3y . On the unanimity rule, society would be indifferent between x and y and between y and z , since in neither case is one option unanimously preferred, yet x is unanimously and hence socially preferred to z .

²³ For a summary of some recent results, see W. Riker, *Liberalism Against Populism*, San Francisco: Freeman 1982. Ch. 7.

²⁴ For this argument see R. Goodin, "Laundering preferences", in J. Elster (ed.), *Foundations of Social Choice Theory*, Cambridge University Press 1986, pp. 75-102, and J. Elster, "The market and the forum", *ibid.*, pp. 103-32.

so as to yield a non-paradoxical social preference order.²⁵ It seems unlikely, however, that there could be any such feedback mechanism, and in any case it would have to be demonstrated rather than assumed. This being said, the likelihood of political inconsistency can be expected to vary across societies, according to the degree of diversity of preferences.

The centralization of information also gives rise to problems which are quite independent of the issue of opportunism. To make rational choices, two sorts of factual knowledge are required. First, one must know the initial conditions: the distribution of individual desires, beliefs, productive capacities etc. Secondly, one must have knowledge about endsmeans relationships, *i.e.* about social causality. The problems related to the second kind of factual knowledge will concern us in (V) below. Here I only want to insist on the difficulties of gathering the first kind of information. Even if individuals tried to report their beliefs, desires and abilities as truthfully as possible, and even if we disregard the opportunity costs of writing these reports and the risk that the information might be out of date when finally used, the center would not find it very useful. The individual's knowledge about his mental states and productive capacities is largely tacit, embodied and personal, rather than explicit, verbal and abstract.²⁶ Firms do not have access to the whole "production function" on which they are operating. They have to know what they are doing, but they have no incentive to know what they could do, until forced to by circumstances.²⁷ Consumers may be quite unable to tell what purchases they plan to make over the next year or years. These familiar objections to

²⁵ This suggestion has been made by Adam Przeworski in unpublished work.

²⁶ K. Polanyi, *Personal Knowledge*, New York: Harper 1962.

²⁷ R. Nelson and S. Winter, *An Evolutionary Theory of Economic Change*, Cambridge, Mass.: Harvard University Press 1982. Ch. 4 and *passim*.

central planning²⁸ apply, in varying degrees, to all unitary-actor conceptions of politics.

These problems, while real, should not be exaggerated. The political process does, as a matter of fact, yield relatively consistent decisions. While to some extent arbitrary with respect to the individual preferences, they are at least non-dictatorial.²⁹ Moreover, with respect to major social choices I am no sure that decisions are all that arbitrary. If we look at a major political decision (or series of decisions) that has shaped history over the last half century, the rise of the welfare state, it seems implausible to argue that, keeping individual preferences constant, this development might not have happened at all, although quite plausible that the timing and specific forms of implementation could have been different. I shall have more to say about this in (VI) below. Also, many kinds of information can be and routinely are, collected by surveys and by requiring firms to provide regular data on sales, employment, etc. Using such data production functions and demand functions can be estimated at the macro level, even if unknown at the micro level. Although fine-tuned central planning is unfeasible, one may have sufficient information for coarsegrained macro-economic planning, which is successful at least if compared with the alternative of doing nothing.

I conclude that actual political systems may approach the unitary-actor model to various degrees, depending on the ambitiousness of the political goals, the honesty of the citizens, the distribution of their preferences, and

²⁸ See in particular the writings of F.A. Hayek, beginning with "Economics and knowledge", *Economica* n.s. 13 (1937), 33-54.

²⁹ Subtle form of dictatorships or oligarchy may, however, arise at the level of agenda manipulation. If a small elite can control the agenda or the voting procedure so as to obtain the result they want in each particular case, the fact that the process is formally democratic is irrelevant. Democracy (i.e. nondictatorship) could be ensured by using the same procedure in all cases or (preferably) by choosing procedures randomly in each case.

the technology of information-collecting. Soviet-type systems, which have very ambitious goals and seem to foster a great deal of opportunism, may, paradoxically, deviate more from the model than pluralist democracies. Perhaps the paradox could be put in the following, rather stark form; the more the self-image of a political system approaches the unitary-actor model, the less it will approach it in practice.

IV. *Precommitment, by-products, and preference formation*

Individuals often find it hard to be rational. They sometimes act rashly, against their interest and better judgement. A long relationship or a life of hard work may be undone by a single act of adultery or gambling. People may smoke, drink and overeat even when they know that they should not. To guard against such behaviour, sheer will-power may be insufficient. It is sometimes better to take one's lack of will-power as given and adapt rationally to that fact. *Precommitment* is the generic technique for coping rationally with one's own lack of rationality. It can take the form of making the undesirable activity physically impossible, as when Ulysses bound himself to the mast. If I know that I may do something stupid at the office Christmas party, I can stay away. Less radically, an individual may precommit himself to a certain course of action by ensuring that he will suffer a penalty if he chooses the undesirable, but tempting alternative. If I know that I have a tendency to cancel appointments with my dentist, I may ask him to bill me in full even if I cancel.

Political rationality may need similar devices for self-binding or precommitment. In an early discussion of this problem, Spinoza makes the point with exemplary clarity:

[It] is by no means contrary to practice for laws to be so firmly established that even the king himself cannot

repeal them. The Persians, for example, used to worship their kings as gods, yet even their kings had no power to repeal laws that had once been established, as is clear from Daniel. Chapter 6; and nowhere, so far as I know, is a king appointed unconditionally, without any explicit terms. This, in fact, is contrary neither to reason nor to the absolute obedience due to a king; for the fundamental laws of the state must be regarded as the king's permanent decrees, so that his ministers render him complete obedience in refusing to execute any command of his which contravenes them. We may clarify this point by reference to Ulysses, whose comrades did execute his command in refusing, in spite of all his orders and threats, to untie him from his ship's mast while he was enchanted by the Sirens' song; and it is put down to his good sense that he thanked them afterwards for carrying out his original intention so obediently. Even kings have followed the example of Ulysses; they usually instruct their judges to have no respect for persons in administering justice, not even for the king himself, if by some odd mischance he commands something which they know to contravene established laws. For kings are not gods, but men, who are often enchanted by the Sirens' song. Accordingly, if everything depended on the inconstant will of one man, nothing would be stable. Hence if a monarchy is to be stable it must be organized so that everything is done by the king's decree alone (*i.e.* every law is the king's declared will), but not everything the king wills is law.³⁰

Similar considerations apply to democracies.³¹ If all is-

³⁰ *Tractatus Politicus*, VII.I. I am grateful to E. Balibar for indicating this passage to me.

