Discourse on Rationality: Rational Choice and Critical Theory

Santhosh Kumar Madiraju

**Dissertation Submitted to the University of Glasgow for
the Degree of Doctor of Philosophy**

**Department of Politics
Faculty of Social Sciences
University of Glasgow
August 1996**

Abstract

The thesis contrasts two hostile and divergent intellectual paradigms in social sciences: rational choice and critical theory. Both rational choice and critical theory offer contrasting perspectives on the structures of social interaction. However, both critical theory and rational choice theory share overlapping concerns i.e., both are preoccupied with determining *what rational can mean* in the realm of social and political interaction.

In the case of rational choice paradigm, instrumental reason forms the cornerstone of the theoretical edifice. Ever since the publication of Jurgen Habermas' *The Theory of Communicative Action Vol. I* (1984) and *Vol. II* (1986) instrumental reason has come under severe attack. His critique anchors on a theory of communicative reason. What makes Habermas' work distinctive is that he does not regard instrumental reason as the single inevitable concomitant of modernity. Habermas sees in modernity an alternative way of conceptualising social interaction in terms of communication rather than strategy. So in a way, his work is a challenge to the defenders of modernity aiming to build a unified social science.

Jurgen Habermas advances the notion of communicative reason as the centerpiece of a social theory as opposed to instrumental reason. By providing a systematic grounding of the concept of reason in human language, he hopes to establish normative basis of critical theory. This model of reaching agreement or consent constitutes a process of dialogue in which reasons are exchanged between participants. This process is perceived to be a joint search for consensus. Such a dialogic concept of collective choice would necessarily work not with fixed preferences to be amalgamated (as rational choice theories do) but with preferences that are altered or modified as competing reasons are advanced in the course of discussion. In rational discussion, the only thing supposed to count is the power of better argument.

Both rational choice and critical theory conceptualise politics in different ways. Rational choice theories critique democratic mechanisms failing to generate general will. Consequently, the political prescriptions offered are limited government or market. On the contrary, the political implications of Habermas' theory of deliberative democracy is anchored in the notion of liberal public sphere envisaging a cognitivist, rationalist vision in which discourse forms a critical normative basis for evaluating the political and moral principles.

Acknowledgements

Analytical Contents

Intellectual histories of rational choice and critical theory

Before contextualising the problem of my thesis, let me briefly run through the mutually hostile intellectual histories of rational choice and critical theory. Critical theory was a product of the Institute for Social Research and the group is referred to as the Frankfurt School, but it is misleading since the members of the Institute rarely shared a readily identifiable collective position and the differences between them overshadowed the similarities. The founding members of critical theory include Max Horkheimer, Theodore Adorno, Herbert Marcuse etc.

One of the significant developments in the trajectory of critical theory was the shift from the project of producing multidisciplinary social theory rooted in the Marxian critique of political economy to that of a new philosophical critique of science, technology and instrumental reason. It found expression in Horkheimer's and Adorno's *Dialectic of Enlightenment* (1947), which combined a novel critique of western civilisation and rationality with a powerful critique of the administered society, which became the target of their social criticism. Although Horkheimer and Adorno extended their critique of fascism and capitalism, they distanced themselves from the Marxian theory of history and critique of political economy, subordinated science and politics to philosophy, and tended to make the domination of nature the fulcrum of their analysis, rather than such classical Marxian themes as political economy, social relations, class struggle or the transition to socialism.

Critical theorists saw themselves giving voice to dissent in the form of negative dialectics etc. At the same time they were critical of Marxist reductionism and replaced the critique of political economy with a critique of instrumental reason. Recognising the changing social and political relations with the advent of

science and technology, critical theorists focused their critique on science and technology. Soon it became an intellectual paradigm offering critical perspectives on culture, art and aesthetics, media, politics, methodologies of natural and social science. Critical theory addresses a full range of issues raised by the impact of modernity on the spheres of economic, political and social and cultural life. Moreover it provides an excellent access to modernity because it has incorporated the positions of some of the major theorists and critics of modernity in its work. Although it has been closely connected with Marxism; in the 1930s it began engaging the positions of Nietzche, Freud, Weber, Heidegger and others who have carried out both rationalist and irrationalist critiques of modernity. Eventually critical theorists incorporated many positions from these thinkers, and thus were able to provide much more complex, multidimensional and stimulating perspectives on modernity than those who took a more one-sided approach. Concepts like critique and ideology have been redefined to offer a critique of advanced capitalist rationality.

My primary concern is with the work of Habermas. What makes Habermas' work interesting is his critical defense of the project of modernity. Unlike his predecessors who ended up in giving a romantic critique of modernity, Habermas anchors his project in a theory of communicative rationality and ethics. Moreover Habermas' critical engagement with the analytical tradition also gives an added advantage in making sense of the esoteric continental debates. Habermas conceptualises enlightenment as a process of differentiation of cultural spheres and the development of autonomous criteria of rationality and universality in the fields of knowledge, morality, law and justice, and art. The project of modernity also has unrealised potential to increase social rationality, justice and morality. For Habermas the project of modernity resulted in part in the colonisation of the life

world by the logic of scientific-technological rationality and domination by the culture of experts and specialists.

Critical theory and rational choice theory provide two radically different versions of modernity. Rational choice theory is firmly rooted in the project of modernity and it has unreflectingly assumed that there is a unique rational answer to most questions. Like the critical theory, rational choice paradigm is not a homogeneous theoretical enterprise. It typically includes multiple theories like rational choice, public choice, social choice and game theory etc. Although there is an underlying common theme running through their works there are deep differences in their solutions to the problems. The classics of rational choice include work by intellectual figures like James Buchanan, Anthony Black, Anthony Downs, and Mancur Olson.

In this thesis, I contrast these two theoretical traditions and assess the implications of these theories for democratic politics. The two traditions can also be grouped under *homo economicus* and *homo sociologicus*. Economics and rational choice are commonly thought to be focused on the narrow concern of what the actor calculates is best to do. Sociology is commonly thought to be about matters in which individuals do not have choices, cannot calculate, and often are guided for odd or even in rational reasons. Although this contrast is pristine, it makes for a seemingly neat disciplinary dividing line between the two paradigms.

Defining Collective Choice: the theoretical dilemma

Rationality has always been at the centre of the debates in social and political theory. In trying to theorise the logics of social interaction, two dominant paradigms emerged offering distinct perspectives on the modes of social interaction. One is broadly called rational choice theory (which includes public

choice, social choice and game theory) and the other Critical theory (as represented in Habermas' writings). The two theoretical paradigms are mutually hostile in their intellectual histories and what demarcates them is their divergent emphases on contrasting notions of rationality - instrumental reason and communicative reason. Rational choice theories attempt to provide a 'theory of rationality' that can be applied in principle to all interactive situations. Theorists working in this paradigm apply economic methods to politics and claim to provide solid microfoundations for the study of social interaction.

The intellectual opposition to instrumental reason has been articulated by the founding members of the Frankfurt school like Max Horkheimer, Theodore Adorno and Herbert Marcuse. Both *Dialectic of Enlightenment* (Horkheimer and Adorno: 1972) and *One-dimensional Man* (Marcuse: 1964) challenged modernity, seeing it as the expansion of instrumental reason. But their challenges were devoid of any normative basis for critical theory. Continuing in this tradition and at the same time critical of his predecessors Jurgen Habermas redefined the definition of critical theory in terms of Communicative reason. The strength of Habermas' two-volume *The Theory of Communicative Action* (1984: Vol. I) and (1987: Vol. II) lies in articulating an elaborate theory of communicative reason showing the limitations of instrumental-strategic rationality. His alternative theoretical programme investigates the potential for organising actions around communicative action rather than merely instrumental or strategic action. Habermas' critical defense of modernity sees 'modernity' as a rationalisation process crystallising in not just instrumental reason but also communicative reason. Habermas precisely challenges this 'identity thinking' and draws our attention to the multiple dimensions of reason. In the light of this critique, I would like to explore the implications, both explanatory and normative, for political theory.

4

But surprisingly not many social and political theorists are enthusiastic to explore the nature of the challenge posed by this work. It is also surprising to see the common concerns of both paradigms. Both are preoccupied with determining what rationality can mean in the social and political world. Both are defenders of enlightenment reason and modernity but in different ways. Both argue for a social science which is empirical: rational choice theory attempts to discover solid micro foundations for social sciences and on other hand critical theory in its present form advocates reconstructive sciences. Rational choice aims to build a unified social science and critical theory attempts a critical social science. Attempts to bring in both the theoretical paradigms to engage in critical conversation are not without problems.

Given their apparently divergent claims about rationality and politics, the aim of this study is to initiate conversation across the two contrasting traditions. I would briefly set out the problem for political theory. The theoretical problematic involves defining 'collective choice.' Insofar as politics involves conflict of interests, disputes among the members of the community, it is a realm of collective choice. In a collective choice the decision of some is binding on all. Moreover the decisions taken are seen as representing the expression of the popular will. The notion that politics involves a general will is widely accepted idea in political theory. This idea that democratic mechanisms represent the general will has come under severe attack from rational choice theories. They criticised the notions such as "general will" or "national interest". So the theoretical dilemma for political theory is: how do we define collective choice? What constitutes a fair and just rule? Two theoretical traditions with hostile intellectual histories have offered contrasting theoretical explanations. One is rational choice and the other is critical theory. I argue that rational choice theories are inadequate in that they fail to provide any conceptual handle to overcome the problem.

Admittedly, the rational choice programme is not a monolithic paradigm. It constitutes multiple theories having their own rationality assumptions. So the term 'rational choice' is employed loosely to include literature that sometimes travels by the names public choice theory, social choice theory, game theory and rational actor theory. The literature which emerged under the title public choice theory includes works such as James Buchanan and Gordon Tullock (1962), Anthony Downs (1957), Mancur Olson (1965), William Riker (1982). Public choice theory works with a model of instrumental rational self-interested maximisation. They claim to have exposed the myth of 'social choice' through a series of discoveries i.e. e.g. aggregation from individual to collective preferences may not be well defined. Hence there can be no acceptable universal rule to convert any collection of individual preferences into collective preferences. The disconcerting implication of public choice theory is that majoritarian democracy is both conceptually and motivationally flawed. If democracy is defined as the consent of the governed based on majority rule then lack of agreement is the modal problem of democracy. In the words of Barry and Hardin (1982), public choice paints a picture of 'rational man and irrational society'. The tradition claims to have demonstrated the impossibility, incoherence or undesirability of democracy.

Given these dismal conclusions of public choice theories on the functioning of democratic institutions and the legitimacy of the state, social scientists, under the banner of rational choice theory, have been quick to explore the problem of collective choice from a strictly rational self-interest perspective. However, the vacuousness of this model has come under increasing attack from within rational choice theory: especially the concept of preference formation. This critique is most powerfully articulated by Jon Elster (1983a 86a; 89a). Elster (1986b) himself recognised the strengths of Habermas' discourse ethics in modelling collective

6

choice. This provides a good entry point for a productive conversation between rational choice and critical theory.

Unlike rational choice theories, critical theory conceptualises the 'subject' as constitutive of cognitive, moral and subjective aspects and not instrumental ones. Internal critiques of rational choice theorists like Elster is have begun to explore the work of Jurgen Habermas and have recognised the fragility of the subject modelled as homo economicus. By contrast critical theorists have always characterised rational choice as aiming to build a positivistic natural social science. The theoretical inquiry also has normative implications for democratic political theory. Rational choice theorists conceptualise democracy in non-cognitivist political language whereas critical theorists line up on behalf of 'moral-cognitivist political rationalism.' Politically, the public choice programme is hostile to democratisation, conceiving it as inherently unjust since it creates cycles. Public choice theorists have a narrow conception of democracy, interpreting it as a device for preference aggregation under conditions of non-cognitivism in individual preferences, instrumentally rational behaviour and self- interest. Non-cognitivism means simply that values and preferences are like emotions, beyond the reach of rational argument.

The practical implication of Habermas' theory is towards a politics of deliberation. Habermas wants to restore a conception of the political which is constituted by an autonomous public sphere.

Summary of the arguments of the chapters

Chapter 1

In this chapter, I discuss public choice theories by way of reviewing the major works of James Buchanan and Gordon Tullock (1962); William Riker (1982); Mancur Olson (1965); Anthony Downs (1957); All these four works deal with the problem of collective choice. Public choice applies economic methods to politics in relation to two central problems: the problem of preference aggregation (voting) and the collective action problem (free-rider problem). The first problem has shown the undesirability of majority rule as mechanism for aggregating preferences. The second problem has exposed the myth of collective action. Both problems have serious and interesting implications for democratic theory.

Chapter 2

In this chapter, I explore that challenge to instrumental reason articulated in Jurgen Habermas' theory of communicative action. His is an alternative model of collective choice that sees politics not as a process of preference aggregation but as dialogic process of advancing reasons. Habermas' communicative action appeals to a richer range of motivations other than purely instrumental ones. Communicative action is seen as an alternative way of organising social action.

Chapter 3

In this chapter, I contrast two different normative theories of collective choice. One of the public choice version of contract and the other discourse ethics. Contrasting Buchanan and Habermas, I argue that Buchanan's notion of agreement is inadequate. As regards normative collective choice, Buchanan's contractarian

solution can be seen as a model of agreement. On the other hand, Habermas' notion of discourse is not just seen as compromise and bargaining but as rational consensus based on advancing reasons. Interpreting discourse ethics, Benhabib contrasts two lines of interpretation of discourse ethics: one is a legal-juridical interpretation and the other is a moral transformatory interpretation. She argues that discourse ethics is not merely procedural and formal but can be contextualised in specific historical and social conditions.

Chapter 4

In this chapter, I shall discuss some of the internal critiques of rational choice theories relating to the limits of strategic rationality. Although the internal critiques vary in their content, almost all of them point to the problem of indeterminacy. To overcome this problem rational choice theories have appealed to norms as opposed to mere interests in order to better explain some of the cooperative behaviour which occur in the everyday life. Quite apart from the problem of indeterminacy, one other theoretical lacuna in their edifice is the failure to recognise the variance of strategic talents among players.

Chapter 5

Finally, contrasting rational choice theories and critical theory I conclude that rational choice theories have a narrow conception of democracy as merely aggregation of preferences or as a mere legal procedure, reducible to something like the counting of votes in majority rule or the rule of law in liberalism. Like the

one-sided conceptions of enlightenment reason their theories of democracy lack moral-cognitive content. By contrast, critical theory sees modernity as not just instrumental reason but opening up possibilities for normative and expressive realms of reason. Democracy, in this view, assumes institutionalisation of learning and discussion within the public sphere. The idea of a public sphere is a central conceptual category in Habermas' theoretical enterprise.

CHAPTER 1

Theoretical Foundations of Rationality: A Public Choice Approach

Introduction

The purpose of the chapter is to elucidate the rational choice approach to politics. Politics is a realm of collective choice. So what constitutes the *rationality* of collective choice in politics forms the subject matter of rational choice theories. Political theorists for a long time have been preoccupied with the problem of collective choice. This problem has drawn many brilliant minds to examine the conditions of cooperative behaviour. For Thomas Hobbes 'disorder' as absence of co-operation is expressed in the state of nature as 'solitary, poor, nasty, brutish and short'. Hobbes' answer to this dilemma in *Leviathan* (1651) was one of the earliest applications of the idea of rational choice to study the conditions of co-operative behaviour. .

Although the idea of rational choice goes back to Hobbes and utilitarians like Jeremy Bentham and John Stuart Mill, it crystallised as a distinct theoretical programme only with the publication of the classic texts: Kenneth Arrow's *Social Choice and Individual Values* (1951), Anthony Downs's *An Economic Theory of Democracy* (1957), Duncan Black's *The Theory of Committees and Elections* (1958), James Buchanan and Gordon Tullock's *The Calculus of Consent* (1962), Mancur Olson's *The Logic of Collective Action* (1965). As Green and Shapiro (1994: 3) argue, what sets contemporary rational choice scholarship apart is the systematic manner in which propositions about the micro foundations of political behaviour are derived. The theory rests on axioms of 'consistency' rather than on the traditional 'psychological' assumptions of utilitarianism. In this respect rational choice theorists derive inspiration from economic models in their efforts to explain

political phenomena appealing to deductive accounts of incentives, constraints, calculations that confront individuals. Consequently they employed mathematical analytical tools with great technical rigour and sophistication in approaching traditional questions of political theory in novel ways.

Throughout this thesis the term "rational choice" is used loosely to include literature like *public choice theory* and *social choice theory* (economic theories of politics; collective action problems; problem of preference aggregation), *game theory* (prisoner's dilemma problems in strategic interaction), *rational choice theory* (includes broadly internal critiques like Elster). Although rational choice emerged as a distinct theoretical paradigm, it is important to note that it is not a single, monolithic theoretical enterprise but constitutes multiple theories each depending on its own rationality assumptions (Hardin 1987: 67).

Public choice theory is defined as the application of economic methods to politics (Mueller 1989; McLean 1987). Starting from the fundamental postulates of methodological individualism, instrumental rationality and the principle of *homo economicus,* the theory focuses on self-interested individuals as they endeavour to maximise their utility through consistent choices. The underlying explanatory assumption is that people are rational and their choices have a broad cost-benefit structure. This approach to the study of politics emerged as one of the dominant paradigms geared to build a unified social science on such micro foundations. This rather simple but powerful idea lies at the root of the influential discipline of economics. In the recent years, the idea that this theory can be fruitfully applied outside the traditional realm of economics has been very widely held, not only by economists with imperialistic ambitions for the expansion of their discipline, but also by political theorists. Public choice theorists take pride in the parsimonious nature of their theory. The project is also seen as a defense of modernity insofar as the theory aims to build a unified social science.

In this chapter, I focus on public choice theories and postpone the discussion of the other theories of rational choice whose critique exposed the limits of instrumental rationality and the principle of *homo economicus.*

My concern in this chapter is to examine the concept of "instrumental rationality" in public choice theories. The concept of "rationality" is conceived as a matter of means, not ends. It is seen as a relation of consistency between preferences, information and action. One other feature is that public choice theorists essentially conceptualise social interaction as "parametric" which means agents take their environment as "given" or "constant". The rational agent assumes his environment to be made up a) of natural objects ruled by causal laws and b) of other agents whose behaviour makes no difference to him. So the agent thinks of himself as a variable and of all other agents as constants.

The theory of public choice claims to have discovered paradoxes in democratic mechanisms. To put it in a nutshell, the basic theoretical discovery is that individual rational choices can sum to collectively irrational outcomes as when everyone in a sports stadium stands up to get a better view. Although its details are constantly questioned, criticised, defended and reformulated, the rational choice paradigm - founded on methodological individualism and the assumption that individuals are motivated by self interest - forms the thread uniting politics and economics.