³¹ For a discussion of the changing function of the Rule of Law, from a protection against absolute monarchy to a protection against absolute democracy, see the contributions of F. Sejersted in J. Elster and R. Slagstad (eds.), *Constitutionalism and Democracy*, forthcoming from Cambridge University Press.

sues were subject to simple majority voting, society would lack stability and predictability. A small majority might easily be reversed, by accidents of participation or by a few individuals changing their minds. More importantly, the majority might be persuaded by the passions of the moment to act rashly and to override individual rights granted by earlier decisions. All democracies, whether direct or indirect, have had some stabilizing devices to prevent all issues from being up for grabs by simple majority voting at all times. In Athens, new legislation was subject to control by the *nomothetai*, a group of individuals chosen by the Assembly with the authority to approve or reject laws passed by the Assembly.³² Another institution with a similar purpose was the *graphe paranomon*, whereby an individual might be punished for having proposed an illegal law in the Assembly, even if it had already been passed.³³ In fourteenth-century Florence democracy was stabilized by a complicated system that involved long lags between the time at which an individual was nominated for office and the time when he entered office, as well as heavy reliance on lotteries to select among the nominated candidates.³⁴ In modern representative democracies self-binding can take several forms.³⁵ Democratic abdication of power can occur when the assembly irrevocably delegates certain decision power to independent bodies, like the Federal Reserve Board or the International Monetary Fund. Constitutions embody limitations on democratic power, to the extent that they contain (i) substantive rules protecting privacy, property, civil liberties, etc., and (ii) procedural rules that require more than a simple majority to change the constitution.

³² D.M. MacDowell, *The Law in Classical Athens*, London: Thames and Hudson 1978, p. 48 ff.

³³ *Ibid.*, p. 50 ff.

³⁴ See J. Najemy, *Corporatism and Consensus in Florentine Electoral Politics 1280-1400*, Chapel Hill: University of North Carolina Press 1982.

³⁵ See the essays in Elster and Slagstad (eds.), *Constitutionalism and Democracy*.

Precommitment is not without problems of its own, however. To see how they arise, consider the story of Daniel in the den of lions referred to by Spinoza. Here it is told how King Darius was rapped by the enemies of Daniel into issuing a decree "that whosoever shall ask a petition of any God or man for thirty days, save of thee, O king, he shall be cast into the den of lions". When Daniel then proceeded to make his prayers to God, his enemies denounced him to Darius and demanded that he be cast in the den. Darius tried to get out of this predicament, but they confronted him with the law "that no decree nor statute which the king establishes may be changed", upon which he yielded. As we know, the lions did not touch Daniel, but the story nevertheless illustrates the risks of precommitment. By rigidly binding oneself to certain rules or procedures, one may be prevented from making the right choice in unforeseen emergencies.

Thus, when discussing the proposal that economic policy should be taken out of the ongoing political process and entrusted to an independent institution like the Central Bank, William Nordhaus writes that "It may be objected, however, that delegating responsibility to an agency that is not politically responsive to legitimate needs is even more dangerous than a few cycles. This danger is frequently alleged regarding central banks which pay more attention to the 'soundness of the dollar' or the latest monetarist craze than to fundamental policy problems."³⁶ Similar problems might be created by rules which make it very time-consuming to change the constitution, *e.g.* by requiring ratification by two successive parliaments.

Ideally, we would like to be able to distinguish between bad and good motives for wanting to break the rule, the existence of the former being the reason to set

³⁶ W. Nordhaus, "The political business cycle", *Review of Economic Studies* 42 (1975), pp. 169-90, at p. 188.

up the rule in the first place and the latter being legitimate exceptions due to unforeseen circumstances. We need rules, but we also need to know when to make exceptions.³⁷ (After all, sometimes one does have good reasons for canceling an appointment with the dentist.) Individuals can use a variety of devices to make this distinction, although they are always fragile and vulnerable to self-deception.³⁸ By contrast, it is hard to see how a political system could have built into it not only first-order safeguards against impulsiveness but also second-order safeguards against unreasonably strict adherence to the first-order safeguards. It is not a question of guarding the guardians, but of making them lay down their guard when necessary. There may be institutional ways of ensuring this, but I haven't come across any.

Weakness of will is failing to do what you have willed to do. Excess of will, or hyperrationality, is a more subtle form of irrationality. It amounts to "willing what cannot be willed"³⁹ —using instrumental rationality to achieve benefits which are essentially linked to behaviour undertaken for other purposes than that of achieving these benefits. If I succeed in finishing this article, and in giving shape to the vague ideas I had when I sat down to write it, I shall be happy. I know, however, that if I focus on the happy state of having written the article, I shall never finish it and hence never achieve that happiness. (As Clausewitz says somewhere, a general who disposes his forces in terms of how they will be most useful to him after he has won the battle, is likely to lose it.) Moral psychologists have always known that happiness is essentially a by-product of activities undertaken for other

³⁷ A useful but ultimately somewhat disappointing study of this problem is R.B. Edger-ton, *Rules, Exceptions, and Social Order*, Berkeley: University of California Press 1985. Coming from an anthropological rather than a decision-theoretic tradition, he neglects incentive effects of rules and the conflict between long-term stability and short-term expediency.

³⁸ G. Ainslie, "Beyond microeconomics".

³⁹ Farber, *op. cit.*

ends than that of becoming happy. The same is true of self-confidence and self-respect. Many forms of behavioural therapy, including manuals of self-help, fail by encouraging a short-cut to a goal that essentially requires more roundabout, time-consuming methods. "[The] difficulty of denying one's intentions may help explain the limited success of behavioural therapies. Clients are trained to act assertively, but they do not feel like assertive people because they know that the behaviour is a deliberate attempt to create an assertive image and is thus an invalid indicator."⁴⁰

These problems are also relevant for political rationality. Consider Tocqueville's argument that democracy is to be praised "much more on account of what it causes to be done than for what it does."⁴¹ While the democratic political process fails in its immediate goal of making good decisions, "it spreads throughout the body social a restless activity, superabundant force and energy never found elsewhere."⁴² These benefits of democracy would surely evaporate if people ceased to believe in the immediate goal. It is because and to the extent that they take politics *seriously* that their participation has these desirable side effects. Similarly, there are activities which, in the eye of the observer, are justified mainly by the self-respect they induce in the participants, but only on the condition that the latter think they have an independent goal worth achieving. People may achieve self-respect by joining political movements or by getting a job, but not if they know that the movement or the job has been created solely for this purpose.⁴³

There is one apparent difference between individual

⁴⁰ G. Quattrone and A. Tversky, "Self-deception and the voter's illusion", in J. Elster (ed.), *The Multiple Self*, pp. 35-58, at p. 48.

⁴¹ *Democracy in America*, New York: Anchor Books 1969, p. 243.

⁴² *Ibid.*, p. 244.

⁴³ For further discussion of these two examples, see my *Sour Grapes*, Ch. II.9 and "Is there (or should there be) a right to work?", forthcoming in A. Guttman (ed.), *Democracy and the Welfare State*.

and political rationality in this kind of case. To succeed, an individual must hide his intentions from himself, which may be difficult. Deception, on the other hand, is easier than self-deception. The government cannot give people self-respect by setting them to dig ditches and fill them up again, but it can do so by creating jobs which appear to be productive and useful even if in fact they are not. The government may encourage people to participate actively in politics, arguing publicly that this will improve the quality of decisions and privately that it will have good economic side effects to offset the lower quality of decisions. These practices are not, however, compatible with democracy. Following Kant and Rawls, we may impose a *publicity* condition on policy choices that excludes deception and manipulation: the government should not adopt any policy whose efficacy depends on its not being made public. More generally, we may require that no policy should be adopted if defending it to its beneficiaries or victims involves a pragmatic contradiction.⁴⁴ Examples of such pragmatically incoherent policies, violating the publicity condition, are given in (VI) below.