1.1   Instrumental Rationality and the Principle of *Homo economicus*

1.2   Paradox of Rationality: Conflict of Individual and Collective Rationality

1.3   Public Choice Interpretation of Democracy

1.1 Instrumental Rationality and the Principle of *Homo Economicus*

As said earlier, in terms of basic assumptions, public choice works with a model of instrumentally rational 'self interested' maximisation. *Homo economicus* is a instrumentally rational and calculating seeker of preference satisfaction. The theory is based exclusively on the model of 'preference satisfiers'.

This model of instrumental rationality is famously used in politics by Downs (1957: 6); and further developed by Buchanan and Tullock (1962) and Riker (1982). Rationality is cast in a means-ends framework with the task of selecting the most appropriate means for achieving ends (i.e. preference satisfaction). Thus individuals are deemed rational because they select actions which will best satisfy those preferences. Technically then individuals must have a 'preference ordering' because it is only when preferences are ordered that one will be able to begin to make judgements about how different actions satisfy our preferences in different degrees. So acting instrumentally to satisfy one's preferences becomes the equivalent of utility maximising behaviour. Although this maximising, calculating view of instrumental reason is common in economics, it should not be confused with utilitarianism.[1]

The theory assumes that every individual has a preference ordering. The metaphor of utility maximisation then works in the following way. Suppose a person prefers driving to catching the train and so chooses to drive. Individual preferences are measured in ordinal utilities. Public choice theorists arrived at a conclusion that there is no way one person' satisfaction can be compared with another's. It is meaningless across persons.

Instrumental rationality is defined as having the following axioms (Heap 1992: 6):

Reflexivity: It demands that any bundle is always as good as itself.

Completeness: Any two bundles can always be compared and ranked.

Transitivity: It entails that when A is preferred to B, and B to C, then A should be preferred to C.

Continuity: It implies that given two goods in a bundle, it will always be possible to define another bundle which is indifferent to the first.

In these thin-rational accounts, the content of the utility maximisation is not specified. The claim that public choice theory assumes only instrumental rationality and is agnostic about the type of ends (e.g., selfish or altruistic) that individuals pursue (Buchanan and Tullock 1962: 17) is debatable. Where this claim prefaces an exercise in economic theory, Buchanan and Tullock do go on to assume self-interested ends and abandon their thin-rational agnosticism in favour of a thick-rational *Homo economicus* assumption. Moreover since Buchanan and Tullock (1962) juxtapose the economic approach to collective choice against the assumption that the representative individual seeks to find the 'public interest' or 'common good', it requires Buchanan and Tullock to portray political actors as selfish and profit seekers. And once one grants that the same motivations drive people's economic and political behaviour, then it follows that political agents must be motivated by self interest. This is what is called the symmetry assumption (Brennan and Buchanan 1985: 48-50).

Although 'intentional action' and 'means-ends rationality' can in principle be combined with any type of motivation, for most forms of collective action, rational choice theorists usually add the assumption that rational agents are motivated by self-interest, i.e., they are egoistic, utility maximisers. This assumption is made because self-interest seems to be the most easily universalisable motivation: i.e. it

can explain a larger fraction of collective behaviour than any other single motivation.

The argument for the defense of *homo economicus* model in politics is, though the wicked are fewer in number than the righteous, yet because we cannot distinguish them, there is a necessity of suspecting, heeding, anticipating, subjugating, self-defending, every incident to the most honest and fair conditioned (Brennan and Buchanan 1985: 60). The other argument in defense of the use of *homo economicus* in politics is essentially a methodological one. In short there is a clear methodological requirement of motivational consistency of human behaviour, and it is on this basis that *homo economicus* is imported into politics. Any attack on the use of *homo economicus* in politics is seen equally to be an attack on its use in market analysis: the case for *homo economicus* stands or falls across the board. Public choice theory is methodologically individualistic, in the sense that the basic units are choosing, acting, behaving persons rather than organic units such as parties, provinces or nations (Buchanan 1978: 5).

1. 2. Paradox of Rationality: Conflict of Individual and Collective Rationality

Having defined the structure of instrumental rationality and the principle of *homo economicus*, let me turn to examine the seminal works which applied this model to the study of politics. Basically the thrust of public choice paradigm is to offer a new perspective on politics. It is 'new' in the sense that it revealed the conundrums which hitherto went unnoticed in political theory. The conundrums directly relate to the logistics of the institutional mechanisms in a democratic society. To put it in nutshell, public choice theories deconstructed the myth of collectivity or public interest characterised in the benevolent state theories in political science. In other words, Barry and Hardin (1982) say, 'public choice

theories paint a picture of rational man and irrational society'. This problem is defined as the 'paradox of rationality' - meaning the conflict between individual rationality and social rationality. For a long time political theorists thought that the institutional mechanisms of democracy were representing the "will of the people". This myth was exposed by public choice theorists with baffling theoretical results. The main critique is of "majoritarian democracy". If democracy is defined as the consent of the governed then what constitutes this consent has become the problem for public choice theories.

A common theme that unifies the theoretical works of public choice is that of rationality. It is argued that when the preferences or actions of rational individuals are brought together, the outcome is collectively irrational. Starting from such narrow definition of rationality postulate, public choice theories proceed to show the paradox involved in what is called 'collective choice'. The theory offers very powerful insights in understanding these paradoxes. Before proceeding to discuss the arguments, it is important to clarify one important feature regarding the intellectual history of public choice research programme. If one looks at the trajectory of public choice theory, it did not emerge as a coherent and unified research programme, like its counterpart Frankfurt school. As Abrams argues,

> Public choice theories emerged in disparate parts with divergent theoretical results. In one sense the whole body of work constitutes 'discrete theoretical islands', seemingly disconnected with no meaningful relation to each other. In actuality, the disparate parts or fields are essentially variations of a common underlying conceptual theme (Abrams 1980).

It is in this sense that one may consider it as a unified theoretical programme. The single conceptual theme is to demystify the notion of collective choice. The problems I discuss here are as follows: *the problem of preference aggregation*:

Kenneth Arrow (1957) and William Riker (1982); *the collective action problem*: Anthony Downs (1957) and Mancur Olson (1965); James Buchanan and Gordon Tullock (1962) on *the logical foundations of democracy*. Although the birth of public choice is frequently identified with Duncan Black and Kenneth Arrow, in political science it has been solidified as an intellectual programme only with Anthony Downs and James Buchanan.

The Problem of Preference Aggregation : Kenneth Arrow and William Riker

Although this problem had been known earlier, Arrow called attention to its broader significance to the study of politics.[2] In essence, Arrow proved that one cannot design a social choice function such that the social choice will accurately reflect individual values or preferences. The problem of preference aggregation is: how does one define a "popular will"? It says, if an institution is run on democratic lines it may not be clear how an institution's 'will' or 'interest' should be defined. Consider an assembly, in which each member is a perfect representative of the interests of his constituency and suppose that they have to decide among three proposals. For specificity, think of a municipal assembly that has to choose among building an indoor swimming pool, subsidising the local symphony orchestra or setting up a golf course. Now, if there is one alternative that everybody thinks best, the choice of that option can plausibly be called an expression of the popular will. But in politics, unanimity is the exception.

And since politics is about resolving conflict of interests, it is thought that "majority voting" will elicit or indeed constitute popular will. But this is not the case. Suppose there are three blocks in the assembly, of approximately equal size, representing the business community, industrial workers and health and social service professionals. Let us suppose that they rank the options as follows:

	Businessmen	workers	professionals
Golf course	1	2	3
Orchestra	2	3	1
Swimming pool	3	1	2

Suppose that majority voting is taken as expression of the popular will or the community interest.

1. Then the popular will is that it is better to have a golf course than to subsidise the orchestra, since the businessmen and workers together form a majority with this preference.

2. Similarly, the popular will is that subsidise the orchestra is better than building a swimming pool, since business men and professionals together form a majority with this preference.

3. But the popular will also supports the swimming pool over the golf course, since workers and professionals together have this preference.

So the conclusion is that the notion of popular will based on majority voting results in incoherent. Thus it results in cyclical majority preferences.

Briefly stated, Arrow's theorem is worked out at a high level of abstraction, showing the impossibility of devising satisfactory collective choice mechanisms that do not lapse into dictatorship. This theorem is one of finding a 'decision making rule' which aggregates the preferences of individuals. Arrow's impossibility

theorem rendered all democratic rules of collective decision potentially suspect. In response to Abram Bergson's influential analysis of social welfare functions, which appeared to show that the State could maximise social welfare through objective aggregation of individual preferences, Arrow demonstrated that so long as minimal assumptions of rationality and the complexity of choice are granted, no SWF exists that is neither imposed nor dictatorial. This impossibility result had an enormous impact on how rational choice theorists have thought about the basic characteristics of democracy and 'majority rule'. Thus when three voters (1,2,3) attempt to decide among three alternatives (a,b,c) by majority rule, a voting cycle may result. It says, that even though every individual may have a clear preference-ranking of all alternatives, it is not possible to aggregate these individual rankings into a collective rationality given certain reasonably acceptable conditions.

For a more powerful criticism of the mechanisms of democratic institutions, one of the significant contributions has come from William Riker. He tries to show in his book *Liberalism Against Populism* (1982), the superiority of liberalism over populism. Riker's argument is based on the application of the minimal criteria of all rational choice explanations: utility maximisation and consistency of preferences. Riker exemplifies the rational choice approach to the explanation of political behaviour. For Riker, politics is a process of selecting, producing and enforcing the policies to realise individual preferences. These are assumed to be well-formed, in the sense that preferences are complete and consistent.

Riker (1982: 129-135) argues that Arrow's result displays a theoretical invulnerability in which the reasonableness of the conditions is difficult to question. For Riker the consequence of the situation revealed by Arrow's theorem is either that power is concentrated or that the process of voting may be manipulated. Assuming democracies do not tolerate extreme concentration of power, then their methods of aggregating preferences are subject to manipulation. Riker

distinguishes two methods of manipulation (Riker 1982: 137). One is the false revelation of preferences and the other is the arrangement of the agenda. The former method of manipulation is strategic voting (Riker 1982: 156), wherein voters vote contrary to their tastes in order to bring about an outcome more desirable than the outcome from voting truthfully. All methods of voting are manipulable in this way. Since we can never be certain what "true tastes" are - all we ever know are revealed tastes - we can never be certain when voting is strategic. Yet if strategic voting occurs it is hard to say that the outcome is a fair or true amalgamation of voters' values, especially when the successful manipulators are a small minority. It is possible that strategic voting is commonplace in the real world, as for example vote trading. If so, then, Riker argues that all voting is rendered uninterpretable and meaningless (Riker 1982: 237). Manipulated outcomes are meaningless because they are manipulated, and unmanipulated outcomes are meaningless because they cannot be distinguished from manipulated ones.

The second method of manipulation is by control of agenda (Riker 1982: 169) to change the outcome from what it would be in the absence of such control. Leaders can control agendas by establishing the sequence and content of business. Riker argues that political dispute involves control and manipulation of the agenda and is commonplace in the real world. But it is hard to identify it. Since this manipulation is frequent and difficult to identify, all outcomes of voting are rendered meaningless and uninterpretable. In Riker's view, a dynamic method of controlling agenda is the introduction of a new dimension and issues in order to generate disequilibrium in which the previous kinds of manipulation are possible. Suppose that, by processes no individual person controls, the set of alternatives has been reduced to precisely two and that one of these always wins by a simple majority vote. This seems to be a stable equilibrium; yet it can be upset,

dynamically, by the introduction of new dimensions and issues. If new issues result in situations vulnerable to strategic voting and manipulation of the agenda, then even stable equilibria are dynamically unstable. Riker argues that all social decisions obtained by disrupting equilibria with new issues are not fair and true amalgamation of public opinion, even though the disruption may be socially and morally necessary and desirable.

Riker's (1982: 244) theory of democracy is sceptical about the ability of political parties within a liberal democracy to implement the popular will. The main burden of Riker's sceptical challenge rests upon an increasing awareness of the fragility of majority rule as an instrument for translating individual preferences into a social choice. The inadequacy of majority rule carries implications in Riker's eyes to the extent to which we should allow matters of public policy and collective action to be decided by political parties competing for a winning share of the popular vote.

Thus Riker draws upon the body of results in formal social choice theory highlighting two sorts of problems with voting: "instability" and "ambiguity." Voting is unstable because aggregation mechanisms generate cyclical social orderings. Because of this indeterminacy, electoral outcomes are subject to manipulation via strategic voting (Riker 1982: 115-95, 236-38). And secondly, voting is ambiguous because electoral outcomes are artefacts of the process by which votes are counted. Starting from the same initial profile of individual preferences, Riker argues that because the aggregation mechanism violates criteria of fairness or consistency, no independent external standard exists for discriminating between methods of vote counting. In particular, there is no way to determine which voting method most accurately represents the popular or collective will.

Based on these analytical results Riker concludes that voting is meaningless. He argues that electoral instability and ambiguity subvert "populist" theories that, inspired by Rousseau, see the outcomes produced by democratic practices as corresponding to a coherent popular will. By contrast, it bolsters a liberal theory of democracy in which, echoing Schumpeter, elections perform the solely negative task of disciplining elected officials.

Given the pervasive aspects of majority voting cycles, Riker has become suspicious of political parties in a democracy. In polarising the choice between just two alternatives (in the American case), political parties must necessarily be imposing an arbitrary structure on the pattern of citizen preferences. On this account political parties *manufacture* popular will rather than *reflect* or *represent* it. Riker came to the conclusion that the populist conception of democracy is not effective. Instead one should seek for a set of institutional arrangements that allow electorates to reject those political leaders who are intolerably bad rather than elect those leaders who are going to be strikingly good.

Riker's analysis show how deep and pervasive social choice problems are. In this context there is another interpretation which it is possible to place on the apparent nihilism of the results in formal social choice theory. It would be possible to argue that what these results show is how difficult it is to build a liberal theory of social choice on purely procedural foundations. The formal results that Riker presents and discusses involve little by way of a substantive account of human interests and welfare. From critical theorists' point of view, these results underscore the inadequacy of purely aggregative notions of democracy in ways that can sharpen the case for deliberation (Knight and Johnson 1994: 280).

Thus since the 1950s rational choice models have generated a series of theorems about the logic of majority rule. These findings of the logic of majority rule have prompted new types of reflection on the normative foundations of democracy. One of the significant features underlying Buchanan, Brennan and Tullock's argument is a critique of politics seen as truth-discovery process. In other words, to see democracy as discovering public good is a myth. They expose the myth of government benevolence (Brennan and Buchanan 1985: 33-46). James Buchanan's initial reaction to the discovery of Arrow's impossibility theorem was one of unsurprise. He says that since political outcomes emerge from a process in which many persons participate rather than some mysterious group mind, why should anyone have ever expected social welfare functions to be internally consistent? (Buchanan 1978: 8). Democracy better reflects individuals' preferences than do any of the well-known alternatives such as the military junta, the single party dictatorship etc. Politics being a process of discovery aimed at identifying the public good, the advantage of the democratic mode of search activity is that it seeks to consult the maximum feasible constituency and to reconcile the separate interests of the largest possible number of social actors.

The task is not an easy one. Moreover, if each individual has a meaningful conception of what he conceives to be the public interest then there will be as many social welfare functions as there are individuals in the group.

The preferred direction for institutional reform will require the difficult balancing of the values of independence, self-reliance, and liberty on the one hand, against those of community, fraternity and dependence on the other. Individuals simultaneously want to be free and to belong to a community (Buchanan 1978:

239). It is this duality of desire which makes the study of the whole of democratic enterprise worth undertaking.

Buchanan and Tullock (1962) develop a normative theory of collective choice. The critical normative presupposition on which the whole contractarian construction stands or falls is the location of value exclusively in the individual human being. The individual is the unique unit of consciousness (Brennan and Buchanan 1985: 21). Values have no external existence or objective content independent of each individual's choice calculus. By way of examining the apparent ideological dominance of majority rule, Buchanan and Tullock analysed alternatives to majority rule. This led them to construct an economic theory of political constitutions, out of which emerged *The Calculus of Consent* (1962). They adopt a contractarian approach for a construction of a theory of constitutional choice.

James Buchanan and Gordon Tullock (1962) make use of Hobbesian social contract theory in attempting to provide a device or criterion to change the existing rules or institutions. Accepting much of Hobbes's theoretical framework, especially his grim state of nature and the idea that people might escape from this state by drawing up a social contract, Buchanan and Tullock emphasise the dangers of giving power to any one and try to show that rational individuals would build constitutional constraints on the government into the social contract. The use of Hobbesian contract as a device is to tease out the implications to understand the workings of alternative political institutions so that choices among such institutions can be more fully informed.

Buchanan and Tullock's *The Calculus of Consent* (1962) is a path breaking work in systematically applying economic concepts to the analysis of political institutions. Individual's confrontation with alternatives and his "logic of choice" becomes the central part of analysis (Buchanan and Tullock 1962: vi and vii). This

book is not about the 'ideal' political constitution for society but is an attempt to analyse the calculus of the rational individual when he is faced with questions of constitutional choice. They examine the process extensively only with reference to the problem of decision-making rules (Buchanan and Tullock 1962: vi). The central question is

> if individuals have differing preferences and at the same time some aspects of life are inherently collective or social, rather than purely private, then the problem is... How are differing individual preferences to be reconciled in reaching results that must be shared jointly by all members of the community? (Buchanan 1978: 5).

Buchanan and Tullock attempt to construct a theory of collective choice. Collective choice must be composed of individual actions. Their theory basically begins with the acting or decision making individual as he participates in the process through which group choices are organised (Buchanan and Tullock 1962: 3). As the title of the book itself suggests - *The Calculus of Consent* - it is a calculus of the rational utility maximiser as he confronts what can be called constitutional choice (Buchanan and Tullock 1962: 265).

Buchanan conceded that the social contract is not a valid account of the historical origins of government. Rather his idea is that it can provide a criterion for changes in political institutions (Brennan and Buchanan 1985: 22). Social contract is seen as a hypothetical invention which can be used as a standard for judging, and altering, these institutions. Along the lines of contract tradition, the standard is what would be agreed by rational, self-interested individuals (or utility maximisers). Applying such economic concepts to the polity and treating the polity analogously to the market, Buchanan and Tullock hope to construct a theory of political choice. In other words, the question posed is what are the means or device that would enable one to pass from individually identifiable interests to something

that might be called public interest. The theory is economic only in that it assumes that separate individuals are separate individuals and are likely to have different aims and purposes for the results of collective action. Any theory of collective choice must attempt to explain or to describe the means through which conflicting interests are reconciled (Buchanan and Tullock 1962: 3-4).