One might be tempted to think that the democratic political process is only concerned with defining ends, and that the choice of means, including the choice whether to lie or to tell the truth, can be left to the government. Little reflection is needed to see why this view would be inadequate. Unless the government has an *obligation* to lie to the people, about devaluations and other acts whose efficacy depends on their not being anticipated, it is obliged to tell the truth. The citizens would want the government to lie about devaluations, but not to give them self-respect by fraudulent devices. Hence the condition of publicity serves to bring individual and political rationality closer to each other,

⁴⁴ For the notion of pragmatic contradiction see my *Logic and Society*, Chichester: Wiley 1978, Ch. 4 and *Sour Grapes*, Ch. II.

by insisting that information about the means used to implement social choices be publicly shared.

Weakness of will and excess of will arise when the agent's values fails to motivate the proper action or, on the contrary, motivates action when none is appropriate. These phenomena do not reflect on the values themselves, only on their causal efficacy or lack of it. A more deeprooted form of irrationality can inhere in the values themselves. As explained in (II) above, there is no generally accepted theory of rational desires. I shall simply assume that some desires are intrinsically or extrinsically irrational: intrinsically if they lead to self-destructive behaviour, extrinsically if they have the wrong sort of causal history. Granting this assumption, a main goal of politics must be to shape individual preferences so as to purge them of irrationality. Political institutions are not simply an aggregation mechanism from individual preferences to social choices. They also tend to shape and modify the individual preferences that are to be aggregated. To the extent that these effects are predictable, they should enter among the determinants of choice between alternative political systems, subject of course to the publicity condition discussed above.

It is widely assumed that *decentralization, participation* and *discussion* tend to enhance the quality of political preferences. I have discussed this view at some length elsewhere,⁴⁵ and I shall not repeat my arguments here, except for some brief remarks. The underlying idea behind this conception of politics seems to me to be valid, in the sense that participatory democracy based on discussion among equals is much superior to other forms of decision-making *when it works*. Anyone who has taken part in a collegial decision based on mutual respect can testify to this. When it does not work, however, it may be worse than the alternatives. Discussion easily degenerates into bargaining and logrolling, or

⁴⁵ In Ch. I.5 of *Sour Grapes*; see also "The market and the forum".

into the dominance of an informal elite. Since we do not have much robust knowledge about the conditions that would prevent these subversions of participatory democracy, it might seem rational to act as if the worse outcome is certain to occur. Following Hume, one might adopt the "maxim, that in contriving any system of government, and fixing the several checks and controls of the constitution, every man ought to be supposed to be a knave, and to have no other end, in all his actions, than private interest."⁴⁶ One might argue, moreover, that to the extent that the conditions for stable participatory democracy *are* known, they include an extent of economic equality well beyond what is presently realized in any democratic country. In a society characterized by substantial levels of inequality it might be disastrous to adopt institutions which presuppose that equality has been attained —disastrous, moreover, for the very attainment of that goal. "Getting there by acting as if one were there already" may work in some contexts, but surely not in all.

I return to these problems in the concluding section. To anticipate, I shall argue that these instrumental grounds for adopting participatory institutions are insufficient. Rather, they must be grounded in arguments about justice and equality.

V. *Uncertainty*

The major obstacle to rational political decision-making arises from the pervasive uncertainty about outcomes of policy choices. I shall argue that neither theory nor experience can help us predict the *long-term steady-state consequences of global policy changes*. Consider first theory as a guide to action. We do not have a theory of "general social and economic equilibrium" that could form a basis for prediction. We have a number

⁴⁶ Hume, *Essays; Moral, Political and Literary*. Oxford University Press 1963, p. 40.

of partial theories from which we can derive predictions about what will happen when other things can be assumed to stay constant, as they often can if the change is sufficiently small and the time scale sufficiently short. Yet all the partial theories do not add up to a global theory. As interaction effects become more numerous, and are allowed to work themselves out over a longer period of time, our predictive powers diminish rapidly. The belief in the power of reason to simulate and anticipate the course of social change is a form of rationalistic hubris that has been denounced many times, from Burke and Tocqueville to Hayek and Popper. Even disregarding the difficulties discussed in (III) above, *i.e.* assuming that the state has the capacity to form and carry out decisions, it would still be acting in a fog, unable to see more than a few steps ahead.⁴⁷ An individual can plan for the future because and to the extent that most of the relevant factors can be taken as exogenously given. In major social decisions, however, all variables must be taken as endogenous.

The natural response to these difficulties has been to say that even if we cannot use theory to predict outcomes, we can learn from experience, as argued by theorists of incremental planning or piece-meal social engineering. On this view, political reform is a process of gradient-climbing, based on trial and error.⁴⁸ There are several gradients to be climbed, simultaneously or successively. First, there is a policy gradient. If the policy is a package of measures, one may try to implement them one by one. If the policy variable is continuous — a tariff or a tax rate — it can be changed

⁴⁷ There are, then, three objections to the unitary-actor model of political decision-making. (i) From the general principle of methodological individualism, it follows that the community cannot *literally* be seen as an actor with desires, beliefs and the capacity to act. (ii) From the more specific arguments set out in (III) it follows that one cannot even in general speak *as if* the community were a unitary actor. (iii) Even assuming that one could, the scope for rational action would be severely limited because of the problem of uncertainty.

⁴⁸ For a discussion of gradient-climbing, see Ch. I of my *Ulysses and the Sirens*.

in small steps. Secondly, there is an institutional gradient. If the policy is eventually intended to be applied to all institutions of a given kind, one may try it out by implementing it in a small subset. Thirdly, there is a temporal gradient. If the policy is justified by its long-term consequences, one may attempt to gauge these by looking at the short-term effects. The principle of incremental planning tells us to move along each of these gradients as long as the marginal net benefits are positive, and to stop when negative returns set in. Clearly, the method of incremental planning is a good recipe for avoiding disasters. On the other hand, I shall argue that many potentially valuable reforms will never get off the ground if this method is adopted. There are good reasons for thinking that the local optima produced by gradient-climbing are far from the global optima. The implicit belief that all global optima can be reached by step-wise improvements along the three gradients is deeply misguided.