In other words, Buchanan and Tullock try to develop an argument for a just constitutional contract. 'Constitution' is a term referring to a set of rules that is agreed upon in advance and within which subsequent action will be conducted (Buchanan and Tullock 1962: vii) and the selection of decision-making rules is itself a group choice (Buchanan and Tullock 1962: 5). So the question is

> what rules do rational individuals agree upon to formulate collective choice? Starting from the premise of rational utility maximisers, the question is posed as when will a society composed of free and rational utility maximisers choose to undertake action collectively rather than privately (Buchanan and Tullock 1962: 43-44).

As Lessnoff puts it, what areas of life would such individuals agree to subject to collective regulation by the state? and on what terms? What constitution would they accept? (Lessnoff 1986: 124). The theory of political choice that Buchanan and Tullock hope to construct is difficult at the outset because of the fundamental interdependence of individual actions in social choice. The important choice that the group must make is: how shall the dividing line between collective action and private action be drawn? What is the realm for social and for private or individual choice? It is not the function of a theory to draw a precise line; theory assumes meaning only in terms of an *analytical* model which describes or explains the processes through which individuals of the group can make this all important decision (deductive logic). In deriving this model we shall be able to describe, in

general terms, some of the characteristic features of a "solution" (Buchanan and Tullock 1962: 5).

Buchanan and Tullock got over the problem of collective choice by looking at how rules for ordinary parlour games are settled before the fall of the cards is known. There are basically two levels of argument in Buchanan and Tullock. One is the "constitutional stage" and the other is the "parliamentary stage".

At the constitutional stage, Buchanan (1989: 59) uses the assumption of a "veil of uncertainty" to help formulate rules for collective action. In considering constitutional questions, the rational individual though self interested is thus not motivated to favour a rule that promote his interests. Since individuals are ignorant of the future positions they occupy once the constitutional rules are drafted, it is rational from each and every individuals' point of view to accept those rules which are in the interests of the collective rather than purely individual interests. So at the constitutional stage, when the fundamental rules of the game are being forged, there is only one decision making rule and it is unanimity of consensus. So central are the rules that there cannot be any rational individual who would wish to absent himself from participation in the process of discussion, bargaining and compromise that culminates in the contract.

> One means to escape this hopeless methodological dilemma is that of introducing some rule for unanimity or full consensus at the ultimate constitutional level of decision making. (Buchanan and Tullock 1962: 6).

At this stage the rule of unanimity is applied. Only those rules are accepted which have the consent of all individuals. Once the rules are formulated at the constitutional stage based on unanimity where the issues of redistribution and property and legal rights are settled, the only decisions left to decide at the parliamentary stage are allocational efficiency issues.

At the parliamentary stage Buchanan and Tullock do not argue for the principle of unanimity but instead develop their costs-benefits analysis of the decision making framework. Since unanimity is costly and time consuming rational utility maximising individuals accept only those decision making rules which minimise their costs and maximise the benefits. Buchanan and Tullock proceed to examine one principle in particular i.e. the majority principle as a decision making rule in a democratic society.

The principle of unanimity is good, but expensive. Since a rational individual is one who has a set of preferences over a set of alternatives, he has a transitive preference ordering. And moreover he views the merits of decision making rules from the viewpoint of cost-benefit analysis. A cost is seen as a loss of utility and a benefit is seen as a gain in utility. From such a simple set of assumptions, Buchanan and Tullock develop their arguments for 'decision-making rules' at a parliamentary stage. Buchanan and Tullock argue that each individual is in the best position to determine when he or she is suffering a loss of utility from the actions of others. The choice of decision making rule represents a trade off between external costs and decision costs. While individuals would prefer a unanimity rule as the best means of protecting against external costs, they will accept less than unanimity rules in order to avoid decision costs. The exact decision rule would depend on the relative importance of the potential external costs and the potential decision costs for the individual. The important point is that the existence of decision costs makes it rational to consider decision rules other than the unanimity rule. Buchanan and Tullock also say that as an ideal the unanimity rule is desirable and that in a world without decision costs, it should be the choice of rational individuals. They argue that all cases of collective action do not suggest the same decision making rule. There is no reason why that should

prevail. Government officials have consistently failed to recognise this fundamental truth. There is no single decision making rule.

Furthermore, Buchanan and Tullock criticised the majority rule for its oppressive exploitation of minorities i.e. its injustice. Pareto optimality is achievable when all individuals are free to exchange goods, services, money or votes in an effort to improve their lot. Under majority rule, those voting for the minority position receive nothing in return for their vote and so the result is not a pareto improvement.

> The theory of constitutional choice has normative implications only insofar as the underlying basis of individual consent is accepted. The costs and benefits from the collective action, as these confront the choosing individual, can be assessed only on the basis of some analysis of the various choice processes. The central part of the book is an analysis of one of the important rules for collective choice - that of simple majority rule (Buchanan and Tullock 1962: 7).

Buchanan and Tullock argue that the reason why the 'self interest' assumption is given prominence is because one cannot be assured that the individual will always follow the moral rules agreed on by the philosophers as being necessary for harmonious social life. The individual may gain unfair advantage over his fellows. They argue against the assumption that individuals assume roles that are institution dependent: for example, that in politics persons take on character roles as statesmen, whereas in the market they take on character roles as possessive profit seekers.

Buchanan and Tullock argue that the operation of their theoretical model does have explanatory value although not in the sense that it is capable of explaining all conceivable configurations that might be observed in the real world political process The important doctrinal implications of their construction lie in its

implicit rationalisation of a political structure that has never seemed to possess rigorous theoretical foundation.

The traditional intellectual defense of majoritarian democracy is made for its superiority in generating an independently existing public good. By contrast, the contractarian paradigm defense of democratic institutions is based on their ability to facilitate the expression of individual values. These institutions are not understood to be discovering the public good. The contractarian construction itself is used retrospectively in a metaphorically legitimising rather than historical sense. The implications of the contractarian vision of politics can be understood through the exchange analogy. So long as the source of value is exclusively located in individuals and there is no differentiation among persons, the whole enterprise of politics can be viewed only as a complex many person system of exchanges or contracts. Individuals must be conceived to join together to explore and ultimately to agree on the establishment of collective entities or arrangements that prove mutually beneficial.

Politics is concerned with resolving conflict of interests. On their view, political order must be antecedent to economic order. Even if a unanimity basis for establishing the legitimacy of the institutions of social order is acknowledged, on a practical level a requirement for unanimity may seem to be mere utopian romanticising. Individuals come to the contractual process, even in its most idealised form, with separated values and with separately identifiable interests.

However Buchanan's contract cast in the *homo economicus* paradigm is prone to numerous problems which I shall discuss in comparison to Habermas' discourse ethics in chapters 3 and 5.

Social Dilemmas - Free-riding and Paradox of Voting: Mancur Olson and Anthony Downs

The problem of collective action is associated with Mancur Olson's account of *The logic of collective action* (1965) and Anthony Downs' *An Economic Theory of Democracy* (1957). The result says individual motives for action may not fit collective preferences for outcomes even when the latter are well-defined. It is essentially an explanatory model involved with the problem of motivation. It says that even when there is an acceptable rule for aggregating preferences, one finds motivational difficulties in implementing the collective preference. This is explained in situations like the free-rider problem and public goods.

It involves the problem of co-operative action among strategically rational individuals for the purpose of providing themselves with 'public goods' or 'collective goods', i.e. goods which, if they are provided at all, must be provided to all members of a community. The problem occurs at the level of enforcement of a collective decision rule in the sense that individuals do not adhere to the agreed upon rules even when the rule is unanimously agreed.

The problem of collective action is also known as the prisoner's dilemma problem (in game theoretic literature), the free rider's problem or the condition of common fate, depending on the context in which it arises. The multiplicity of names is indicative of the failure to generalise the nature of problem. Before examining the problem, it is important to define certain terms relating to this problem. One is the concept of public goods. A good or service is a public good if it is in some degree indivisible and non-excludable. A good is said to exhibit perfect indivisibility or jointness of supply if once produced any given unit can be made available to every member of the public. A good is said to exhibit non-excludability

if it is impossible to prevent individual members of the group from consuming it. By contrast, a perfectly divisible good is one that can be divided between individuals. Once any part of it is consumed by any individual, the amount available for consumption by others is reduced by the whole of that unit. A loaf of bread or a chocolate are examples of perfectly divisible goods. A good which is perfectly divisible is called a private good. Thus, in order to be public, a good must exhibit some degree of indivisibility or jointness.

A good is nonexcludable to the extent that it is difficult or costly to operate a system of charges, denying access to the good to people who do not pay the charge. Collective goods or public goods are goods which can provide benefits for many people simultaneously. Clearly, the range of collective goods is enormous - clean air, safe streets, national security, a just society and so on. If an individual is not excluded from consumption or use of the public good, it is possible for him to be a free rider on the efforts of others, i.e. he can consume or use the public good that is provided by others; whether or not he will in fact be a free rider is something one has to examine.

Employing the self-interest axiom, Olson attacked conventional pluralist accounts of interest groups. It is characteristic of the traditional theory in all its forms that it assumes that participation in voluntary associations is virtually universal. Olson argued by contrast that since groups organise around collective goals, it is often not rational for self-interested individuals to join in such activity; they can better pursue their interests by free-riding on other people's efforts. Hence large groups are often under-organised. This is because people reason that their individual participation will make little or no difference to a group's success and failure, and they will gain the same level of collective benefits whether they participate or not.

The defining characteristic of a collective action problem is very roughly that 'rational egoists are unlikely to succeed in co-operating to promote their common interests'. The argument of collective action is as follows: that if all individuals refrained from doing A, every individual as a member of the community would derive a certain advantage. But now if all individuals less one continue refraining from doing A, the community loss is very slight, whereas the one individual doing A makes a personal gain far greater than the loss that he incurs as member of the community. This is the problem of collective action: If any individual is narrowly self-interested he would presumably not refrain from doing A. Suppose smog in metropolis could be prevented if all individuals would voluntarily pay to have a certain anti-pollution devices installed in their cars. Voluntary action would not solve the smog problem if too many individuals were narrowly rational. In a similar way, much less voluntary activity in many realms such as the green movement can be explained by this logic.

Evidence of this logic at work is seemingly everywhere. So the argument of the logic of collective action is based on the strong assumption that individual actions are motivated by self interest. The appeal of the assumption of narrow rationality is methodological: it is relatively easy to assess in generalisable behaviours.

Olson challenges the view that groups of individuals with common interests usually attempt to further those common interests.

> Groups of individuals with common interests are expected to act on behalf of their common interests as much as single individuals are often expected to act on behalf of their personal interests. This view is widely accepted in the theories which explain state, class, group behaviour in farm lobbies etc. The view that groups act to serve their interests presumably is based upon the assumption that the individuals in groups act out of self interest. So if the members of a group have a common interest then individuals would be rational and self interested to act to achieve that objective. But it is not true that the idea that

groups will act in their self interest follows logically from the premise of rational and self interested behaviour. Unless the number of individuals in a group is small, unless there is coercion or some other special device to make individuals act in their common interest, rational self interested individuals 'will not act to achieve their common or group interests'. In other words, individuals voluntarily will not act to achieve that common or group interest (Olson 1965: 1-2).

Olson explores this inconsistency between group interests and individual interests. These points, Olson (1965: 2) argues, would hold true even when there is unanimous agreement in a group about the common good and the methods of achieving it. The widespread notion in the social sciences is that groups tend to further their interests on the assumption that groups act in their self interests because individuals do.

Olson draws on the tools of economic theory to study individual behaviour in a group, but the conclusions of the study are relevant to political science. The kinds of organisations that are the focus of this study are expected to further the interests of their members: labour unions; farm organisations; cartels; the corporation; and finally even the state is expected to further the common interests of its citizens. The important aspect Olson brought to light is the free rider problem in public goods and collective action problems. The free-rider is one who gets benefits independent of his contribution to the collective good.

Olson's main contention is that the larger the group is, the farther it will fall short of providing an optimal supply of any collective good, and the less likely it will act to obtain even a minimal amount of such a good; In short, the larger the group, the less likely it will further its common interests.

There are three arguments in support of this conclusion to be found in Olson's work. Before examining these issues it is important to clarify some definitions. A group is said to be "privileged" if it pays at least one of its members to provide some public good unilaterally, i.e. to bear the full cost of providing it

alone. Any group which is not privileged is said to be "latent". Where the group is privileged, there is, in Olson's view, a presumption that the public good will be provided; but there should be no such thing in the case of a latent group. However Olson argues that there are groups which are small and capable enough to provide themselves with public goods. He calls such groups intermediate groups. Here in this group an individual's contribution to the provision of a public good will not go unnoticed by other members. In the latent groups - i.e. large groups - it is impossible to notice individual contributions to public good and so Olson argues that there is a need for 'selective incentives' which in turn motivate individuals to pay for public goods. So the individual receives benefit if and only if he contributes and incurs private cost if and only if he fails to contribute. Thus, for example, trade unions which are funded primarily to provide for their members certain public goods such as higher wages and better working conditions have also had to offer prospective members sickness, unemployment and dispute benefits and other positive selective incentives.

So in deciding whether or not to participate, the individual weighs the costs-benefits calculation to see the gains of participation. So there are three arguments which Olson offers in support of his claim that the larger groups are less likely than smaller groups to provide any amount of the public good.

1. The larger the group, the smaller is each individual's net benefit from the public good.

2. The larger the group, the less the likelihood that it will be privileged or intermediate.

3. The larger the group, the greater the organisation costs of providing the public good (costs of communication and bargaining).

It is often taken for granted, at least where *economic* objectives are involved, that groups of individuals with common interests, usually attempt to further those common interests. This view is based upon the assumption that the individuals in groups act out of self interest. If the individuals in a group altruistically disregarded their personal welfare, it would not be very likely that collectively they would seek some selfish common or group objective. Such altruism is considered exceptional, and self interested behaviour is usually thought to be the rule, at least when economic issues are at stake. The idea that groups tend to act in support of their group interests is supposed to follow logically from this widely accepted premise of rational self interested behaviour.

Olson disputes the idea that groups will act in their self interest follows from the premise of rational and self interested behaviour. It does not follow from the fact that all of the individuals in a group would gain if they achieved their group objective, that they would act to achieve that objective, even if they were all rational and self interested.

Olson's study focuses on organisations or groups with significant economic interests. The kinds of organisations that are the focus of this study are expected to further the interests of their members. Although the logic of the theory can be extended to cover communal, ethnic, religious and philanthropic organisations, the theory is particularly useful in studying groups whose interests are primarily economic. Olson also shows the reasons why some organisations or groups are possible. The fact that other large groups have been organised would seem to contradict the theory of latent groups offered in the study. So how does one explain the organisation of some large groups? The argument is as follows:

The large economic groups that are organised do have one common characteristic which distinguishes them from those large economic groups that are not, and which at the same time tends to support the theory of latent groups

offered in this work. The common characteristic which distinguishes all of the large economic groups with significant lobbying organisations is that these groups are also organised for some *other* purpose (Olson 1965: 132). So the large and powerful economic lobbies are in fact the by-products of organisations that obtain their strength and support because they perform some function in addition to lobbying for collective goods.

Similarly, Olson applies his argument to "special interest" groups like business groups which form an important segment of society and has the largest number of lobbies working on its behalf. The power wielded by business in American politics may puzzle the person of democratic predilections: a comparatively small minority exercises enormous power (Olson 1965: 142). The number and power of the lobbying organisations representing American business is indeed surprising in a democracy operating according to majority rule (Olson 1965: 142).

Olson goes on to argue that the power that the various segments of the business community wield in this democratic system, despite the smallness of their numbers, has not been adequately explained (Olson 1965: 142). There have been many vague, and even mystical, generalisations about the power of the business and propertied interests, but these generalisations normally do not explain why business groups have the influence that they have in democracies; they merely assert that they always have such an influence, as though it were self evident that this should be so. Olson poses the question: But why? Why is it natural and necessary, in democracy based on the rule of majority, that the political power should fall into the hands of those who hold the property? (Olson 1965: 143).

The high degree of organisation of business interests, and the power of these business interests, must be due in large part to the fact that the business community is divided into a series of (generally oligopolistic) industries, each of

38

which contains only a fairly small number of firms. Because the number of firms in each industry is often no more than would comprise a privileged group, it follows that these industries will normally be small enough to organise voluntarily to provide themselves with an active lobby - with the political power that naturally and necessarily flows to those that control the business and property of the country (Olson 1965: 143).

There may be a sense in which the narrow "special interests" of the small group tend to triumph over the (often unorganised and inactive) interests of the people. At other times, practical observers may be sensing the fact that the organised and active interest of small groups tend to triumph over the unorganised and unprotected interests of larger groups. Often, a relatively small group or industry will win a tariff, or tax loophole, at the expense of millions of consumers or taxpayers in spite of the ostensible rule of the majority (Olson 1965: 144) .

Thus the large or latent groups have no tendency to voluntarily act to further their common interests. For the unorganised groups, the groups that have no lobbies and exert no pressure, are among the largest groups in the nation, and they have some of the most vital common interests. Migrant farm labourers are a significant group with urgent common interests. The white collar workers are a large group with common interests, but they have no organisation to care for their interests. Even taxpayers and consumers belong to such groups.

Nor can such groups be expected to organise or act simply because the gains from group action would exceed the costs. The existence of large unorganised groups with common interests is therefore quite consistent with the basic argument of this study.

The advocates of group theory like Commons, Bentley, Truman, Latham, and some of the pluralist and corporatist writers have emphasised the pressures of the different economic groups. Many of these writers have taken for granted that

large economic groups working for their economic interests are absolutely fundamental in the political process. They affirmed the existence of groups with something other than self interested economic purposes, but still self interested economic groups always play the largest role in their writings.

Since relatively small groups (privileged and intermediate) will frequently be able voluntarily to organise and act in support of their common interests, and since large groups (latent groups) normally will not be able to do so, the outcome of the political struggle among the various groups in society will not be symmetrical (Olson 1965: 127). Practical politicians and journalists have long understood that small "special interests" groups, the "vested interests", have disproportionate power. The somewhat too colourful and tendentious language with which the men of affairs make this point should not blind the scholar to the important element of truth that it contains. The small oligopolistic industry seeking a tariff or a tax loophole will sometimes attain its objective even if the vast majority of the population loses as a result.

The smaller groups can often defeat the large groups which are normally supposed to prevail in a democracy (Olson 1965: 127). The former triumph because they are generally organised and active while the latter are normally unorganised and inactive.

So Olson's conclusion is that majority rule does not necessarily serve the interests of the majority but serves the interests of the minority.

However not all collective action problems are alike in nature. There is a very important class of problems which arise in connection with the use of resources to which there is open access i.e. resources which nobody is prevented from using. These resources need not be public goods. Garret Hardin's "tragedy of the commons" (1968) concerns resources of this kind. This problem is one where there is 'non excludability' but not 'indivisibility' which means that access to public

good is indivisible. Hardin asks us to imagine a common pasture open to all. So each herdsman is assumed to seek to maximise his own gain. As long as the total number of animals is below carrying capacity of the common, a herdsman can add an animal to his herd without affecting the amount of grazing of any of the animals. But beyond this point, the tragedy of the commons is set in motion. A herdsman adds one more animal thinking that this entails for him a gain and a loss: on the one hand, he obtains the benefit from this animal's yield and on the other hand, the yield of each of his animals reduced because there is now overgrazing. The benefit obtained from the additional animal accrues entirely to the herdsman. The effect of overgrazing is shared by all the herdsmen. The result is that the herdsmen collectively bring about a situation in which each of them derives less benefit from his herd than he did before the carrying capacity of the common was exceeded. So in both the cases above, individual strategic rationality brings out collectively irrational outcomes. This paradox led to the thinking of the discussion on states, private property, community and norms as the alternatives for the effective provision of public goods.

Even the so-called Prisoner's dilemma is closely allied to Mancur Olson's free rider dilemma.

In a similar way, Anthony Downs' *An Economic Theory of Democracy* (1957) grasped the logic of collective action in the context of 'voting' and 'party competition' within a democratic political system. In his model, only a government, parties and voters inhabit the political scene. The government is made up by the party which maximised votes at the last general election. Political parties are assumed to be completely united behind the leader and solely concerned to maximise votes as their only goal is to reap the rewards of holding governmental office. To that end it is rational for parties to be flexible and adopt whichever policies will advance electoral success. In Downs' world it is not rational for parties

to be committed to policies and to be active in persuading voters to give them their support. For their part, voters see elections as a chance to choose a government and because they too are assumed to be rational and self interested individuals they will vote in the way which they consider will offer them the best policy package at the lowest tax cost. The trouble is that informing oneself so as to vote rationally involves costs (both time and money). So it is seen as making sense for voters to attend less to the detail of policy and performance and more to the ideologies for the parties since these give a rough guide to probable policies. Given these assumptions about voters and parties, Downs was able to predict that the two party competition will produce a kind of consensus policies in which parties will tend to develop policy positions near to each other in order to win the marginal votes, although the threat of abstention by the committed party extremists will encourage parties to keep a little away from the centre. This economic theory of democracy clearly parallels the neo-classical picture of the economy. Voters are political consumers using votes as politicians are seen as supplying policies to win votes and secure the fruits of government office.

The other problem which Downs analyses is the aspect of voting in democracy. Using economic assumptions and cost-benefit analysis, Downs came to the conclusion that voting costs time and money and so most people would rationally calculate not to vote. The difficulty involved with co-operative action to provide collective goods revolves around the results of a rational actor's calculations as to whether the costs of participating in such action outweigh the benefits. So in regard to voting, it appears that a rational individual would decide not to vote, for the simple reason that the cost of time and effort of that act is far too high when measured against the benefit of having his favoured party win. He thinks the likelihood that his one vote will make the difference between his party winning or losing. By not voting the individual does not measurably change the

collective outcome of the results. If his party loses, it would have done so even if he had voted for it. By contrast, if it wins, he gets the benefits of the collective good without any cost. So here the individual is a free rider. But the conclusion that it is not rational to vote has disturbing consequences for both political philosophers and rational choice theorists. So the problem which arises regarding voting arises with all types of collective action situations.

The most significant contribution of Downs' analysis is the "rational ignorance hypothesis": it holds that the information-seeker continues to invest resources in procuring data until the marginal return from information equals the marginal cost (Downs 1957, 215).

Downs (1957: 238) argues that citizens acquire political information for two main reasons: 1) to help them decide how to vote, and 2) to form opinions with which they can influence government policy formation during the period between elections. In Downs' model voting is a means of selecting the best possible government from among the parties competing for the job. Therefore a rational man votes because he would rather have one of these parties than any of the others. The margin of his preference is his party-differential.

Since information involves costs, Downs argues that a rational man may invest in information for three reasons: 1. he wishes to influence the government's policies, 2. his prediction of how other voters will act indicates that the probability is relatively high that his own vote will be decisive or 3. he derives entertainment value or social prestige from such data. So it might seem that political information is useful to voters because it enables them to have specific preferences which in turn influence government policies that affect them. But Downs comes to a different conclusion.

Downs argues that information is relatively useless to those citizens who care which party wins and those citizens for whom information is most useful do

not care who wins. In short, then, nobody has a incentive to acquire political information. In so far as voting is concerned, rational behaviour implies both a refusal to expend resources on political information *per se* and a definite limitation of the amount of free political information absorbed.

Every rational voter realises that he is not the only person voting. This knowledge radically alters his view of the importance of his own vote. The party which eventually wins will probably be elected no matter how one casts his ballot, as long as the other citizens vote independently of him (Downs 1957: 244). Because nearly every citizen realises his vote is not decisive in each election, the incentive of most citizens to acquire information before voting is small (Downs 1957: 299). A large percentage of citizens - including voters - do not become informed to any significant degree on the issues involved in elections, even if they believe the outcomes to be important (Downs 1957: 299).

Contrasting ethical models of democracy and his model, Downs argues that they differ in two ways. The contrast arises from the simultaneous truth of two seemingly contradictory propositions (Downs 1957: 246): 1) Rational citizens want democracy to work well so as to gain its benefits, and it works best when the citizenry is well informed, but 2) it is also individually irrational to be well informed. Here individual rationality conflicts apparently with social rationality; i.e., the goals men seek as individuals contradict those they seek in coalition as members of society. This paradox exists because the benefits men derive from efficient social organisation are indivisible. Downs (1957: 246) argues that every citizen benefits in the long run if government is truly run "by consent of the governed"; i.e., if every voter expresses his true views in voting. By his true views, Downs means, the views he would have if he thought that his vote decided the outcome.

But in fact his vote is not decisive: it is lost in a sea of other votes. Hence whether he himself is well informed has no perceptible impact on the benefits he gets. If all others express their true views, he gets the benefits of a well-informed electorate no matter how well-informed his is; if they are badly informed, he cannot produce these benefits himself. Therefore as in all cases of indivisible benefits, the individual is motivated to shirk his share of the costs: he refuses to get enough information to discover his true views. Since all men do this, the election does not reflect the true consent of all governed (Downs 1957: 246). In a democracy, government cannot force people to become well-informed. This view of rationality conflicts sharply with the traditional idea of good citizenship in a democracy. Indeed, the whole concept of representative government becomes rather empty if the electorate has no opinions to be represented (Downs 1957: 245). The tension which exists between individual and social rationality prevents the governed from expressing their true consent when they select a government.

Downs argues that there are significant differences between acquiring information in order to "vote" and acquiring it in order to "influence policy-making." In the first place, voters automatically communicate their decision to government in the act of voting, but influencers must transmit their opinions to government by specific act in order to get results. Although this also involves cost, the cost or benefit depends on the position of the citizen of society. If the citizen's interest in a policy area stems from his business, he can charge the costs of transmitting his views to his firm, which will probably deduct them from its taxable income.

There is a distinction between two types of return from information: almost everyone at least considers voting, but relatively few citizens ever consider exerting influence in any particular area of policy. A voter's party differential is subject to heavy discounting because of the great number of other voters. In

contrast, an influencer's intervention value may suffer hardly any discount because only a small number of others are interested in the policy he wants to influence (Downs 1957: 253). Perhaps many people are affected by this policy, but since most of them do not realise in advance the source of these effects, they cannot seek to alter policy pursued at that source. Such ignorance is not the result of mere apathy; rather it stems from the great cost of obtaining enough information to exert effective influence. Each influencer must be acquainted with the situation at least well enough to be in favour of a specific policy. Many people voice strong policy preferences without benefit of much information, and the ballots these people cast are just as potent as those of the well-informed. Nevertheless, the government knows that its behaviour in a given policy area will affect many people who show no immediate interest in that area. Consequently, it must be persuaded that these presently passive citizens will not react against whatever policy an influencer is promoting. A would-be influencer has to be knowledgeable enough to carry out this persuasion.

Thus formulation of policy requires more knowledge than choosing among alternatives which others have formulated. As a result, influencers need more information about the policy areas they operate in than even the most well informed voters; hence their data costs are higher. The complexity of these areas often forces influencers to become experts before they can discover what policies best suit their own interests. And because many influencers with different goals are competing with each other for power, each must 1. produce arguments to counter any attacks upon him, 2. assault the others' contentions with data of his own 3. be informed enough to know what compromises are satisfactory to him.

In contrast, a voter need find only the differential impact on him of a few alternatives formulated by others. He does not have to examine all the possible alternatives, since not all are open to his choice - though all are open to the choice

of a policy maker. Also a voter need not be well informed enough to think of compromises since either one party or the other will win.

The gist of the analysis is that influencers are specialists in whatever policy areas they wish to influence; whereas voters are generalisers trying to draw an overall comparison between parties. Specialisation demands expert knowledge and information and most people cannot afford to become expert in many fields simultaneously. Therefore influencers usually operate in only one or two policy areas at once. This means that in each area, only a small number of specialists are trying to influence the government. Thus the potential returns from influence are high enough to justify a large investment of information. In almost every policy area, those who stand the most to gain are the men who earn their incomes there.

So for all these reasons, producers are more likely to become influencers than consumers. The former can better afford to invest in the specialised information needed for influencing and to pay the cost of communicating their views to the government. However every man is both a producer and consumer at different moments of life. So men are more likely to exert political influence in their roles as income earners than in their roles as income spenders, whether acting as private citizens or as members of a corporate entity (Downs 1957: 254-55).

The cost of acquiring information and communicating opinions to government determines the structure of political influence. Only those who can afford to bear this cost are in a position to be influential.

A striking example is the failure of consumers at large to exercise any cogent influence over government decisions affecting them. For instance legislators are notorious for writing tariffs laws which favour a few producers in each field at the expense of thousands of consumers. On the basis of votes alone, this practice is hardly compatible with Downs' central hypothesis about government behaviour. But once we introduce the cost of information, the explanation springs full armed

from Downs' theory. Each producer can afford to bring great influence to bear upon that section of the tariff law affecting his product. Conversely, few consumers can bring any influence to bear upon any parts of the law, since each consumer cannot even afford to find out whether tariffs are raising the price they pay for any given product.

Under these conditions, government is bound to be more attentive to producers than to consumers when it creates policy. In general, economic decisions of a rational government in a democracy are biased against consumers and in favour of producers (Downs 1957: 239). Economically speaking, governmental policy in a democracy almost always exhibits an anti-consumer, pro-producer bias. And this bias exists not because the various agents concerned are irrational, but because they behave rationally (Downs 1957: 256).

"Division of labour" and "uncertainty" inevitably cause men to be informed to different degrees. The inequality of information always results in a corresponding inequality of influence over governmental policy formation. This conclusion emphasises once more the inherent inequality of political power in democratic societies (Downs 1957: 257).

Democracy is well defined as *government by consent of the governed.* To go a step further, government by consent of the governed is defined as decision making in which the decider makes each choice on the basis of the preferences of those affected by it and weights the preferences of each in proportion to the degree to which he is affected. Though this complex definition is still ambiguous, it is clear enough to compare with the method of weighting preferences used by the rational government in our model.

The comparison shows that the *cost of information* prevents Downs's model government from ever functioning by *consent of the governed* in a pure sense. This does not mean that government makes decisions without considering

the desires of the people affected by them; on the contrary, it is extremely sensitive to the wishes of the electorate. Nevertheless, because of the very structure of society, each government decision cannot result from equal consideration of the wishes of men who are equally affected by the decision. To this inherent disparity of influence, if the inequalities of power caused by the uneven distribution of income are considered, then we have moved a long way from political equality among citizens (Downs 1957: 257).

Downs (1957: 258) argues that this conclusion by no means implies that democracy cannot work, or that it is without benefits, or that it embodies only sham equality. A contrast between Downs's model and the communist model would probably show that democracy is relatively successful at achieving political equality. Nevertheless his model does tend to verify the following assertion: even if society's rules are specifically designed to distribute political power equally, such equality will never result in an uncertain world as long as men act rationally. In short, perfect political equality is irrational when uncertainty exists, unless there is no division of labour.

In an appendix (Downs 1957: 274) titled: *the possible existence of irrationality in the model*, Downs argues that every citizen who wishes to vote but is indifferent about who wins chooses a party at random and votes for it. From the point of view of the individual, there is no reason why random selection is not preferable to certain other methods of choice. Since he cannot distinguish between the parties on the basis of their policies, he might as well use any other basis which pleases him. For instance, he might vote for the party whose leader has the most charming personality, or the one his father voted for. Thus a rational man may employ politically irrational mechanisms to decide for whom to vote. The difficulty with these arguments is that they rationalise everything. If it is rational to vote for prestige, then why is it not rational to vote so as to please one's employer or one's

sweetheart? Soon all behaviour whatsoever becomes rational because every act is a means to some end the actor values. Though the use of such devices is individually rational, it is socially irrational (Downs 1957: 274).

1. 3. Public Choice and Democracy

In this section, I shall discuss the implications of these results for democracy. To sum up, the theoretical conclusions of public choice theories of democratic politics have shown the paradoxes of individual and collective choice. I have roughly discussed three problems: One, following Arrow and Riker, is that the notion of *social preferences* is in general not well-defined: the problem of preference aggregation. Second, Buchanan and Tullock's construction of a normative theory of collective choice by way of a critique of the "majority principle." Finally, Olson and Downs argue that social action is likely to be distorted by the private interests of the agents: the collective action problem with reference to public goods.

What unites all these works of public choice theory is a powerful critique of democratic institutions, describing in great detail the nefarious effects of economic individuals pursuing advantage through political mechanisms and thereby exposing the myth of "general will." Hence the normative problem for public choice theorists has been to open up new reflections on the normative foundations of democracy. But it is important to remember that normative responses are varied. Buchanan, Tullock and Riker call for a limited government. Olson argues for "selective incentives" to individuals in groups as a way of overcoming the free rider problem. Of all the positions, Buchanan and Tullock provide the most concrete normative foundation of a theory of collective choice

Democracy is both an ideal and a method. As a method, it deals with the procedural characteristics of the ballot box and all that goes with it. And the ideal is justice and individual self-realisation involving essentially free human persons. Populist interpretations of democracy see it as a just political device for representing the "general will." Public choice theorists, by contrast, contend that the failure to apply the "self-interest" axiom to politics was because of the dominance of Rousseau's ideas fostering the view that governments in a democracy have no real existence apart from the "will of the people." The result is an unconscious assumption that all governments are run by altruists: this criticism of democratic consent as "general will" operating on the principle of majority rule lies at the heart of all public choice theorists (Buchanan and Tullock 1962; Downs 1957: 279-294; Riker 1982; Olson 1971).

Thus public choice theories have come up with two findings about democracy: one is the conceptual incoherence of majoritarian democracy and the second is the collective action problem. A common objection to the method of majority decision is that it is illiberal. If apparent majorities are often chimerical, if minorities can manipulate democratic decision rules to generate the results they desire, and if there is no way to amalgamate individual desires into a general will, as Rousseau had claimed in *The Social Contract* (1762), then the nature and desirability of democracy require re-evaluation.

The theory of public choice is a theory about the way the preferences of individual persons are amalgamated into a collective choice of society. The crucial attribute of democracy is "popular participation" in government. Although the institutions of participation have been many and varied, they have always revolved around the simple act of "voting." - As Riker (1982: 5) says, voting is not equivalent to democracy, only voting that facilitates popular choice is democratic. This condition excludes voting both in oligarchic bodies and in plebiscites in

communist and military tyrannies, where voting is no more than forced approbation. Thus one can say that democracy implies voting but voting does not imply democracy. So voting is a necessary, but not sufficient condition of democracy. To render them equivalent, voting must be surrounded with numerous institutions like political parties and free speech which organise voting into a genuine choice. Thus public choice theorists see voting as a central act of democracy.

Because "voting" is one method of aggregating values, public choice theorists have come up with a theory of voting. And although participation can take many forms, it invariably includes "voting." Therefore, the theory of choice is highly relevant to the theory of democracy. Voting as a fair democratic procedural ideal assumed importance in the twentieth century when it became more frequently used as a method of social decision. Public choice theory is an analytic theory about the main institution of democracy - namely voting (Riker 1982: 4).

Contrasting liberal (or Madisonian) and populist (Rousseaunian) interpretations of "voting" Riker (1982: 239) argues populism as a moral imperative depends on the existence of popular will discovered by voting. But if voting does not discover or reveal a will, then the moral imperative evaporates because there is nothing to be commanded. By contrast liberals interpret voting as a democratic device as inadequate and fear the tyranny of majority. Using the results of social choice theory, Riker tries to show the superiority of liberalism over populism. Riker uses these results to draw a conservative normative conclusion: that democracy cannot possibly live up to many of the claims made for it, and that only what he calls liberalism (democracy restricted by checks and balances) is supportable. So the only alternative is liberalism, which is a form of democracy with entrenched rules to prevent the tyranny of the majority.

The normative response of Buchanan and Tullock is to model collective choice on the basis of the contractarian paradigm. Contrasting rationalist democracy and individualist democracy Buchanan and Tullock argue that:

> Political theorists do not seem to have considered fully the implications of individual differences for a theory of political decisions. The choice making process has been conceived of as the means of arriving at some version of "truth"; some rationalist absolute which remains to be discovered through reason or revelation and which, once discovered will attract all men to its support. The conceptions of "rationalist democracy" have been based on the assumption that individual conflicts of interest, will and should, vanish once the electorate becomes fully informed (Buchanan and Tullock 1962: 4).

The critical potential of public choice theories is seen in its critique of such concepts as *general will* which are shown to be deeply restrictive of liberty and supposedly impossible without the "sword of leviathan", to use Buchanan's phrase. The theory has been used to uncover the hidden operation of force and coercion in modern centralised political institutions and to point out the illusory character of many communitarian political ideals tied to the radical democratic tradition.