Consider first the policy gradient. It is a bit of a truism to say that when a set of measures mutually presuppose each other, one cannot learn about their efficacy by implementing them one by one. Each of them might have bad effects in isolation, yet be useful when taken together. Consider the belief in the efficiency of competition. On the basis of economic reasoning, one might believe that the world would be better if all trade unions, employers' associations and other organized economic groups were forbidden, so as to get rid of monopoly effects and deadweight losses. Yet one might have sufficient insight into the limit of abstract reasoning to abstain from implementing all these changes simultaneously. Instead, one might try to eliminate monopolies in sector by sector, to see whether the alleged benefits of unrestricted competition are born out by experience. Yet "It is *not* true that a situation in which more, but not all of the optimum conditions are fulfilled is necessarily, or is even likely to be, superior to a situation

in which fewer are fulfilled. It follows, therefore, that in a situation in which there exist many constraints which prevent the fulfilment of the Paretian optimum conditions, the removal of any one constraint may affect welfare or efficiency either by raising it, by lowering it or by leaving it unchanged.”⁴⁹ For another example, consider the introduction of market mechanisms in Eastern Europe. Markets for consumption goods and even for capital goods might not have the anticipated effects in the absence of the complementary financial markets.⁵⁰

Consider next the institutional gradient. There are many reasons why effects observed when a reform is implemented locally need not generalize to the global case. As an example, I shall use the introduction of worker-owned, worker-controlled firms. First, there is the Hawthorne effect: the mere fact of participating in an experiment may motivate people to efforts which would not be forthcoming were the practice to become general. Next, there can be positive or negative self-selection. The first firms to introduce workers’ self-management might attract exceptionally motivated and skilled workers—or they might attract “unstable individuals, excessive risktakers, and people lacking in pragmatic orientation.”⁵¹ Furthermore, there can be positive or negative discrimination. Workers’ cooperatives might have less easy access to credit than their capitalist counterparts—but they might also attract ideological supporters who would help them out in times of crisis. Finally, there can be positive or negative externalities. Isolated workers’ cooperatives might go under by failing to internalize

⁴⁹ R.G. Lipset and K. Lancaster, “The general theory of the second-best”. *Review of Economic Studies* 24 (1956-57), 11-32, at p. 12. See also A. Margalit, “Ideals and second bests”, in S. Fox (ed.), *Philosophy for Education*, Jerusalem: Van Leer Foundation 1983, pp. 77-90.

⁵⁰ See T. Bauer, “The unclearing market”, in J. Elster and K.O. Moene (eds.), *Alternatives to Capitalism*, forthcoming.

⁵¹ L. Putterman, “Some behavioral perspectives on the dominance of hierarchical over democratic forms of enterprise”, *Journal of Economic Behavior and Organization* 3 (1982), 139-60, at p. 152.

some of the positive externalities, such as the effect of training on workers who later leave the firm to take a job elsewhere. Conversely, if cooperatives are good at imitating but bad at innovating, they might do better in isolation than they would as part of a system of such cooperatives.

Consider finally the temporal gradient. There are two main reasons why short-term effects might differ systematically, in size and even in sign, from long-term effects. First, the immediately observed effects could be transitional ones, which would not be present when the system settles down to its new steady state. If outcomes are produced by institutions and individual goals in conjunction, and if institutional change give rise to endogenous changes in goals and preferences, then any reform should be judged by the outcome produced by the new institutions jointly with the new preferences, not by what happens when existing preferences operate within new institutions.⁵² The transient effects might be worse than the steady-state outcome, as argued by Tocqueville when he distinguished between the effects of democratization and the effects of democracy on key social variables, such as social cohesion, dogmaticism of beliefs, moral standards, or ambitiousness.⁵³ The transient effects might also be superior to the steady-state effects. If one system is better at creating resources and the other better at using them efficiently, a transition from the former to the latter might yield temporary improvements but the new steady state could be inferior. This points to the need for a second distinction, *within* the steady state, between short-term and long-term effects. Tocqueville, Schumpeter and others have insisted on the fact that static waste and inefficiency may be a condition for long-term growth, both considered as steady-state features of the system.

⁵² *Critique of the Gotha Programme*; Putterman, *op. cit.*, p. 149.

⁵³ See my "Consequences of constitutional choice: Reflections on Tocqueville", forthcoming in Elster and Slagstad (eds.), *Constitutionalism and Democracy*.

The upshot of this discussion is that for society to achieve global maxima by trial and error, it might be necessary to initiate large-scale, long-term experiments. Small experiments are simply too inconclusive. The question then becomes whether citizens would ever accept to participate in reforms the outcome of which is shrouded in uncertainty. Why would they accept to be guinea pigs in experiments the benefits of which might come much later if at all? The next section addresses this issue.

VI. Conclusion: Justice as guide to political action

The main political reforms of the last century have not been supported by instrumental considerations. Rather, they have been carried by social movements anchored in a conception of justice. I shall illustrate this proposition by three examples: the extension of suffrage, the rise of the welfare state, and some current proposals for economic reform. The thrust of my argument is that to the extent that the principle underlying economic reforms is perceived as being fundamentally just, people are willing and motivated to put up with the costs of transition and of experimenting with different modes of implementing the principle. Those who find this statement excessively idealistic may be more persuaded by an alternative formulation; if a reform is widely perceived as fundamentally just, it is very difficult to oppose it in more than a half-hearted way. It is usually easy to distinguish real opposition to reform from rearguard actions which are mainly designed to delay the inevitable.

In democratic societies, suffrage is of necessity restricted by age and citizenship (or residence). Beyond these, no restrictions appear to be inherently necessary, and in most democratic societies there are today very few other limitations on suffrage. In the past, however, restrictions have been numerous and strong. They can be distinguished according to their substantive content or,

more usefully, according to the underlying motivation.

Consider first economic restrictions, such as ownership of property or payment of taxes. These have been justified in at least four ways.⁵⁴ First, prior to the introduction of the secret ballot economic well-being was often seen a guarantee for *integrity*, which in turn was thought to be necessary to prevent the voters from being bribed.⁵⁵ Next, ownership of property has frequently been seen to provide the owners with special *qualifications* to take part in politics, either because ownership was seen as a proxy for education (which is why property owners often have been exempted from literacy tests), because it was thought to ensure the indispensable free time, or because it was thought to induce in the owners an interest in the long-term welfare of society, as distinct from a desire for immediate gain. Landowners, in particular, have been favoured on this ground. Furthermore, poll taxes have been advocated on the grounds that the willingness to pay them demonstrates a higher *motivation* and concern for political matters.⁵⁶ (The main argument for poll taxes, however, has been as a criterion for qualifications to vote.) Finally, economic restrictions have been justified on grounds of *commutative justice*: no taxation without representation *and vice versa*. Of these arguments, the first three are clearly instrumental, in the sense that they aim at bringing about substantively good decisions. In John

⁵⁴ The following draws heavily upon C. Williamson, *American Suffrage: From Property to Democracy 1760-1860*, Princeton University Press 1960; D.O. McGovney, *The American Suffrage Medley*, University of Chicago Press 1949; C. Seymour and D.P. Frary, *How the World Votes*, vols. 1-2, Springfield, Mass.: G.A. Nichols 1918; H.C.H. Metthew, R.I. McKibbin and J.A. Kay, "The franchise factor in the rise of the Labour Party", *English Historical Review* XCI (1986), 723-52.

⁵⁵ Actually, the argument only shows that bribing rich voters is more costly, which may be offset by the fact that with suffrage restricted to the rich there are fewer voters to be bribed.

⁵⁶ Stephen Holmes has pointed out to me that the Romans imposed economic conditions on the right to vote in order to elicit information from the citizens about their taxable property. In theory, this rule could also serve the purpose of sorting out the citizens who were sufficiently concerned about voting to risk bringing their property to the attention of the authorities.

Ely's terminology, they would pass "the rational basis test" and might even pass the "special scrutiny test" of what constitutes an admissible classification.⁵⁷ The last argument is grounded in considerations of justice, but of a very special and restricted kind, as I argue below.

Most other restrictions fall in one of these categories. Thus the link between universal suffrage and universal military service is also grounded on considerations of commutative justice.⁵⁸ Disfranchisement of serving soldiers, on the other hand, has been justified on the grounds that they are transient members of the local population with no interest in its long-term welfare.⁵⁹ The argument has also been used against student representation on the governing bodies of universities, and to support stringent residence requirements for the right to vote in local elections. Literacy tests are supposed to sort out the qualified voters from the less qualified. Disfranchisement of the mentally ill is similarly justified on grounds of competence. Disfranchisement of felons, during the period of confinement or for a longer period, may be justified on grounds of commutative justice, as an inverse of the principle "No taxation without representation". It is likely, however, that legislators may also have been influenced by the idea that the political opinions of convicted criminals tend to be twisted or unsound, and hence should not be represented.⁶⁰ Exclusion of women, finally, has been justified on grounds of competence or on grounds of commutative justice (since women do not do military service).

The argument from commutative justice rests on the vision of society as a joint-stock company, with the citizens cooperating for mutual advantage. Although taxpayers

⁵⁷ J. Ely, *Democracy and Distrust*, Cambridge, Mass.: Harvard University Press 1980, pp. 31, 120 ff., 146 ff.

⁵⁸ Athenians citizens were disfranchised for cowardice in war and for unpaid debts to the state (MacDowell, *op. cit.*, pp. 160, 165).

⁵⁹ Ely, *op. cit.*, p. 120.

⁶⁰ Thus Aiskhines in the speech *Against Timarkhos*: "The legislator considered it impossible for the same man to be bad privately and good publicly", (cited after MacDowell, *op. cit.*, p. 174).

may be willing to have some of their taxes spent on non-taxpayers, they would insist on taking part in the decision to spend the money that way and, crucially, they might also insist on *excluding non-taxpayers from that decision*.⁶¹ "No representation without taxation." I shall later consider a crucial ambiguity in the term "non-taxpayer," which may include those who are permanently unable to work and thus to pay taxes, as well as those who are temporarily out of work. For the time being, it is sufficient to note that on either acceptance, the denial of the right to vote to non-taxpayers (or, for that matter, to those who cannot perform military service) rests on a very narrow conception of justice. It is a vision of democracy as the result of a bargain among self-interested individuals: we shall pay the salary of politicians if they allow us to kick them out of office; we shall defend the country if the country gives us a say in defining what we defend.

Universal adult suffrage rests on a simpler and more compelling conception. Society is indeed a joint venture, but the bond between the members is not simply that of mutual advantage, but also one of solidarity. Just as the first step in the development of democracy was the idea that no minority of persons can be assumed to be inherently superior to others, so the second step was the idea that no minority can be assumed to be inherently inferior. If a low-income individual is excluded from voting on the grounds that the overall quality of decisions will be improved if low-income persons are excluded, any given person in this category might be justifiably offended by being subsumed under a statistical generalization which will inevitably have many exceptions. Even assuming the exclusion of individuals in a certain category to be instrumentally justified in each and every case, this procedure would not respect the publicity constraint. One cannot coherently try to make individuals understand that they should be kept out of politics because they are in-

⁶¹ This is, of course, how rich countries today decide on aid to poor countries.

capable of understanding. More simply, exclusionary procedures would no respect the self-respect of the excluded individuals. In any case, the excluded individuals would suspect, with good reason, that the decisions taken by the enfranchised would not be guided by the concern of eventually incorporating them.⁶²

It is often argued that the extension of suffrage must be understood in terms of legitimacy requirements.⁶³ Governments and ruling classes have successively removed restrictions on suffrage as concessions they were forced to give in order to retain legitimacy. Because of the risk to themselves involved they have usually done so reluctantly, with the exception of exceptionally far-sighted leaders like Bismark who minimized the risk by granting voting rights well in advance of any legitimacy crisis. If universal adult suffrage had not eventually been granted, there would have been massive discontent and social unrest; hence governments elected by a smaller electorate acted instrumentally rationally in extending the suffrage. This conception may be true as far as it goes, but it falls well short of a full explanation. Arguments from legitimacy presuppose other arguments. Unless the extension of suffrage was perceived by the population at large as desirable on other grounds, governments would not lose legitimacy by not acceding to it. My claim is that among these other grounds, arguments from justice have been important. The granting of voting rights to women may provide the clearest example. This was not carried by a political party with the purpose of using the vote to promote women's economic or social interests. Rather, the disfranchisement of women was seen as inherently insufferable and degrading. In the case of the working-class movement, this motivation was intertwined, sometimes inextricably, with the struggle for economic interests.

⁶² Ely, *op. cit.*, p. 120 ff.

⁶³ See for instance J.R. Freeman and D. Snidal, "Diffusion, development and democratization: Enfranchisement in Western Europe", *Canadian Journal of Political Science* XV (1982), 299-329.

It should be clear, however, to any reader of E.P. Thompson's *The Making of the English Working Class* that the struggle for manhood suffrage was largely motivated by arguments from simple justice.⁶⁴

The rise of the welfare state is analogous to and intertwined with the extension of suffrage. First, let me distinguish between two aspects of the welfare state. On the one hand, there are activities which take the form of compulsory saving or compulsory risk-pooling,⁶⁵ with no redistributive elements. On the other hand, there are essentially redistributive activities. Although most welfare services combine risk-pooling and redistributive aspects, it will nevertheless prove useful to distinguish them.

The welfare-state element in the first activities derive from their compulsory character. Individuals can and do save privately for their old age and take out private insurance against illness, accidents or (more controversially) unemployment. Increasingly, however, payments of insurance premiums are removed from the free choice of individuals and become a matter of compulsory payroll deduction.⁶⁶ Sometimes, compulsory compensation schemes retain the actuarial basis of private schemes. In that case, the arguments for introducing them can only be paternalist or self-paternalist ones. Through their politicians, people may bind themselves to measures they would want to take as private citizens were it not for their predictable weak-

⁶⁴ One might consider a purely instrumental-utilitarian argument for extension of the suffrage, by arguing that the removal of this degrading discrimination *ipso facto* represented a gain in welfare. Yet again this instrumental consideration would be parasitic on a non-instrumental one, namely the perceived inherent injustice in the unequal treatment.