Although the theoretical conclusions of public choice expose the myth of collective action, it does not mean that they do away with any notion of collective action. Public choice theorists oppose collective action if it means discovering some objective notion of public interest other than what individuals desire. On this account the goal of politics is instrumental, in the sense that politics has only to do with rules and not with any correct measures of individual interests. Buchanan (1978: 9) says the task of politics is to devise institutions that would enable us to pass from individually identifiable self-interest to something that might take the place of public interest.

Collective rationality is seen as a question of co-ordinating or aggregating individual preferences. Buchanan and Tullock (1962) make "unanimity" a condition

for political decisions. They argue for unanimity on the libertarian grounds that anything short of it will violate the rights of the minority. Democracy is interpreted as a device for preference aggregation under conditions of non-cognitivism in individual preferences, instrumentally rational behaviour and self-interest. Non-cognitivism means simply that values and preferences are beyond the reach of rational argument.

As Buchanan (1986: 41-54) argues politics is a process in which conflicts among individual interests are settled. In this enterprise there is no independently existing interest analogous to "truth" towards which the interaction process converges. Rules become the reference point against which the justice of individuals' behaviour can be assessed. A fair rule is one that is agreed to by players in advance of play itself. So a rule is fair if players agree to it. Agreement in its purest form is unanimous agreement.

Brennan and Buchanan (1985: 43) defend democratic institutions such as "majority rule" (because "unanimity" is costly and time consuming) on economic grounds. By contrast, Buchanan and Brennan (1985: 43) argue that proponents of majoritarian democracy defend it because of its alleged efficacy in discovering the "public good." Thus, on this view, democratic institutions stand or fall on their alleged superiority in generating the attainment of an independently existing public good. Moreover, proponents of democracy in the noncontractarian paradigm acknowledge that majority rule is less imperfect than its alternatives in part because the public good is not readily discernible. And, indeed, these very proponents may argue that non-democratic methods of reaching collective decisions are more efficient when the public good is more clearly outlined, as, for example, in war or emergency.

By contrast the contractarian defense of democratic institutions, as constitutionally derived, is based on their ability to facilitate the expression of

individual values. These institutions are not understood to be discovering the public good, even if at some deeper philosophical level the existence or non-existence of such a public good may be a subject of further inquiry. The argument is that since there is no *external* standard against which institutions can be evaluated in efficiency terms, any evaluative criteria must be applied directly to the institutional processes themselves. In the contractarian perspective, constitutional democracy, which may embody majoritarian voting rules in specified circumstances, is categorically distinct from all other political forms of authority.

For the noncontractarian, majoritarian democracy is aimed at finding or discovering the public good. Thus politics is seen as the search for the grail of the public good and the activity of discovery is carried on under the continuing presumption that what is sought exists independently of the expression of individual values in the search itself. Hence any constitutional constraints on the "will of majority" will tend to be opposed.

Although public choice theorists (like Buchanan and Riker) are obsessed with overcoming the paradoxes of voting, they fail to recognise that voting does not exhaust democratic practice. This is mainly because they share the view that the political process is instrumental rather than a public action, viz. the individual and secret vote. With these usually goes the idea that the goal of politics is the optimal compromise between given and irreducible interests (Elster 1986b: 103). Critics claim that public choice theories embody a confusion between "market" and "forum" (Elster 1986b). The average rational citizen is modelled on the lines of consumer, and they fail to recognise the distinction between the domain of the market and the domain of the forum. In the former, the consumer chooses between courses of action that differ only in the way they affect him. But in political choice situations, the citizen's choice inevitably affects other people who might differ.

One can question the adequacy of the model of preferences that is built into

the approach when we are considering matters of political choice. In these models, judgements about the relative merits of alternative social states are represented in the same way in which tastes are represented in the choice of commodities by consumers. Although this principle can be applied in markets, the same cannot be true of political judgements (Weale 1992: 215). In any political argument people offer reasons for favouring their preferred alternative. So the logic of advancing a political preference must be understood in the light of these underlying reasons, and not simply as claims to have one's preferences satisfied whatever they happen to be.

The conception of preferences based on instrumental reason presented in public choice theories has come under attack from both internal and external critiques. In the next chapter, I shall turn to the external critique articulated by Habermas in his theory of communicative action who distinguishes instrumental reason and communicative reason. This alternative model sees collective choice not as a process of preference aggregation but as a process of dialogue in which reasons are exchanged between participants in a process that is perceived to be a joint search for a consensus.

Notes

[1] The key difference is that Bentham's social philosophy envisioned a universal currency of happiness for all people. Everything in people's lives either adds to the sum total of utility in society (i.e. it is pleasurable) to subtracts from it (i.e. is painful) and the good society is the one that maximises those utilities. This view is plainly controversial now because it presumes as we can compare one person's utility with another's.

[2] The foundations of the theory of choice were laid down in the eighteenth century. In 1785, the French mathematician Marie de Caritat better known as the Marquis de Condorcet, discovered the interesting "paradox of voting". This paradox was rediscovered in the nineteenth century by Charles Dodgson, better known as Louis Carol, and in the early 1950s Kenneth Arrow used it as part of his "impossibility theorem".

Instrumental Rationality and Communicative Rationality: Habermas on Social Interaction

Introduction

In the previous chapter, I examined rational choice theory's defense of the concept of instrumental reason in modelling social interaction. It can be said that the origins of rational choice theory belong firmly in the project of modernity. In social sciences, the completion of the project of modernity is aimed at developing a rational theory of society. By contrast, critiques of modernity argue for the abandonment of the project of modernity. Although the critique of instrumental reason was at the heart of early Frankfurt school thinkers like Max Horkheimer, Theodore Adorno and Herbert Marcuse, they failed to provide critical theory with any alternative normative standpoint. Modernity is seen as constituting only one form of reason and that is instrumental reason. As Benhabib says,

> The Frankfurt school's critique of instrumental reason was an aporetic project precisely because once the *identification* of emancipation with the increased technical mastery of nature was rejected, there seemed no other instance of human rationality to appeal to besides aesthetic reason (Benhabib 1986: 224).

Jurgen Habermas maintains that the Frankfurt school's critique of instrumental reason need not appeal to a utopian reconciliation with nature; the true negation of instrumental reason is not utopian but communicative reason. For Habermas the earlier versions of critical theory were inadequate because:

> In the first place, critical theory never really took the theoretical contributions of the social sciences and analytical philosophy seriously.

It never engaged with them systematically, as it should have done, given its own intentions. Hence, secondly, it took refuge in an abstract critique of instrumental reason and made only a limited contribution to the empirical analysis of the over-complex reality of our society. And finally, it failed to give an unambiguous account of its own normative foundations, its own status (Dews 1986: 49).

What makes Habermas' project unique and challenging is his emphasis on an enlightened suspicion of enlightenment, a reasoned critique of western rationalism. The point is that reason can only be defended by way of a critique of reason. McCarthy (1984: vi) aptly calls him the "last great rationalist". The central preoccupation of Habermas in his two volumes of *The Theory of Communicative Action* (Habermas 1984: Vol. I; 1987: Vol. II) is to give an exposition of the normative foundations of critical theory by way of a critique of subject-centered reason. From the standpoint of critical theory, rational choice theory conceives modernity merely in terms of instrumental reason and defends this form of reason. On the other hand, Habermas gives a critical exposition of modernity as multidimensional and redefines critical theory in terms of a theory of communicative rationality and ethics. Communicative action is not a metatheory but the beginning of social theory concerned to validate its own critical standards (Habermas 1984: xxxix).

In this chapter, I shall examine Habermas' critical theory of modernity and discuss his theory of communicative reason as an alternative way of conceptualising social interaction.

2.1   Communicative Rationality and Lifeworld

2.2   Habermas' Transformation of Epistemology: Reconstructive Sciences and Social Evolution

2.3   Habermas on Social Interaction

Strategic rationality and Communicative rationality

2.1 Communicative Rationality and Lifeworld

Habermas's central aim in *The Theory of Communicative Action Vol. I* (1984) *and Vol. II* (1987) is to sketch a critical theory of modernity that suggests a redirection rather than an abandonment of the project of the enlightenment. It is by way of examining Weber's theory of the rationalisation process that Habermas elaborates his theory of communicative action. He takes Weber's analysis of rationalisation as a point of entry to discuss modernity.

It is important to note the differences between Weber's and Habermas' characterisations of the cultural legacy of modernity. Habermas uses the notion of communicative reason as the key to a systematic reconstruction of the development of modern structures of consciousness and he thinks this development can be understood as a gain in rationality for the human species. Habermas' point of entry to address this complex question is Weber's theory of rationalisation. He then modifies Weber in accordance with the communicative model. Habermas (1984: 140) characterises his approach to Weber as a "flexible exploration and deliberate exploitation". Undoubtedly specialists will find much to contest in his treatment of Weber. For present purposes, I am not concerned with such problems, but only with the role it plays in communicative rationality.

The starting point of Habermas is that Weber employed an unclarified concept of rationality in analysing the process of disenchantment in the history of religion which is said to have fulfilled the necessary internal conditions for the appearance of occidental rationalism. In his analysis of societal rationalisation as it

makes its way in the modern period, Weber allows himself to be guided by the restricted idea of purposive rationality.

> Because Weber construed rationalisation in terms of the increasing dominance of purposive rationality, so he did not adequately grasp the selectivity of capitalist rationalisation or its causes. Following him in essential respects, Horkheimer and Adorno were led to deny any trace of reason in the structures and institutions of modern life. In Habermas' view this diagnosis misinterprets the very real distortions of modernity and underestimates its equally real accomplishments and as yet unrealised potentials (McCarthy 1984: xxxvi).

> The discontents of modernity are not rooted in rationalisation as such, but in the failure to develop and institutionalise all the different dimensions of reason in a balanced way (McCarthy 1984: xxxvii).

Elucidating the concept of rationalisation Weber traces that long process of development from the breakdown of magical-mythical ways of seeing the world to the emergence of the Protestant ethic which allowed purposive rationality to be motivationally and institutionally anchored in such a way that capitalist rationalisation could finally take off. It is no simple matter to link all of Weber's statements together into a cohesive, multi-dimensional account of the different senses and degrees of rationalisation which are involved in that process.

Habermas does not think that such an account can be given unless significant changes are made. Habermas' contention is that the root problem is that Weber's theoretical framework is too restricted to grasp adequately the range of phenomena he hoped to take account of. The task Habermas sets for himself is to rethink Weber's theory, using the communicative model's resources to overcome its difficulties. The ultimate purpose of this undertaking is not simply the interpretive one of generating greater coherence; it is rather the provision of a richer account of what Weber saw as the costs of modernisation or rationalisation: the loss of

freedom in an increasingly bureaucratised society and the loss of meaning or unity in a fully disenchanted world.

From Habermas' (1984: 186-215) perspective, Weber's analysis of rationalisation is interesting on three different levels: worldviews, societal rationalisation, differentiation of various cultural value spheres. In his sociological study of religion, Weber analysed that process of disenchanment, whereby the magical-mythical view of the world broke down under the influence of what he called "world religions" (Christianity, Judaism, Hinduism, Buddhism, Islam). The universal hallmark of such religious-metaphysical worldviews is that they represent the cosmos as a coherent, meaningful whole, within which an explanation for 'suffering' is given as well as directions for the sort of life conduct which is necessary to earn salvation from that suffering. Weber was particularly interested to show how this universal rationalisation process happened, in the west, to give rise also to a process of societal rationalisation, that is, a rapid increase in the degree to which areas of social life, economy and administration, were organised according to criteria of formal, purposive rationality. For Weber the answer was the Protestant ethic.

Weber connected rationalisation at the level of worldviews with, on the one hand, the societal form of rationalisation and, on the other, the rationalisation of worldviews which also results in a differentiation of various value spheres. Habermas (1984: 143, 279-286) argues that Weber's analysis of modernity as rationalisation process was restricted on two counts:

In the first place, in his analysis of the emergence of the modern world view, primarily in the light of the evolution of world religions, Weber focused on the dimension of the ethical transformation of such world religions and overlooked developments in the scientific cognitive and the aesthetic-practical spheres.

Habermas (1984: 233-34) strongly emphasises that the decentration of the worldview in modernity results not primarily in the dualistic ontology which juxtaposes a normatively regulated social life - "a sphere of values" - to objectivity - "a sphere of facts" - but in the differentiation of three realms. These are "world", "society" and "self". These categories are not ontological but referential ones: they designate an objective domain of facts, a normatively regulated intersubjective sphere and a subjective domain of the self's inner experiences. Such a decentered worldview allows the acting and knowing subject to adopt different basic attitudes toward the same world. These differentiated spheres become accessible for cognition and action in different modes: their separation from one another also seems to render fluid the attitudes with which the knowing and acting self can relate to these worlds attitudes which are named by Habermas (1984: 237) the "objectifying", the "norm-conformative", and the "expressive". Only when these three basic attitudes toward the three worlds become cognitive possibilities are we presented with the full range of the rationalisation potential contained in the demise of world religions.

In the second place, this cognitive potential becomes anchored in social life in the institutionalisation of cognitive, normative, and aesthetic ideas in cultural action systems. The differentiation of science and technology, law and ethics, and aesthetics in modernity is accompanied by their institutionalisation in universities and research academies, legal and juridical professions, and an autonomous realm of art production and criticism. Through such institutionalisation, modes of appropriating cognitive, social, and subjective values become reflexive, i.e., increasingly subject to discourse validation and argumentation. As argumentative modes of generating beliefs and procedures for testing them are institutionalised, the rationality of such processes is seen to lie increasingly in their procedural

characteristics rather than in their substantive content. The rationality of belief systems is attributed, first and foremost, to the procedures through which they are acquired and through which they can be revised or refuted.

However, in Weber's mind, societal rationalisation had only one possible course to take: the spread of spheres of societal rationalisation was identified with purposive rationalisation. But, for Habermas, such an identification is not necessary. Habermas argues one can open up the question of whether purposive rationalisation is the only possible way of developing that broader potential for the rationalisation of action which is made available with the culture of modernity. Habermas argues that western modernisation constitutes only a one-sided utilisation of the rationality potential of modernity. The important implication that would follow is that one could see the dilemma of loss of meaning and loss of freedom against the background of counterfactually projected possibilities for organising social action differently.

The idea is that the modern social world is composed of a number of distinct provinces of activity, each having its own dignity and its own immanent norms. Instead of conflating 'enlightenment reason' with instrumental reason, Habermas sees in modernity increasing potential for more congenial kinds of social and political interaction organised around "talk" rather than strategy. This contrast between instrumental reason and communicative reason is a constant theme in Habermas' work.

In order to open up the conceptual space for such inquiry, Habermas introduces the concept of rationalised lifeworld to complement his theory of communicative action.Habermas' specific interest to flesh out the notion of rationalisation of worldviews requires that one possess a clear theoretical framework, in terms of which one systematically generates insights. Weber does not seem to have had one. Disenchanment signifies a breakdown of a sociocentric

consciousness of a seamless magical-mythical world and the construction of a decentered consciousness which recognises clear demarcations between the natural, social, and subjective worlds. And again the demarcation of formal world concepts also means increasing recognition of the differentiated system of validity claims corresponding to the three worlds. This change can be described as a process of rationalisation because it enhances the learning capacity of mankind. It does this because it provides actors with the conceptual means of constructing a reflexive or self critical perspective.

Habermas stresses that the destruction of the unity of reason does not mean a general loss of rationality. Even if, as regards the substantive content of ideologies and worldviews, the loss of meaning is irreversible, the unity of reason can be preserved via those communicative processes through which validity claims can be redeemed. The new paradigm of rationality is not substantive but *'discursive'*; it thematises not the content but the form of those necessary argumentation procedures in light of which the validity claims of "truth", "rightness", and "truthfulness" are settled.

Habermas (1984: 234-5) connects each of the three cultural value spheres (cognitive, normative, and aesthetic) with corresponding interest positions, or "orders of life", that regulate the possession of "ideal goods": science/knowledge, religion/morality, and art/taste and "material goods": economy/wealth, politics/power, counterculture/love. These value spheres are logically irreducible to one another. So construed, none of the three rationality complexes is capable of exerting a structure-building influence on society as a whole; as a consequence, value-rational orientations are entirely uprooted from their religious base and are replaced by instrumental and hedonistic modes of conduct. The resulting loss of freedom is in part symptomatic of the disappearance of moral autonomy. What emerges is a model of selective rationalisation in which society, culture, and

personality are overburdened with integration problems while at the same time being threatened with absorption into the combined economic-bureaucratic apparatus.

> "...the tensions among the ever more rationalised spheres of life go back in fact to an incompatibility of abstract standards of value and aspects of validity, or rather to a partial and therefore imbalanced rationalisation - for example, to the fact that the capitalist economy and modern administration expand at the expense of other domains of life that are structurally disposed to moral-practical and expressive forms of rationality and squeeze them into forms of economic or administrative rationality" (Habermas 1984: 183).

Habermas does not deny that the freedom to adopt differing attitudes toward the three domains of reality

> can become a source of conflict as soon as different cultural value spheres simultaneously penetrate the same institutional domains, so that rationalisation processes of different types compete with one another in the same place (Habermas 1984: 244-45).

But Habermas adds that

> cognitive-instrumental, moral-practical, and the aesthetic-expressive orientations of action ought not to become so independently embodied in antagonistic orders of life that they overburden the personality system's average capacity for integration and lead to permanent conflicts between lifestyles (Habermas 1984: 244-245).

The three value spheres - cognitive/instrumental, moral/practical, and aesthetic/expressive - are stabilised in permanent, cumulative learning processes, and these demarcate value spheres possessing their own "inner logics" and capable of being organised into professional discourses.

Reflexivity is a general characteristic of modern belief systems, institutionalised in the specialised discourses of science, jurisprudence, and aesthetics. For Habermas, increased reflexivity and the differentiation of value spheres result in a loss of meaning for the modern individual, but this loss also strengthens those communicative processes through which alone a sense of validity can be regained. Suppose that one were to raise the following objection: whatever one's evaluation of this process the argument concerning the binding nature of reflexivity begs the question. Certainly self-questioning, the justification of one's standpoint through reasoned argumentation, analysis of implicit and explicit presuppositions, and the like have been ideals in western culture.

Like Weber, Habermas maintains that cultural modernisation processes are in some sense irrevocable (Benhabib 1986: 275-7). In establishing this point, he often utilises arguments drawn from cognitive-developmental theories, which maintain that in a normal course of development, the individual will make the transition from one stage to another in a sequence which is irreversible. Once having attained stage four, adolescents do not fall back to stage two unless functioning as a normal human being is impaired by conditions of severe trauma, breakdown, shock, and the like. Of course, this covers a number of problems. One would want to know more precisely what these debilitating conditions are or could be.