⁶⁵ Strictly speaking, there are no instances of compulsory saving. Yet we may distinguish between the welfare services, such as old age pensions, in which the element of saving is large and the element of risk-pooling is small, and those in which the latter predominates.

⁶⁶ Formally, this is often presented as contributions of the employer. Economists agree, however, that these are *de facto* payroll deductions, in the sense that without compulsory employer contribution the salaries of employees would have been higher by the same amount. (B. Page. *Who Gets What from Government?*, Berkeley: University of California Press 1983, p. 28.)

ness of will. Usually, however, compulsory schemes deviate from private insurance in two ways: they are neither actuarially correct at the individual level nor self-financing at the collective level.

Compulsory insurance is often accompanied by redistributive measures, as when people do not get back in old age the actuarial equivalent of what they have paid in over the years. True, redistributive features also characterize most private insurance schemes, as an unavoidable consequence of the fact that "because no risk class is completely homogeneous, there always appears to be some subsidy of the slightly higher risks within a class by the slightly lower risks".⁶⁷ Redistributive aspects of social insurance deliberately go beyond these unavoidable effects, usually in an equalizing direction. This is increasingly also true of private insurance companies, when they are forbidden by law to use certain classifications to distinguish between risk classes. For instance, if sex-classifications were disallowed, "males would subsidize female pension rates, and females subsidize male life insurance rates".⁶⁸ Yet this policy easily leads to absurdities. For example, "it seems inappropriate to ask disability insureds to bear the full costs of subsidies to hemophiliacs."⁶⁹ Similarly, as soon as society decides to use compulsory insurance for redistributive purposes, it becomes inappropriate to require each separate program to be self-financing. In fact, there is no point in requiring even the whole set of programs to be self-supporting, since there is no reason to keep this form of redistribution neatly separate from redistribution through taxation. The upshot is "the welfare state": a system in which the original correlation between premiums and benefits has almost totally disappeared.

For these reasons, compulsory risk-pooling and redistribution are almost inseparable in the modern welfare state.

⁶⁷ K. Abraham, *Distributing Risk*, New Haven: Yale University Press 1986, p. 84.

⁶⁸ *Ibid.*, p. 92.

⁶⁹ *Ibid.*, p. 99.

To see that the distinction is not useless, however, it suffices to note that the welfare state covers many disabilities against which one could never get private insurance. People with congenital blindness or easily detectable genetic defects cannot insure themselves against these disabilities, since one cannot insure against an event that has already taken place. At the other end of the spectrum, some parts of social insurance still obey actuarial principles, at least approximately.⁷⁰ Hence I shall keep referring to risk-pooling and redistribution as two separate aspects of the welfare state, corresponding respectively to the values of security and solidarity.

The distinction can also be stated in a different language, which will prove more useful for the present purposes. Many theories of distributive justice agree on the formal point that a just distribution is that which would be chosen behind "the veil of ignorance", but are in substantive disagreement as concerns the "thickness" of that veil. In different but essentially equivalent terminology, the theories may agree that the distribution of goods and welfare should not be affected by "morally arbitrary features" of individuals, but disagree about the criteria for what is arbitrary and what is relevant. Risk-pooling takes place behind a very thin veil, which allows people to know their actual skills, preferences and wealth, but not their future earning power and earning opportunities. Under these circumstances rational individuals will agree to insure against risk, *i.e.* to pay a premium into a common pool out of which compensation can be made. Redistribution takes place behind a much thicker veil, which denies people knowledge about most, perhaps all of their personal qualities and property. Behind thick veils of ignorance people ask themselves how they would want society to be organized if they did not know which assets they will turn out to have. Rational individuals would want to protect themselves against the risk of being born poor or poorly endowed with productive skills.

⁷⁰ Page, *op. cit.*, pp. 67, 75.

The notion of a thin veil of ignorance can be understood quite literally. We do not know what the future will bring, so it makes sense to take precautions. The thick veil, by contrast, cannot be taken literally, since we do know our skills, preferences, wealth, etc. The thick veils are only devices to express the idea that that the welfare of individuals ought not to be affected by certain morally arbitrary properties —those, precisely, from which abstraction is made behind the veil of ignorance in question. The thinnest of these thicker veils corresponds to a meritocratic conception of social justice, according to which people are entitled to the fruits of their skill and effort but not to fruits of inherited property. A somewhat thicker veil is that proposed by Ronald Dworkin, according to whom distribution of welfare should be “ambition-sensitive” but not “endowment-sensitive”.⁷¹ The most impenetrable veil is proposed by John Rawls, according to whom ambitions, including time preferences, risk aversion and the like, are just as morally arbitrary as skills.

The redistributive aspects of the welfare state presuppose that some qualities of individuals are seen as morally arbitrary. Minimally, these include inborn abilities and disabilities. The welfare state expresses a widespread belief that it would be unfair to let individuals suffer from genetic accidents outside their control. In this perspective the meritocratic conception appears inconsistent. If social luck is to be eliminated from the determinants of welfare, why should genetic luck be respected? Yet Dworkin’s position may also be criticized as inconsistent.⁷² How can one defend the view

⁷¹ R. Dworkin, “What is equality? Part 2. Equality of resources”, *Philosophy and Public Affairs* 10 (1981), 283-345. For an illustration, consider three individuals, A, B and C. A and B have the same skills. B and C produce the same output. A and C work the same number of hours. In other words, the unskilled B is able to produce as much as the skilled C because he is willing to work more hours than his unskilled colleague A. Dworkin would endorse higher income to B, and relegate A and C to the same, lower income level. The principle “To each according to his unweighed labour time” was also used by the early British socialists (see U. Pagano, *Work and Welfare in Economic Theory*, Oxford: Blackwell 1985, Ch. 2.3.)

⁷² See notably J. Roemer, “Equality of talents”, *Economics and Philosophy* (1985).

that a low level of ambition is not also the product of social and genetic luck? If it is, should it not also provide grounds for compensation? This seems to be the central philosophical issue in current controversies over the welfare state.

Again, the condition of publicity may provide the beginning of an answer. To tell an individual that he is entitled to welfare because he is not responsible for his preferences is pragmatically incoherent. One cannot in one and the same breath treat an individual as rational and open to arguments, and treat him as moved by psychic causal forces outside his control. Perhaps one could justify this to third parties, but in a democratic society the policy must be rejected if it cannot coherently be explained to the individual in question. By withholding material benefits one may protect the crucial value of self-respect. Yet, as I said, this austere principle is only the beginning of an answer. Applied to contemporary societies, it would often be perceived as unfair, because of expectations that have been shaped by the existing welfare states and the massively unequal distribution of the ability to form autonomous preferences. This inequality, in turn, stems from differences in wealth which even meritocrats would admit to be morally arbitrary. As long as the influence of genuinely arbitrary properties has not been eliminated, justice may require us to count as arbitrary some properties which would be considered non-arbitrary if the former were removed.