Habermas claims that *factually* there has been an increase of rationality, cumulative learning process, in these three spheres - the cognitive-instrumental, the moral-practical, and the aesthetic-expressive ones. But equally he means that normatively one ought not to equate loss of meaning with a loss of rational potential in modern culture. The loss of meaning can be attributed to the cultural impoverishment of the lifeworld in the wake of the emergence of a culture of

experts who are in turn incapable of transmitting the achievements of science and art back to the lifeworld.

This is described as a process of rationalisation because it enhances the learning capacity of mankind. It does this because it provides actors with the conceptual means of constructing a reflexive or self critical perspective; that is, the categorical scaffolding constituted by the system of three world relations and corresponding validity claims makes possible an articulated consideration and evaluation of alternative interpretations of what is the case, what is legitimate and what is authentic selfexpression. It is this complex multidimensional learning potential of modernity that Habermas wants to emphasise, not just the mastering of formal, operational modes of cognition, leading to the capacity to do science and technology. The latter logocentric emphasis is in fact what has often led to an uncritical self-interpretation of the modern world that is fixated on knowing and mastering external nature.

In order for this enhanced learning potential which emerges with the rationalisation of worldviews to be set free, it must be crystallised into separate, "specialised forms of argumentation" which are institutionalised in corresponding cultural spheres. The communicative model gives Habermas what he thinks is a deeper understanding of that process of cultural rationalisation which begins with disenchantment and comes to fruition in the institutional anchoring of specialised forms of argumentation. This institutionalisation takes the following form.

> a) the establishment of a scientific enterprise in which empirical-scientific problems can be dealt with according to internal truth standards, independently of theological doctrines and separately from basic moral-practical questions; b) the institutionalisation of an artistic enterprise in which the production of art is gradually set loose from cultic-ecclesiastical and courtly-patronal bonds, and the reception of works of art by an art-enjoying public readers, spectators, and listeners is mediated through professionalised aesthetic criticism and c) the professional intellectual treatment of questions of ethics, political

theory, and jurisprudence in schools of law, in the legal system, and in the legal public sphere (Habermas 1984: 165-66, 340).

It is in relation to this overall process of cultural rationalisation that Habermas wants to stake his universalist claim. He asserts that

If we do not frame occidental rationalism from the conceptual perspective of purposive rationality and mastery of the world, if instead we take as our point of departure the rationalisation of worldviews that results in a decentered understanding of the world, then we have to face the question, whether there is not a formal stock of universal consciousness expressed in the cultural value spheres that develop, according to their own logics, under the abstract standards of truth, normative rightness, and authenticity. Are or are not the structures of scientific thought, posttraditional legal and moral representations, and autonomous art, as they have developed in the framework of Western culture, the possession of that "community of civilised men" that is present as a regulative idea? The universalist position does not have to deny the pluralism and the incompatibility of historical versions of "civilised humanity"; but it regards this multiplicity of forms of life as limited to cultural contents, and it asserts that every culture must share certain formal properties of the modern understanding of the world, if it is at all to attain a certain degree of "conscious awareness" or "sublimation." (Habermas 1984: 180).

Habermas argues that once this universally significant process of cultural rationalisation is conceptualised then one can properly understand the process of societal rationalisation. Habermas introduces another equally important concept, called the 'lifeworld' to supplement his theory of communicative action. The distinctiveness of Habermas's concept of lifeworld enters in modifying it in terms of rationalised lifeworld. The formal system of world relations and corresponding validity claims come to constitute, in the modern world, general structures of the lifeworld, that is, structures which remain the same even within different particular lifeworlds and forms of life. As these basic structures of modern consciousness are

institutionalised in differentiated cultural spheres, crucial change occurs in the relationship between action and lifeworld.

> To the degree that the institutionalised production of knowledge that is specialised according to cognitive, normative, aesthetic validity claims penetrates to the level of everyday communication and replaces traditional knowledge in its interaction-guiding functions, there is a rationalisation of everyday practice that is accessible only from the perspective of action oriented to reaching understanding - a rationalisation of action systems like the economy and the state. In a rationalised lifeworld the need for achieving understanding is met less and less by a reservoir of traditionally certified interpretations immune from criticism; at the level of a completely decentered understanding of the world, the need for consensus must be met more and more frequently by risky, because rationally motivated, agreement (Habermas 1984: 340).

In a rationalised lifeworld then, the formal scaffolding of modern structures of consciousness can increasingly be used by the individual as a framework in terms of which new experiences are accommodated to the stock of unproblematic, substantive background convictions which constitute his lifeworld. As this occurs, each agent's own critical capacities are increasingly rationalised. And this means actions are not guided by some authority or tradition but are subjected to discursive reasoning. In sum, Habermas introduces a second dimension to societal rationalisation, one that transpires "more in the implicitly known structures of the lifeworld than in explicitly known action orientations".

The question is, how does a society reproduce itself both symbolically and materially? Within Habermas's theoretical framework, the question of symbolic reproduction is the same as the question: how is the lifeworld reproduced? Communicative action plays an important role in the reproduction of the lifeworld, that is, in how communicative action generates ongoing patterns of social relations and the integration of individuals into them. Communicative action now becomes

interesting as a "principle of sociation". And one of the hallmarks of modernity is the enhanced role of this principle in organising the symbolic reproduction of society.

In order to grasp the reproduction process adequately the notion of the lifeworld must be further articulated. The diffuse notion of "unproblematic background" is simply too undifferentiated as it stands. In searching for a way to elaborate this notion, Habermas finds that no existing theories of the lifeworld satisfactorily grasp the breadth of what actually constitutes the unproblematic background of action. Different thinkers have focused on the lifeworld as a cultural storehouse, or as a source of expectations about the ordering of social relations, or as a milieu out of which individual competences for speech and action are formed. Habermas (1984: 167) wants to emphasis the fact that part of what constitutes a rationalised lifeworld is its "structural differentiation" of precisely these three dimensions: culture, society and personality .

The consequence of the structural differentiation of the 'lifeworld' into culture, society and personality also leads one to think in terms of differentiated processes of reproduction. These processes are identified by Habermas as "cultural reproduction, social integration, and socialisation". Habermas maintains that

> the unity of rationality in the multiplicity of value spheres rationalised
> according to inner logics is secured precisely at the formal level of the
> argumentative redemption of validity claims (Habermas 1984: 249).

Given his admission that arguments play different roles with different degrees of discursive "binding force" depending on whether they are 'cognitive', 'normative' or 'aesthetic', such a claim can hardly be expected to carry much conviction. The rational decentering of consciousness that allows actors to adopt different attitudes with respect to different domains of reality provides Habermas

with the reference points required to construct a nonselective model of social rationalisation. Such a model depicts those rationalisable action systems that must not be subordinated to laws intrinsic to heterogeneous orders of life if institutionalisation of the three value spheres is to proceed toward balanced equilibrium.

The concept of lifeworld (Habermas 1987: 119) has its roots in the phenomenological and hermeneutical tradition before it underwent reconstruction in Habermas' critical social theory. Habermas agrees with the phenemenological tradition that the lifeworld designates the horizon of unthematised, intuitive, and always already assumed expectations, definitions, and modes of orientation. Social action always occurs against the background of such a horizon; in each case, it is only a specific and limited segment of the lifeworld which is drawn into the action situation. Only the framework of communicative rationality can do justice to the three dimensionality of the lifeworld. This means that the first, lifeworld is the domain of cultural reproduction, social integration, and socialisation. Second, the lifeworld must not be viewed as transcendentally constituted, but as one that is reproduced over time and whose structures change. Against the atemporality of phenomenological analysis, Habermas insists that symbolic reproduction must be understood as a dynamic and temporal process. Both these theses are grounded in a third: linguistically mediated communicative action can fulfil all three functions of symbolic reproduction. Communicative action serves the transmission of cultural knowledge, of action coordination, and identity formation. The concept of lifeworld is complementary to communicative action. If one does not interpret the term 'constitution' as the achievement of a transcendental ego but as symbolic reproduction, one could say that in a sense the lifeworld is both constitutive of, and constituted by, communicative action.

Communicative action unfolds against this background of a semantically interpreted social lifeworld. It is the uniqueness of language that it can constitute it own reflective medium. In speaking, any of the validity claims which we take for granted can be challenged. When this background consensus breaks down, it can only be re-established via special argumentation procedures. It is possible to reach an understanding only by giving reasons to re-establish the validity of criticisable claims. Such reasons can be cited with reference to the three domains of external reality, which first become clearly differentiated from one another in modernity: these are the objective, the social, and the subjective worlds. In speaking we make reference to the world about us which we conceive of as the arena in which to carry out our action plans; to the social world, the rules and norms of which constrain us to act in certain ways rather than in others; and to our own subjective world of feelings, desires, and intentions. Communicative action has a threefold world reference (Habermas 1984: 88-9). These frameworks of reference, named the world, society and self are the *pragmatic presuppositions of our speech acts:* a cognitively interpreted external reality, a normatively interpreted social one, and an individually interpreted subjective sphere. In each of these domains agreement, once lost, can be reestablished by argumentation processes. We can thematise the validity claims of the truth of propositions which refer to the external world, the rightness or correctness of the norms which are invoked by them, and the authenticity or sincerity of a speaker's reference to his or her inner world.

Language can serve as a medium of action coordination only because it allows us to continue and reestablish, via argumentation, a background consensus of the institutionalised value spheres, i.e. science, morality and aesthetics (Habermas 1984: 94-5).

The key to Habermas' notion of reaching understanding is the possibility of using good reasons or grounds to gain intersubjective recognition for criticisable

claims. In fact, Habermas' claim is even stronger: he maintains that not only reaching understanding but understanding as such is only possible if we would know hypothetically what it would mean to redeem the claims to validity of certain utterances.

The claim has been named the cognitivist or rationalist core of Habermas' concept of action. For he maintains that the linguistic achievements of ordinary social actors presuppose a core of communicative competence, the essential aspect of which is the ability to continue and reestablish a lost consensus via argumentative processes in which reasons are advanced, debated, and evaluated. This communicative competence is said to be the essential medium through which the coordination of social action, the reproduction of cultural meanings, and individual socialisation take place. This means that social individuals can accomplish all three functions because they are able cognitively to judge the validity of certain claims on the basis of reasons advanced to back them up.

> There are number of ways of settling disputed claims - for example by appeal to authority, to tradition or to brute force. One way, the giving of reasons-for and reasons-against has traditionally been regarded as fundamental to the idea of rationality (McCarthy 1984: x).

In developing the model of communicative action, Habermas finally intends to break with the paradigm of the "philosophy of consciousness". From Descartes to Husserl, from Feuerbach to Adorno, the philosophical tradition has offered two models of the self: either the thinking, cogitative self or the active one appropriating and transforming nature. Either a lonely self cogitates upon an object, or an active self shapes the world. At least since Hegel's revival of Aristotle, attempts have been made in the modern tradition to understand intersubjectivity and the relation between selves as well. But the focus has been on consciousness, not on language-in-use. As Hegel put it, "a consciousness faces another

consciousness". Habermas maintains that neither the approach of continental philosophy which reduces interaction to a form of inter-consciousness, nor the debates concerning "other minds" predominant in analytic philosophy, can grasp the integrity of *social* interaction. Both approaches proceed from my mind to your mind, from my consciousness to your consciousness. Following an insight of Mead, Habermas claims that the "self" becomes an "I" by interacting with other selves.

> The Cartesian paradigm of the solitary thinker as the proper, even unavoidable, framework for radical reflection on knowledge and morality dominated philosophical thought in the early modern period. The methodological solipsism it entailed marked the approach of Kant at the end of the eighteenth century no less than that of his empiricist and rationalist predecessors in the two preceding centuries. This monological approach preordained certain ways of posing the basic problems of thought and action: subject versus object, reason versus sense (McCarthy 1984: vii).

Neither has grasped the suppressed traces of reason that provide the better grounding for a critical social theory of modernity (Pusey 1987: 15) The explication of the conceptual relationship among communication, argumentation, and reaching understanding through 'reasons' serves Habermas empirically to explicate the meaning of the rationalisation of the lifeworld in modernity. By rationalisation of lifeworld is meant nothing other than the increase in argumentative practices within the everyday world in the three crucial domains of action coordination, the reproduction of cultural tradition, and socialisation. Thus the rationalisation of the lifeworld takes place via these three institutionalised interrelated processes.

In the first place, the decentration of the modern worldview and the emergence of a dualistic ontology sharply distinguishes nature from culture, the external from the social world.

74

In the second place, this decentration of the worldview is accompanied by an increasing differentiation of the once-unified value spheres of science, morality, religion, law, aesthetics, and by institutionalisation of discourses aiming at working out the internal logic of these spheres.

In the third place, the rationalisation of the lifeworld results in an increase in reflexivity. The rationalisation of the lifeworld brings with it an increase in reflexive, argumentation modes of coming to grips with the contents of the socio-cultural universe as well as a reflexive reappropriation of the resources of the lifeworld.

In his *Legitimation Crisis* (1975: 45-50) Habermas attempts to reveal the immanent crisis tendencies of rationalisation processes. Since his early essays on *Theory and Practice* (1973), one of Habermas' chief concerns has been to show that the subject of modernity entailed a moral and political potential which could not be exhausted by the achievements of a primarily technical reason. In breaking with the utopian project of reconciliation with inner and outer nature as postulated by Adorno, Horkheimer and Marcuse, Habermas turned to this practical legacy of modernity. The thrust behind the distinction between the rationalisation of "systems of action" on the one hand and that of the "lifeworld" on the other is to suggest that the latter contains processes and achievements whose normative potential we have not yet exhausted.

Habermas introduces a differentiated sociological model which in the first instance distinguishes between 'system' and 'social' integration. Briefly explicated, the distinction between system integration and social integration is the following: by the former is meant a mode co-ordinating social action through the functional interconnection of action consequences, whereas social integration refers to the coordination of action through the harmonising of action orientations. The institutional differentiation of modern societies into the polity, the economy, and

the family meant that in two domains in particular - the economy and state administration - the coordination of social action was functionalised. Both the capitalist marketplace, and the modern state, relying on the medium of money as well as formal, juridical power, generate a series of social actions that influence one another through their unintended consequences. Behind the back of individuals and unintended by them, their actions give rise to other actions and reactions such that a domain of quasi self-regulating mechanism arises. But for Habermas, system integration refers only to one of the modes in which modern societies coordinate the action of individuals. The system perspective must be completed by the perspective of social integration (labour and interaction; work and politics).

Social integration means that individuals orient their actions to one another because they cognitively understand the social rules of action in question. Put in a nutshell, the thesis of *The Theory of Communicative Action* (Habermas 1987) is that rationalisation processes are paradoxical because they undermine the very rationality of the lifeworld which first made societal rationalisation possible. The rationalisation of the lifeworld, initiated by modernity, contains an emancipatory potential which is threatened by the dynamics of societal rationalisation spurred on by capitalist growth. The distinction between system and social integration in the legtimation crisis is the forerunner of the distinction between system and lifeworld in the TTCA (1987). As in Legitimation Crisis, in the present work as well, by 'system', Habermas means that social life can also be viewed as quasi-purposive whole, where the unintended consequences of social actions come together to yield functional interdependencies. Such functionally interdependent systems of action can regulate themselves, adapt to the environment, assume capacities of problem solving, and the like. Habermas maintains that the systemic perspective on society, which is always developed from the point of view of an observer, must be supplemented by the internal perspective of social actors.

The real methodological contribution of the TTCA (1987) consists in the introduction of the concept of lifeworld as the correlate of the concept of communicative action and in the explication of how communicative action can carry out the three functions of social integration, cultural reproduction, and socialisation.

Making this distinction between the rationalisation of "systems action" and of the "lifeworld" allows one to see that the "iron cage" of modernity is not as sturdy as it might first appear, and that late capitalist societies contain many contradictions. The pathologies of the lifeworld (Habermas 1987: 143) arise in three domains: in the sphere of culture reproduction, the consequence is a loss of meaning; in the sphere of social integration, anomie emerges; and as regards personality, we are faced with psychopathologies.

Habermas (1987: 153; 170-72) provides an analysis of social rationalisation from the dual perspective of society as "system" and "lifeworld" and a clarification of the way in which the normative principle of discourse ethics, via the model of communicative action, is reflected in institutions of the life-world.

The system/lifeworld distinction is introduced by Habermas both as a distinction between two analytic perspectives that the social theorist can adopt and as a real distinction between two different institutional orders in modern societies: institutions that are largely integrated through consensually accepted norms (society as life-world) and institutions that are largely defined by their capacity to respond to the functional requirements imposed by the environment (society as system). Habermas (1987: 145-47) has insisted that this distinction is crucial for his analysis of the 'crisis tendencies' that can be found in modern societies: in short, 'social pathologies' arise whenever the attempt to meet the requirements of 'system' maintenance spill over into domains of the 'lifeworld' properly integrated on the

basis of normatively secured or consensual interactions. Habermas refers to this process as 'colonisation of the lifeworld'.

The distinction between system and lifeworld (Habermas 1987: 153-170) is also important in that it enables Habermas to identify two different sides to the process of social rationalisation: on the one hand, rationalisation is conceived as the differentiation of subsystems from the sociocultural lifeworld (and from each other) and an increase in system complexity-this is the aspect of social rationalisation predominant in the work of classical sociologists such as Weber, Durkheim, and Parsons. On the other hand, Habermas also speaks of the potential for increased (communicative) rationalisation within a lifeworld that has to a large degree been relieved of tasks taken over by the subsystems. What is central to this aspect of social rationalisation is not the expansion of formal instrumental reason to further dimensions of social life, but an opening up of processes of symbolic reproduction to consensual agreement among autonomous individuals in light of criticisable validity claims.

In a similar vein, at the end of *Legitimation Crisis* (1975), Habermas argued that the impact of the monetary bureaucratic system upon the life-world would lead to an increased demand for participation and legitimation whose possibility would depend upon the presence of meaning and value orientations in the culture that could be reactivated such as to generate new motivational patterns for individuals. A crisis of legitimation was likely when expectations were generated which could not be satisfied within available patterns of value or through other compensatory mechanisms. Legitimation crisis presupposes a motivation crisis. Habermas pointed to certain transformation in the culture of late-capitalist societies, like the decline of civic, familial relations and maintained that bourgeois and pre-bourgeois traditions had been replaced by ideologies like scientism and universalist morality. To say that the structures of communicative rationality are

irrevocable can mean only the following: the legacy contains a potential which we would like to see realised, and for which we are ready to engage ourselves. The theory of cultural modernity is not a philosophy of history but a critical theory with practical-emancipatory intent. The fulfilment of this legacy is a practical question, not a theoretical one. The question here is: does such a demand for the fulfilment of modern reason project the image of a future we would like to make own?