These are, however, minor variations on a more general theme, which is the need for a veil of ignorance that excludes personal wealth and skills as morally relevant determinants of welfare. It is impossible to understand the development of the welfare state unless we see it as sustained by some notion of this kind. Just as political democracy is more than a joint venture for mutual advantage, in which individuals consent to taxation and military service in return for the right to vote, the welfare state is more than a system of mutually beneficial risk-pooling. Both systems comprise individuals who have no chips to use in bargaining and exchange, but who are included because it would be arbitrary, de-

grading and unjustifiable to exclude them. The overriding argument for democracy is that it makes certain arguments impossible to state in public, whereas others become endowed with almost irresistible power. Grumbings about democracy and the welfare state may often be justified, in terms of the short-term and transitional costs associated with them. Yet, as I said, these grumbings do not amount to more than a rearguard action. Because of the perceived injustice of hierarchy and inequality, or at least the pragmatic impossibility of defending them in public, the long-term trend towards participation and equality seems irreversible. The built-in commitment of democracy to these values is so strong that the instrumental language of costs and benefits does not really apply. If there are costs—and there will be—they will be accepted as inevitable parts of a learning process, not as reasons for halt or retreat.

Let me conclude by some brief comments on five recent proposals for economic reform. These are (i) the Swedish proposal for a wage earners' fund,⁷³ (ii) James Meade's proposal for a property-owning democracy, further elaborated recently by Richard Krouse and Michael McPherson.⁷⁴ (iii) proposals for a "social dividend" or guaranteed income at a level sufficient to provide a decent living without any obligation to work in return.⁷⁵ (iv) Martin Weitzman's model of a "share economy" in which workers and management would bargain over the relative shares in the net product

⁷³ I shall discuss the development of this proposal up to 1982, relying mainly on S.L. Albrecht and S. Deutsch, "The challenge of economic democracy: The case of Sweden", *Economic and Industrial Democracy* 4 (1983), 287-320 and H.G. Myrdal, "Collective wage earner funds in Sweden", *International Labour Review* 120 (1981), 319-34.

⁷⁴ J. Meade, *Efficiency, Equality, and the Ownership of Property*, London: Allen and Unwin 1964; R. Krouse and M. MacPherson, "A 'mixed'-property regime: Equality and liberty in a market economy", *Ethics* 97 (1986). See also my "Comments on Krouse and McPherson", *ibid.*,

⁷⁵ See Philippe van Paris and R. van der Veen, "The transition from capitalism to communism", *Theory and Society* 15 (1986) as well as my "Comments" on their proposal, *ibid.*

rather than over the absolute wage level,⁷⁶ and (v) economic democracy at the enterprise level, with full workers' control as the immediate or ultimate goal.⁷⁷ All of these would involve major changes of the capitalist organization of production as it currently exists. They are advocated on the grounds that they will promote efficiency, equality or participation —indeed, for many of the proposed systems it is argued that they will be superior on all of these counts. Against this I shall contend that with the exception of economic democracy, all the proposals are non-starters, because they do not rest on a simple, compelling conception of justice. They are engineering blueprints for Utopia, technocratic dreams or nightmares without the potential for animating a social movement. They have no chance of being adopted since people would feel, correctly, that they were being asked to participate in a large-scale experiment of no intrinsic value and highly uncertain extrinsic value. The general argument for uncertainty was sketched in (V) above. I shall now adduce more specific arguments for each of the proposals.

Consider first the Swedish proposal for creating wage earners' funds which would have given profitable firms a labour majority ownership within a few decades. Between 1975, when it was first put forward, and 1981, the proposal underwent considerable changes, as a result of political and economic objections. In the last version, the 24 regional funds would be financed partly by a supplementary pension contribution to be paid by the employer and amounting to 1% of the salaries, partly by a transfer of 20% of firms' "excess profits". The funds would buy shares in existing firms, forcing them to issue new shares if necessary. When the fund purchases shares, the voting rights would be divided between the fund and the employees in the enterprise in which

⁷⁶ M. Weitzman, *The Share Economy*, Cambridge, Mass.: Harvard University Press 1984.

⁷⁷ For discussion and further references, see the Introduction to Elster and Moene (eds.), *Alternatives to Capitalism*.

the shares are bought. The voting rights are divided equally between the fund and the employees until each has acquired 20% of the voting right; after that, all voting rights from new shares go to the fund. The administration of the funds remains unclear, but is likely to involve a majority trade union and a minority "social" representation, by appointment or election.

On efficiency grounds, the proposal seems doubtful at best. It would certainly take much out of the bite of the profit motive if firms knew they would be taken over if they made "excessive" profits. Also, the investment criteria of the regional funds are likely to represent regional interests rather than profitability. Rather than having a limited and welldefined portion of society's resources used to support ailing regions, one might now expect to see massive support for lame-duck industries. From the point of view of promoting social justice, it seems perverse to give employee voting rights only to workers in firms which for some reason happen to be chosen as investment objects for the funds. In any case, the ceiling of 20% on employee participation truncates those rights before they would be seen as worth while having. To argue that "the working class" as a whole would have control over the firms, through trade union representatives on the funds, is just as ridiculous as saying that the workers in the Soviet Union are the "real owners" of their plants. The real power would be vested in the trade union bureaucracy.

Consider next Meade's proposal for property-owning democracy, which rests on a combination of a progressive property tax and a radical reform of the inheritance tax system. The latter shall induce large holders of property to bequeath their wealth to a large number of relatively poor individuals. This could be achieved in either of two ways: either by taxing "each individual gift or bequest not solely according to the size of the individual gift or bequest, but also according to the existing wealth of the beneficiary", or by taxing the beneficiary "when he received any gift or bequest neither according to the size of that gift or bequest

nor according to the size of his total property at the time of the receipt of that gift or bequest, but according to the size of the total amount which he had received over the whole of his life by way of gift or inheritance".⁷⁸ According to Krouse and McPherson, this scheme would ensure that "everyone would start out in life with substantial property income". At the very least, "this would create psychological attitudes importantly different from current experience", since "workers of one firm would be part owner of others: under the authority of managers in one firm they will help oversee managers in others". It would also create the material means for workers to form cooperatives, while not requiring this to be the obligatory form of ownership.

Again, the alleged consequences of the proposal are highly doubtful. Either inheritance scheme would have perverse incentive effect mind-boggling problems of implementation or both.⁷⁹ Moreover, even if we disregard these problems, there is no reason to expect that everyone would be chosen by someone as a recipient for a gift or bequest. One can easily imagine the negative consequences for the self-respect of those to whom nobody had chosen to give or bequeath property. Furthermore, even if that difficulty is assumed away, the suggestion that owning shares in other firms could somehow offer a power which would compensate for being subject to managerial authority in one's own workplace is absurd.⁸⁰ The idea that widespread property-owning could facilitate the formation of workers' cooperatives is more attractive. If one believes in economic freedom as well as in economic democracy, voluntary formation of workers' cooperatives might seem preferable to a system with mandatory workers' control (as in Yugoslavia). I believe, however, that such control is best conceived as an inalienable right, on a line with the right to vote or the right to choose one's occupation or place of residence.