Habermas wants to claim that the decentered worldview and the reflexive differentiation of value spheres are constitutive of communicative rationality, and that such a structure of rationality and its constituents are quasi-transcendental, irrevocable, and binding upon us. Consequently to deny this leads to "performative contradiction": critiques employ reason in order to attack reason.

Following Habermas, it can be argued that modernity has potential for different forms of rationality. Therefore, to presume strategic rationality and its universality as the only form of rationality which best explains human behaviour is a naive conclusion. Against the unidimensional and monological version of reason in rational choice theory, Habermas put forth his multidimensional rationalisation thesis. Modernity does not constitute only instrumental reason. Of course, recognising this plurality of modernity in differentiation of value spheres he also prescribes language as a universal mode of regulating different spheres. The question is, is rational choice conception of rationality unidimensional?

The project of an emancipated society does not entail the rejection of the cultural legacy of the moderns, but its completion. One must distinguish the distortions caused by the one sided rationalisation of the economy and the administration under capitalism from the rationalisation of life world. It is the destructive dynamics which must be reversed. The constituents of cultural modernity - decentration, reflexivity, and the differentiation of value spheres - are

binding criteria of rationality. The project of emancipated society implies the fulfilment of communicative rationality, not its transfiguration.

2.2 Habermas's transformation of epistemology: social evolution and the reconstructive sciences

Habermas' account of modernisation as rationalisation relies upon the concept of *"reconstructive sciences"* (Habermas 1979: 178-179). The aim of such 'reconstructive sciences' is to make individuals aware of the rule competencies which they always already practice as an implicit know-how. Reconstructions serve the task of self-reflection in the sense of making individuals aware of what they already know. However, unlike transcendental philosophy, reconstructive theorems do not assume that such deep structures are ahistorical, non-evolving frameworks (Rasmussen 1990: 20-7). On the contrary, Habermas views such deep structures as patterns of rule competencies which evolve in the history of the individual and of the species.

Social change cannot be observed from the standpoint of the observer alone. There are aspects of social evolution which must be viewed as sequences in a developmental logic and which can be reconstructed internally, that is, insightfully recapitulated from the perspective of the participants. Social development presupposes social learning: at any given stage, it must be possible to identify problems to which social agents respond in meaningful ways; these responses subsequently may be institutionalised. Social innovation occurs through the answers that social agents give to the ever new problems of their life-world. Through institutionalisation, the experiences leading to a specific answer set become part of the material and cultural history of society. These previous answers are available to social agents as the legacy of the past; they reproduce their life-

world by recapitulating these already available answers while creating and seeking new ones. The concept of the reproduction of the lifeworld thus presupposes the related notion of a learning process as an internally reconstructable sequence meaningful to participants themselves.

Habermas maintains that if it can be shown that criteria of communicative rationality are the results of learning processes that can be meaningfully reconstructed, then their quasi-transcendental status can be redeemed. They would then be shown to be deep structures underlying cognitive and interactive human competencies. These competencies are indeed changing and evolving but their evolution represents an internally compelling sequence.

Since Kant, the goal of transcendental arguments has been to establish the necessity and uniqueness of certain conceptual presuppositions (e.g., the pure concepts of understanding) and structural conditions (e.g., the necessary synthetic unity of consciousness). Such necessity may be logical in nature, or it may be "quasi empirical", as for example when we claim that unless we can distinguish between subjective experiences and intersubjectively shared common ones, no human self-consciousness ascribable to a continuing self is possible. While arguments concerning logical necessity can be developed in a straightforward manner, the greatest difficulty faced by transcendental philosophies, and which have led many critics to question their viability altogether, is the claim that there can be such "quasi-empirical" necessity.

Reconstructions in the sense suggested by Habermas can by no means meet the stringent requirements of transcendental arguments. First, since such reconstructions of action and speech competencies are developed by synthesising the conclusions of generative linguistics, genetic epistemology, and cognitive psychology, they are more empirical than transcendental.

There are three possible lines of response Habermas may provide in defense of the cogency of his rational reconstructions. The first claims that rational reconstructions are more *empirical* than transcendental; the second emphasises the *form/content* distinction and the third points to the empirical fruitfulness of reconstructive studies. In fact the very concept of quasi-transcendental may be a misnomer. What distinguishes rational reconstructions from other accounts like hermeneutic and deconstructionist accounts is not their special philosophical status, but their empirical fruitfulness in generating further research.

Communicative action is not a novel theme in Habermas' work. In his earlier work *Knowledge and Human Interests* (Habermas 1970b), Habermas attempted to follow the Kantian transcendental route, giving an analysis of the formal, universal and necessary conditions and structures of linguistic interaction. Such a transcendental analysis is supposed to reveal the basic universal norms underlying all possible speech. This analysis of universal and unavoidable conditions leads to the insight that speech is oriented to unconstrained consensus; as a norm, consensus functions both as a foundation and a regulative ideal, a basis and a goal, for actual acts of communication. It can become the moral and epistemic basis for emancipatory interests in autonomy, beyond cognitive interests in control. In his latter formulation, Habermas rejected this philosophical and transcendental justification of communicative action.

Habermas became increasingly skeptical of a purely transcendental justification of his claims about communication, although not about their content and political implications. He shifted to the second approach in order to transform the search for reflective transcendental foundations into the social scientific search for certain empirically universal processes of learning. While the claim to 'reason' made by Kant and the enlightenment needs to be defended, the transcendental role of philosophy itself as the final tribunal of reason cannot be.

In the broadest sense, Habermas's project is that of resolving what he takes to be the underlying problem of modern self consciousness which first emerged at the end of the eighteenth century, after the high water of the enlightenment. Modern consciousness is marked from this point on by a search for self-assurance, or a search for some standards which are both made available by that consciousness and yet can provide some normative guide for it in modern life. Modernity, in other words, "must create its normativity out of itself" (White 1988: 91). And it is Habermas's intention to develop just those standards which will allow modernity to interpret itself in a way which is self-critical, but which gives some basis for normative self assurance. To achieve these aims he must, in all three discussions, succeed in persuading us that reason, rationality and truth, confront us in world views, in culture, in history, and in the norms and values that guide ordinary human action.

McCarthy aptly describes Habermas' enterprise as constituting the

> idea of a critical social theory that would be empirical and scientific without being reducible to empirical-analytic science, philosophical in the sense of critique but not of presuppositionless 'first philosophy,' historical but without being historicist, and practical in the sense of being oriented to an emancipatory political practice but not to technological-administrative control (McCarthy 1979: vi).

In its post-Kantian form, philosophical thought remains epistemological, insofar as it originates in reflection in reason as it is embodied in cognition, speech and action. But the only way to approach each of these human capacities, whose performance manifests the basic competences of human reason, is through a new relationship to the human and social sciences. Epistemology can no longer be satisfied with second-order reflection, but must begin with first-order theoretical

statements of these abilities themselves, in the self-reflection of those sciences that try to understand and explain the products and structures of these abilities, as they are evidenced in their exercise and development in human societies. While such a critical theory does surrender the historical metaphysical claim to formulate a first philosophy, the epistemological claims for reason are now established, discovered and warranted in an empirical social inquiry. A foundationalist position in philosophy is one which claims that philosophy can, by some method, demonstrate the absolute, universal validity of some conception of knowledge or morality. This view that there are ahistorical conceptual or moral frameworks existing, as it were, above the heads of concrete historical actors, is one which has increasingly fallen into general disrepute. Habermas's work has often been suspected of harbouring some variant of foundationalism, and yet he has just as often claimed that his is a non-foundationalist position.

Habermas focuses his attempt to retain the normative rather than the foundational status of epistemological notions like truth or reason in an empirical research program he calls 'the reconstructive sciences'. Their aim is to develop a modest, fallibilistic, empirical and yet universal account of the rationality manifested in human activities. Habermas' attempt is to reconstruct the intuitive and pretheoretical know-how underlying practices and performances of basic human competence. Once the structure of these abilities is understood, their acquisition can be analysed as a learning process on both individual and social levels. These reconstructions yield universal knowledge about human competence, but yet are 'hypothetical' instead of necessary, 'empirical' rather than a priori; they raise universal, though defeasible, normative claims to validity useful in social criticism of various practices.

The general result of Habermas' reconstructive sciences is a *theory of social evolution*, which organises and shows the interrelationships of the sub-

theories of various individual competences in a general theory of species-wide learning. Such a theory is guided by two hypotheses: first, that learning is the basic evolutionary mechanism in culture; and second, that homologous patterns exist on ontogenetic (the sequence of events involved in the development of an individual being) and phylogenetic (the sequence of events involved in the evolution of species) levels for the cognitive development of individuals and of the species as a whole. Such development is not evolution in Darwinian sense. It is the competences and not societies which develop, so that we may count as evolutionary any cumulative learning process which has a direction and becomes embodied in individual members. Both social institutions and individual personalities are related to the current stages in the acquisition of various competences of cognition, speaking and acting. These same cognitive abilities can be shown, on the level of social theory, to be precisely the mechanisms for the reproduction and integration of society. Habermas' theory of social evolution is not about the teleological unfolding of a species subject, but rather about processes of rationalisation leading in the direction of some epistemic achievement. Habermas' theory is multidimensional, and includes a cognitive dimension, a moral dimension and a subjective dimension.

Habermas describes his theory of social evolution by way of a critique of the economic reductionism of historical materialism. His thesis is that developments in the sphere of social integration have their own independent logic:

> I am convinced that normative structures do not simply follow the path
> of development of reproductive processes...but have an internal history
> of their own (Habermas 1979: 117).

Invoking Aristotle, Habermas insists that *praxis* cannot be reduced to *techne*, nor rationality to purposive or instrumental rationality. Rationalisation

processes in the sphere of communicative action are neither identical with nor an immediate consequence of rationalisation processes in the sphere of productive forces. The rationalisation of action takes effect not only on productive forces but also, and independently, on normative structures (moral representations, norms, worldviews, identity formation, law etc.).

In this connection reconstruction means that one takes a theory apart and puts it back together in a new form, in order better to achieve that goal which it set for itself. The reconstruction of historical materialism aims at a general theory of social evolution capable of explaining particular evolutionary developments - above all, the transition from primitive societies organised around kinship relations to civilisations organised as class societies with a differentiated political system, and the transition from developed premodern civilisations to liberal capitalist society, as well as the evolution of the latter to the stage of advanced capitalism. The term 'evolution' implies a conception of cumulative learning processes in which a direction can be perceived.

Habermas argues that

> development of productive forces depends on the application of technically useful knowledge; and the basic institutions of a society embody moral-practical knowledge. Progress in these two dimensions is measured against the two universal validity claims we also use to measure the progress of empirical knowledge and of moral-practical insight namely, the truth of propositions and the rightness of norms (Habermas 1979: 142).

> Social evolution is conceived as a bidimensional learning process (cognitive/technical and moral/practical), the stages of which can be described structurally and ordered in a developmental logic. The emphasis is not on the institutionalisation of particular contents, but on the "institutional embodiment of structures of rationality", which makes learning at new levels possible, that is, on learning applied to the structural conditions of learning. In one sense it is only socialised

individuals who learn; but the learning ability of individuals provides a "resource" that can be drawn upon in the formation of new social structures (McCarthy 1979: xxii).

The categorical distinction between purposive rational action and communicative action thus permits us to separate the aspects under which action can be rationalised. As learning processes take place not only in the dimension of objectivating thought but also in the dimension of moral-practical insight, the rationalisation of action is deposited not only in forces of production, but also - mediated through the dynamics of social movements - in forms of social integration. Rationality structures are embodied not only in amplifications of purposive-rational action-that is, in technologies, strategies, organisations, and qualification - but also in mediations of communicative action- in the mechanisms for regulating conflict, in the worldviews, and in identity formation (Habermas 1979: 120).

He defends the thesis that the development of these 'normative' structures is the pacemaker of social evolution, for new principles of organisation mean new forms of social integration and the latter, in turn, make it possible to implement available productive forces or to generate new ones, as well as a heightening of social complexity. Communicative action has to be understood to analyse the symbolic structures that underlie law and morality, an intersubjectively constituted world, and the identities of persons and collectivities. In working out the logic of development of normative structures, Habermas' strategy is to employ structural comparisons with the developmental logic worked out for ontogenetic processes in the framework of his theory of communicative competence. The fundamental question is, how does the evolutionary step to a new form of social integration come about? How is it possible? The descriptive answer of historical materialism is through social contradictions and conflicts, political struggle, and social movements. But a theory of social evolution requires an *analytic* response. The research program designed to carry through on this suggestion is sufficiently articulated to permit a sketch of its main outlines. Habermas proposes that

the species learns not only in the dimension of productive forces but also in the dimension of moral-practical consciousness decisive for structures of interaction. The rules of communicative action do develop in reaction to changes in the domain of instrumental and strategic action; but in doing so they follow their own logic (Habermas 1979: 148).

The leading idea is that social evolution can be comprehended as a learning process, not in the sense of behaviouristic psychology - which is not complex enough to grasp more than peripheral learning mechanisms - but in the sense of cognitive developmental psychology. Central to this approach is the notion of a developmental logic that incorporates a distinction between formally characterised levels of learning and the learning processes that are possible at each level. Drawing on these ideas, Habermas construes organisational principles of society as sociocultural innovations that institutionalise developmental logic of learning; they establish the structural conditions for technical and practical learning processes at particular stages of development. Principles of organisation circumscribe ranges of possibility within which institutional systems can vary, productive forces can be developed and utilised and system complexity and steering capacity can be increased. The concrete embodiments of these abstract principles are the institutional nuclei that function as relations of production and determine the dominant form of social integration (McCarthy 1978: 246).

Habermas draws on various sources to explain his theory of social evolution as constituting a learning process in the development of normative structures of social integration. Central to his theory of social evolution is the notion of *developmental logic*. Developmental studies are well underway in a number of different areas - psycholinguistics, cognitive psychology (including moral consciousness), psychoanalysis (including ego psychology), and social interactionism, among others. The task, as Habermas sees it, is to work out an integrated framework in which the different dimensions of development are not only analytically distinguished but their interconnections systematically taken into account. Habermas's proposals are to show the basis of his theory of

communication drawn among different dimensions in which an utterance can succeed or fail: 'comprehensibility', 'truth', 'rightness' and 'truthfulness'.

> Each of these specifies not only a dimension of communicative action, and thus of rationality, but a 'region' of reality - 'language', 'outer nature', 'society', and 'inner nature' - in relation to which the subject can attain varying degrees of autonomy (McCarthy 1979: xx).

In both dimensions i.e., the structures of the ego and of worldviews, development apparently leads to a growing decentration of interpretive systems and to an ever clearer categorical demarcation of the 'subjectivity' of internal nature from the 'objectivity' of external nature, as well as from the 'normativity' of social reality and the 'intersubjectivity' of linguistic reality.

Using this universal-pragmatic classification of validity claims and corresponding 'regions' of reality as a guide, Habermas has advanced some tentative suggestions for unifying developmental studies.

Habermas' explication of the key notions of a developmental logic and of levels or stages of learning are adapted from the Piagetian tradition in cognitive psychology. The idea underlying ontogenetic studies of this type is that the various abilities of the adult subject are the result of an integration of a maturational and learning processes. These run through an irreversible sequence of discrete and increasingly complex developmental stages, whereby no stage can be passed over and each higher stage implies or presupposes the previous stages (child; youth and adolescent; adult). Stages are constructed wholes that differ qualitatively from one another and no later phase can be attained before earlier ones have been through, and elements of earlier phases are preserved, transformed, and reintegrated in the later. In short, the developmental logical approach requires the specification of a hierarchy of structured wholes in which the later, more complex, and more encompassing developmental stages presuppose and build upon the earlier.

Admittedly, ontogenetic (the sequences of events of involved in the development of an individual organism) and phylogenetic (the sequence of events involved in the evolution of a species) models of developmental processes cannot be transposed without problems to the domain of social evolution. However, the suggestion is that the learning ability of individuals is the basis of societal learning: ontogenetic learning processes provide a resource that can be drawn upon in the formation of a new social structures. At the same time, these processes are themselves conditioned by the developmental level of society.

Under 'society' Habermas understands all systems which - through linguistically coordinated (instrumental and social) actions - appropriate outer nature in production processes and inner nature in socialisation processes. Accordingly a theory of evolution of the type Habermas envisages would have to construct a developmental logic for both dimensions: 'productive forces' and 'forms of social integration'. Habermas' argument is that the development of productive forces cannot be grasped independently of developments in the forms of social integration. In making the distinctions between labour/language; production/communication; material/symbolic, Habermas' aim is to sketch out the normative structures pertaining to processes of social integration. There are a number of different aspects to the evolution of forms of social integration that would have to be analytically distinguished and investigated before a unified account of their interrelations could be worked out. And his own work has been concerned primarily with the development of legal and moral systems. Habermas uses Kohlberg's schema for the development of moral consciousness (more precisely, of its cognitive side, the ability to make moral judgments) as a clue to the development of moral and legal systems, for these represent attempts to resolve morally relevant conflicts on a consensual basis and without resort to manifest violence (McCarthy 1978: 251)

Even if the tentative sketch could be worked out in satisfactory detail and in accord with available empirical evidence and even if it were possible to show that it represented a developmental logic (that is, to demonstrate the relevant hierarchical relations of dependence and interdependence among the different stages of morality and legality), it would clearly not yet be a theory of social evolution. Hierarchical structural patterns do not themselves supply an account of how and why developments actually come about.

The explanation of a transition from one social formation or principle of social organisation to another requires - in addition to a structural description of the relevant levels of learning - recourse to systems problems (which overburden the capacities of the old social formation), to societal learning processes (in which the surplus learning capacities of individuals are exploited and institutionalised in a new form of social organisation), and to contingent initial and boundary conditions (which stimulate or prevent, support or hinder, further or limit these processes).

The theory of social evolution is still left with the problem of explaining how what is learned by individuals or marginal social groups could eventually become the basis for a new principle or organisation of a society as a whole. Habermas' suggestion is that the results of evolutionarily relevant learning processes find their way into the cultural tradition, the worldviews and interpretive systems of society; in the form of empirical knowledge and moral-practical insights, they comprise a kind of cognitive potential that can be drawn upon by social movements when irresolvable systems problems require a transformation of the basic form of social integration.