⁷⁸ Meade, *op. cit.*, pp. 56-57.

⁷⁹ See my "Comments on Krouse and McPherson".

⁸⁰ Here Meade is more realistic, when he argues that "investments [would have to be] chosen by specialists on behalf of the man in the street" (*op. cit.*, p. 40).

Economic freedom can be respected by leaving workers free to choose (but not irreversibly) any particular organization of the enterprise they want, including the possibility of leaving all decisions to a hired manager who receives part of his salary as a percentage of the net income.

Consider now proposals for a negative income tax, social dividends, universal grants and the like. There are obvious objections to the economic feasibility of a guaranteed, substantial and unconditional income for everybody. Here I shall simply argue that any such proposal would fail because it would be seen as unfair, indeed as exploitative.⁸¹ People who choose to work for an income rather than to live in a commune on the unconditional grant would have to pay higher taxes in order to support those who took the second option. They would think, correctly in my opinion, that they were being exploited by the latter group. Against this, the following counterargument has been made.⁸² By assumption, everyone would be free to choose the unconditional grant. If some people choose not to do that, they can hardly complain if others do. Their choice reflects a preference for consumption over leisure which is no reason for preventing others from acting on different preferences. To this I have two replies. First, some people might remain in the work force simply because they believe someone has to be in the work force. When contemplating the happy commune members they might mutter, angrily, "What if everyone did that?" Next, even if they do in fact prefer to work because they value consumption, this is not a reason for taxing them more heavily. They might prefer the forty-hour week over the fifty hour week which they have to work because of the higher taxes imposed on them by the people who choose the uncon-

⁸¹ For a related argument, see R. H. Frank, *Choosing the Right Pond*, Oxford University Press 1985, p. 256 ff.

⁸² See P. van Parijs and R. van der Veen, "Reply to six critics", forthcoming in *Theory and Society*.

ditional grant. Because the latter prevent them from achieving their desired level of welfare with the forty-hour option, the argument from freedom of choice doesn't work.

Consider now the proposal for a share economy. This is justified exclusively on grounds of efficiency, more specifically on the grounds that it will eliminate unemployment. The argument is that when the total workforce in the firm receives a fixed percentage of the net income, the management will always have an incentive to hire more workers as long as the marginal product of labour is positive, since the management will get the agreed-upon percentage of that product. By contrast, if collective bargaining takes place over the wage, management will stop hiring labour when the marginal product equals the wage. For my purposes, I do not need to argue that the share economy will fail to have the alleged consequences, only that it is highly uncertain whether it will work and that, moreover, it does not rest on a conception of justice that could provide the motivation to make it work. To justify my scepticism, consider the following arguments. In a situation with full employment, workers have a very strong position, even in the absence of labour unions. They can shirk at little risk to themselves, since if they are detected and laid off, they can easily get a new job at the same wage.⁸³ Weitzman's share economy might exhibit some of the sluggish features of Soviet-type economies. Moreover, if labour is organized, the bargaining power of workers is strongly enhanced in a situation of full employment. It is hard to imagine that they will not demand a say in the decisions of the firm —especially in the decision whether to hire more workers and thereby reduce wages.⁸⁴ Adding to

⁸³ For this argument see C. Shapiro and J. Stiglitz, "Equilibrium unemployment as a worker discipline device", *American Economic Review* 74 (1984), 433-44.

⁸⁴ D.M. Nuti, "The share economy: Plausibility and viability of Weitzman's model", European University, Florence: Working Paper No. 85/194 from the Department of Economics, 1985.

this the predictable opposition of union leaders, and keeping in mind the absence of non-instrumental arguments for the proposal, it seems doomed to failure.

I shall not say much about the final proposal, the introduction of workers' management and ownership, except to argue that it possesses the feature which the other proposals lack: a simple, compelling argument from justice. Hierarchy and subjection at work are just as arbitrary and degrading as similar authority structures in other walks of life. The abolition of economic inequality is, it seems to me, another built-in commitment of democracy.⁸⁵ No doubt this will be a long process, marked by trial and error, with progress taking the form of "two steps forward, one step backward". The exact timing and the exact forms it will take cannot be predicted. The possibilities for building tactical alliances to block it are numerous. Yet, once again, these will be rearguard actions, visibly motivated by vested interests. Over time, they will be eroded by the relentless arguments from justice, equality and participation —unless democracy itself as we know it ceases to exist.

⁸⁵ R. Dahl, *A Preface to Economic Democracy*, Berkeley: University of California Press 1985.

RESUMEN

Sabemos, más o menos, lo que quiere decir la racionalidad cuando se aplica a elecciones y decisiones individuales. No resulta tan claro lo que quiere decir cuando se aplica a elecciones políticas. En este artículo discutiré dos preguntas. Primero, ¿puede uno asignar un significado a la idea de elección política racional? En segundo lugar, suponiendo que esta asignación sea viable, ¿cuál es el alcance de las decisiones políticas racionales? En tercer lugar, si este alcance resulta ser limitado (como sucederá), ¿podría haber una guía alternativa a la acción política?

La sección (II) establece, muy brevemente, la teoría de la elección racional a un nivel individual. Supondré que la racionalidad política se define mediante una extensión de las elecciones individuales a las elecciones políticas que, en un sentido, están hechas por y para la "sociedad". Esto es, excluyo las concepciones de política que corresponden al napoleónico "Todo para el pueblo, nada por el pueblo". Dentro del marco de la teoría de elección social, esto significa imponer la condición de no-dictadura.

La sección (III) discute si la noción de racionalidad política es significativa, mediante la consideración de diversas objeciones que se derivan del principio del individualismo metodológico. Los individuos actúan sobre la base de sus deseos y creencias. Las colectividades, en cuanto tales, no actúan; tampoco tienen deseos o creencias. Por tanto, surgen problemas en relación con las preferencias agrupadas, la información centralizada y la instrumentación de elecciones. La sección (IV) considera el problema de la irracionalidad política, observando las analogías políticas de diversas formas de irracionalidad individual que pueden describirse brevemente como debilidades de la voluntad, excesos de la misma y la formación de preferencias adaptativas. Para que la política sea racional, debe diseñar maneras de evitar o manejar estas tendencias. La sección (V) se pregunta si la racionalidad política siempre producirá prescripciones claras y sugiere en términos generales una respuesta negativa, argumentando que la *incertidumbre* es el mayor obstáculo para las elecciones políticas racionales. El argumento no es simplemente que es difícil anticipar los resultados de las decisiones políticas. De manera más fundamental, consiste en que la planeación relativa al incremento, la ingeniería social y tácticas semejantes son respuestas inadecuadas a este problema. En la sección (VI) concluyo argumentando que la *justicia*, más que la *racionalidad*, debe guiar las elecciones políticas fundamentales. La racionalidad tiene un lugar que desempeñar, pero éste es subsidiario más que primario.

[J.E.]