In carrying this evolutionary model over to social development, Habermas calls for a word of caution. Its usefulness consists in its directing our attention to the evolutionary learning mechanism. The socio-cultural stages of developmental learning processes are socially organised from the start, so that the results of

learning can be handed down. Thus cultural tradition provides a medium through which variety-generating innovations can operate after the mechanism of natural evolution has come to rest. In the case of social evolution the learning process takes place not through changes in genetic makeup but through changes in knowledge potential. The identity of a society is normatively determined and depends on cultural values; on the other hand, these values can change as a result of learning processes.

The cognitive developmental psychology has shown that in ontogenesis there are different stages of moral consciousness, stages that can be described in particular as preconventional, conventional and postconventional patterns[1] of problem solving. Habermas argues that the same patterns turn up again in the social evolution of moral and legal representations.

Habermas argues that all three complexes (*worldviews* and *collective identities and law and morality*) lead back to structures of linguistically established intersubjectivity.

> a) Law and morality serve to regulate action conflicts consensually and thus to maintain an endangered intersubjectivity of understanding among speaking and acting subjects.
> b) The demarcation of different universal object domains - one of which appears in the propositional attitude of the observer as external nature, a second, in the performative attitude of the participant in interaction as normative social reality, and a third in the expressive attitude of one who expresses an intention as his own subjective nature - makes possible the differentiation of those validity claims (truth, rightness, truthfulness) that we implicitly tie to all speech actions.
> c) the construction of personal and corresponding collective identities is a necessary presupposition for taking on the general communication roles, which are provided for in every speaking and acting situation and which find their expression in the employment of personal pronouns (Habermas 1979: 116).

Habermas considers the acquisition of interactive competence, the ability to take part in increasingly complex interactions, to be "the core of identity formation". Thus, with the advent of modernity, there is a differentiation of value spheres into *cognitive-instrumental, moral-practical, aesthetic-expressive*. Communicative action analyses the symbolic structures that underlie *law and morality*, an intersubjectively constituted world, and the identities of persons and collectivities. Communicative action is oriented to observing intersubjectively valid norms that link reciprocal expectations. In communicative action the validity basis of speech is presupposed. The universal validity claims (truth, rightness, truthfulness), which participants at least implicitly raise and reciprocally recognise, make possible the consensus that carries action in common. This forms the background consensus which is institutionalised. Institutionalisation again means the organisation of consensual action resting on intersubjectively recognised validity claims. It is in this sense that communicative action is seen as oriented toward reaching an agreement.

Communicative action can be rationalised neither under the technical aspect of the means selected nor under the strategic aspect of the selection of means but only under the *moral-practical* aspect of the responsibility of the acting subject and the justifiability of the action norm.

Rationalisation means extirpating those relations of force that are inconspicuously set in the very structures of communication and that prevent conscious settlement of conflicts, and consensual regulation of conflicts, by means of interpersonal communication barriers. Rationalisation means overcoming such systematically distorted communication in which the action-supporting consensus concerning the reciprocally raised validity claims- especially the consensus concerning the truthfulness of intentional expressions and the rightness of underlying norms - can be sustained counterfactually.

To the extent that action conflicts are not regulated through force or strategic means but on a consensual basis, structures emerge that mark the moral consciousness of the individual and the legal and moral system of society. They comprise the core domain of the aforementioned general action structures - the representations of justice crystallising around the reciprocity relation that underlies all interaction. Thus, Habermas elaborates communicative action as a moral-practical insight.

2.3 Strategic rationality and Communicative rationality - Habermas on Social Interaction

Given this exposition of a critical theory of modernity based on the theory of communicative rationality, I shall now discuss Habermas' contrast of two different forms of social interaction: one is instrumental-strategic rationality and the other communicative rationality.

As can be seen from the discussion of rationality, Habermas articulates an. expansive critical social theory anchoring his project in a comprehensive concept of communicative rationality intended to remedy the concept of cognitive-instrumental rationality. In this section the focus is on the typology of action that he constructs around the concept of communicative rationality following from his larger question of the modernity. Although Habermas' work has been extensively discussed with reference to theoretical paradigms like hermeneutics, post-structuralism, philosophies of science etc., unfortunately not much attention is given to the rational choice paradigm.

One of the central lessons of Habermas TTCA (vol. I 1984; vol. II 1986) is that unlike the rational choice model of instrumentally rational self-interested maximisation, critical theory expects individuals to act in different ways in different

circumstances. Their rationality has 'cognitive', 'moral' and 'aesthetic' aspects, not just instrumental ones. Thus it invokes different motivations to explain social interaction. The universality of communicative reason lies in the fact that it unifies all the three value spheres of life in devising argumentative procedures. From the standpoint of critical theorists, strategic rationality conceptualises action in a monological model of subjectivity.

The discussion of rational action is the fulcrum for Habermas' broader empirical and normative concerns. Johnson (1991a) gives a critical account of this typology. Johnson's discussion centers on the place Habermas assigns to strategic action in his typology. This issue is typically neglected in treatments of Habermas' work. Before proceeding to a critical discussion of the typology, it is important to note that Habermas, unlike early Frankfurt school theorists, does not deny the role of strategic action in the social and political world. His objection is its claim to 'universality'. Although Habermas himself has not engaged in discussion with rational choice theory, critical theorists under the banner of communicative action do, for example White (1988: 10-13, 44-47).

It is important to see how far Habermas is successful in accommodating both domains of action into his theory. Given their apparently divergent approaches to the questions of rationality, both critical and rational choice theory can be viewed as competing research traditions. It is my task to encourage constructive exchange between advocates of each tradition. There are three reasons for advancing such an enterprise.

First, because both are preoccupied with determining what rationality can mean in the realm of social and political interaction.

Second, Habermas' aims are not narrowly philosophical and are concerned explicitly with 'empirical usefulness' contributing to concept formation in the social sciences. He is thus engaged in a similar enterprise as rational choice theorists.

Third, rational choice theorists typically concentrate on interactions between rational actors within a prespecified context. Habermas also focuses on the efforts of agents to define the context within which their ongoing interactions will transpire. Habermas is addressing a conceptually notorious issue problematic for rational choice theory.

Within the broad category of rational action Habermas differentiates between "action oriented to success" and "action oriented toward reaching understanding" (Habermas 1984: 286-87). Success is evaluated relative to states of affairs purposefully generated by intervention in the world, whereas reaching understanding is a process by which participants seek agreement concerning the nature of their interaction. Such an agreement defines the context within which actors pursue their individual plans.

Habermas uses this basic distinction between orientations to generate a complex typology of action. He firsts sets off a range of social action from instrumental intervention in the physical environment. Instrumental action is oriented to success, but it is not unique in that regard. Strategic action is social action oriented to success in influencing the actions of other rational actors.

Communicative action consists of attempts by actors to co-operatively define the context of their interaction in such a way as to enable them to pursue their individual plans. It is the paradigmatic form of social action oriented toward reaching understanding. In norm-regulated action, participants conform to socially expected modes of behaviour. In dramaturgical action, agents constitute a public for one another before whom they can present their selves.

It is to the experience of achieving mutual understanding in communication that is free from coercion that Habermas looks in developing his idea of rationality (Habermas 1984: x).

Habermas views all the above types of action as rational in a specific sense. To be rational action, an action minimally must be capable of being "defended against criticism" (Habermas 1984: 8, 16). Habermas argues that each type of action is tied to a characteristic "validity claim" in the light of which particular actions of that type, in principle, can be criticised, defended, and hence regarded as potentially rational. Moreover, he treats the connection between rational action and validity claims as a matter of conceptual necessity (Habermas 1984: 9, 19, 42, 38). Instrumental and strategic action raise claims to truth or effectiveness. Normative action raises a claim to rightness. Dramaturgical action raises a claim to sincerity or authenticity. Parties to communicative action can raise validity claims of each sort.

Validity claims are in turn grounded in what Habermas calls formal concepts. These concepts, shared by participants to an interaction, specify the character of the objective, social, and subjective world. Instrumental and strategic action presuppose only an objective world constituted by pre-existing states of affairs. In addition to this objective world, normative action relates to a social world of legitimately ordered relations. Dramaturgical action publicly expresses an inner or subjective world of desires and feelings to which an agent has privileged access. Parties to communicative action potentially relate to all three worlds. The unity of rationality in the multiplicity of value spheres rationalised according to their inner logics is secured at the formal level of the argumentative redemption of validity claims. Habermas depicts communicative reason - embodied in validity claims to truth, rightness, and sincerity that are implicitly and necessarily raised in human speech - as an "unavoidable pragmatic presupposition" of language use and hence of social interaction.

Communicative action involves participants in "the cooperative negotiation of common definitions of the situation" in which they are interacting. In everyday communicative practice, this process of mutual interpretation remains implicit. In

more reflective forms of communicative action - what Habermas refers to as discourse or argument - it is made explicit. At both levels, participants in communicative action seek an argument which can admit of consensus. In doing so, they advance and respond to validity claims, using the formal world-concepts as a commonly supposed system of coordinates.

Parties to communicative action seek to reach an understanding regarding the situation within which they pursue their individual plans in one or another direct operation on the objective, social, or subjective world. Actors have two options when the process of reciprocal interpretation implicit in everyday interaction deteriorates. They can have recourse to discourse or argument "as a court of appeal that makes it possible to continue communicative action with everyday routines."

> Alternatively, they might resort to strategic action, abandoning in the process, their joint endeavour after consensus. The resulting interactions will be coordinated by quite distinct mechanisms - by consent in the case of communicative action and by influence, arbitrary choice or complementarity of interest in the case of strategic action (Habermas 1982: 237, 264-65).

For Habermas, this contention is not just an academic issue but has grave political implications. He claims that

> to the degree that interactions cannot be coordinated through achieving understanding, the only alternative that remains is *force* exercised by one against others (in a more or less refined, more or less latent manner). The typological distinction between communicative action and strategic action says nothing else than this (Habermas 1982: 269).

Habermas aspires to articulate a comprehensive concept of rationality that will dislodge teleological action from the central role it traditionally is accorded by

social and political theorists. He suggests that pre-eminent social scientific reflections on rational action are founded in the writings of Max Weber. But he criticises Weber for privileging teleological action, thereby classifying "all other actions... as specific deviations from this type. Habermas seeks to overcome this one-sidedness in his own typology of action.

Habermas characterises reaching understanding as the inherent telos of human communication. Yet he also recognises that not all "linguistically mediated interaction" in fact manifests that telos. Theoretical commitments thus compel him to demonstrate that these other sorts of interaction are parasitic on the "original mode of language use". In attempting to surmount what he perceives to be the one-sidedness of traditional social and political theory, Habermas dislodges "teleological" action from its privileged theoretical position only to replace it with a similarly privileged concept of communicative action.

But Habermas not only presents strategic action as derivative of communicative action, he is committed to viewing it as peripheral insofar as strategic action neither potentially relates to all three worlds nor operates reflexively as does communicative action. It thus is difficult to see how he might incorporate it as other than a residual category in his typology of action.

This shift in theoretical focus is not a major failing. It is problematic only if 1) it produces conceptual distortions parallel to those Habermas decries in traditional accounts of rational action or 2) Habermas is unable to persuasively explain the force of the now privileged concept of communicative action.

Habermas argues that social actions can be distinguished according to the mechanisms for co-ordinating individual actions, for instance according to whether a social relation is based on 'interest positions' alone or on 'normative agreements' as well. It is in this way that Weber distinguishes the sheer facticity of an 'economic' order from the 'social' validity of a legal order. In the one case, social

relations gain stability through the factual intermeshing of 'interest positions'; in the other, through an additional recognition of normative validity claims. A coordination of actions secured only through the complementarity of interests can be normatively restructured through the superaddition of validity based on agreement, that is through the "belief that a certain behaviour is required by law or convention". Weber elucidates this connection with the development of traditions through the transition from custom to convention: It is by way of conventional rules that merely factual regulations of action- that is, mere custom - are normally transformed into binding norms, guaranteed primarily by psychological coercion (Habermas 1984: 282-3).

Interaction based on complementarity of interests exists not only in the form of custom - that is, of insensibly accepted habituation - but also at the level of rational competitive behaviour, for example in modern commerce, in which participants have formed a clear consciousness of the complementarity as well as of the contingency of their interest positions. On the other hand, interaction based on normative consensus does not only take the form of tradition-bound, conventional action; the modern legal system is dependent on an enlightened belief in legitimation, which rational natural law - in the idea of a basic contract among free and equals - traces back to procedures of rational will-formation (Habermas 1984: 283). Although appreciating Weber's typology of action, Habermas argues that Weber did not fully extend the typology of action for the problematic of societal rationalisation. If processes of societal rationalisation are to be investigated in their entire breadth, other action-theoretical foundations are required. Habermas employs the concept of communicative action to explicate the complex concept of rationality.

Social actions are distinguished according to two action orientations - corresponding to the coordination of action through interest positions and through

normative agreement (Habermas 1984: 285). The model of purposive rational action takes as its point of departure the view that the actor is primarily oriented to attaining an end, that he selects means that seem to him appropriate in the given situation, and that he calculates other foreseeable consequences of action as secondary conditions of success. Success is defined as the appearance in the world of a desired state, which can, in a given situation, be causally produced through goal oriented action or omission.

Habermas argues that a categorical distinction needs to be made between the rationalisation of communicative action on the one hand, and of instrumental rationality (with regard to nature) and strategic rationality (with regard to social sphere) on the other. He calls an action oriented to success 'instrumental' when applied to non-social realm of activity. By contrast, he calls an action 'strategic' when understood in a social sphere wherein the interaction is coordinated in a two way process. Habermas speaks of communicative action as an action which is coordinated not through egocentric calculations of success but through acts of reaching understanding (Habermas 1984: 285-6).

In communicative action participants are not primarily oriented to their own individual successes; Classifying these two actions as 'action oriented to success' and 'action oriented to reaching understanding', Habermas deploys Austin's speech act theory, in order to explicate the differences (Habermas 1984: 288:9) Strategic action is related to perlocutionary acts and communicative action is related to illocutionary acts. Social actions can be distinguished according to whether the participants adopt either a success-oriented attitude or one oriented to reaching understanding.

A communicatively achieved agreement has a rational basis, it cannot be imposed on either party, whether instrumentally through intervention in the situation directly or strategically through influencing the decisions of opponents.

Agreement can indeed be objectively obtained by force, but what comes to pass manifestly through outside influence or the use of violence cannot count subjectively as agreement (Habermas 1984: 287).

The unity of rationality in the multiplicity of value spheres rationalised according to their inner logics is secured precisely at the formal level of the argumentative redemption of validity claims (Habermas 1984: 249) .

This distinction between strategic action and communicative action parallels another similar problem discussed widely in his works which is 'labour' and 'interaction' or 'work' and 'politics' or 'production paradigm' and 'symbolic or language paradigm'. Habermas is critical of the orthodox political economist who tend to regard the domain of politics as a reflection of underlying material forces instead of recognising that values and norms shape the world. Politics is creative and it has been responsible for a reorganisation of relations among groups in society.

Thus Habermas' typology of actions is a broad one. Communicative rationality is appropriate to the interactions of competent, reflective critical, and social individuals. Critical theory has no parsimonious paradigm of subjectivity but instead has cognitive, moral and subjective aspects, not just instrumental ones. In other words, the development of world views, ethical positions and identity contributes to and helps constitute an individual's rationality. Above all, rationality means the competence to decide when it is appropriate to act instrumentally, or in conformity with social norms, or dramaturgically as an expressive subject; and the ability to judge these qualities in others. Thus, from the standpoint of critical theory, public choice's model of subjectivity misses the richness of human motivations and appears impoverished and emblematic of instrumentally rationalised modernity. To Marcuse this is a "one-dimensional man".

Conclusion

Habermas' alternative model of collective choice is conceived not as a process of preference aggregation, in which there is a mapping from a set of individual orderings to a social ordering, but as a process of dialogue in which reasons are exchanged between participants in a process that is perceived to be a joint search of consensus. Such a dialogic concept of collective choice would necessarily work not with fixed preferences to be amalgamated, but preferences that were altered or modified as competing reasons were advanced in the course of discussion. Such a dialogic exchange would not necessarily preclude occurrences of paradoxes of collective choice, since differences of judgement might remain after debate and dialogue, leaving voting as the only practicable way of overcoming disagreement. Yet, in this context, the paradoxes would be seen not as the proof that the popular will was a meaningless concept, but as revealing the as yet unresolved imperfections of a process of discussion that characterised an adequate concept of collective choice.

In this chapter, I have shown how collective choice is conceptualised from a communicative action perspective. In the next chapter, I shall discuss the internal critiques of rational choice theories and make way for a critical engagement between critical theory and rational choice theory.

The contention from the critical theorists' point of view is that public choice theories are reductionist in the sense of treating individual behaviour in all realms on the basis of *homo economicus* principle. Unlike the reductionist variants of political theory, it is important to recognise that values and norms also shape the world. Politics is a creative and dynamic process in so far as the aim is to deliberate over norms and rules which regulate the social conduct of everyday lives of the individuals. This is the concern which Habermas addresses by bringing back to life

Aristotelian distinction between *praxis* and *theoria*. Habermas is concerned with generating a consensus in social interaction of individuals through dialogue. Moreover his enterprise is aimed at providing a reference point to unravel questions of power, domination and coercion. On his account, the rational choice enterprise evades the question of ideology, power, domination in social interaction.

Critical theory is largely an external critique and so is uninformed by the theoretical developments in the rational choice paradigm. In the next chapter, I shall consider the internal critiques of rational choice who have come to recognise the limits of instrumental reason. I shall return to the discussion of strategic and communicative action in the light of internal critiques. My aim is to encourage a productive conversation between the two discordant traditions in the light of these critiques.

Notes

[1] These stages are associated with Lawrence Kohlberg' stages of moral consciousness:
Preconventional: It means the responsiveness to cultural labels of good and bad and are interpreted in terms of physical consequences of action (punishment).
Conventional: In this stage, allegiance to a group e.g. family or nation is perceived as valuable in its own right. Here the attitude of the individuals is conformity to existing norms.
Post-conventional: At this level the moral values and principles are defined independent of any authority of the groups. This stage includes a free and equal persons coming to an agreement on moral norms.

Chapter 3

Contract or Discourse: Modes of Collective Choice

Introduction

In chapter 1, I discussed two problems of rational choice: one is the problem of preference aggregation and the other collective action problem. Here, I am primarily concerned with the first problem i.e. the inadequacy of "majority rule" as a fair procedure for resolving conflict of interests. Rational choice findings about the logic of majority rule have prompted new reflections on the normative foundations of democracy (Green and Shapiro 1994: 4). One such attempt came from Buchanan who developed a contractarian model of constitutional choice. Interestingly, even critical theorists are critical of democratic institutional arrangements that rely solely or even primarily on electoral mechanisms, that is, on ways of aggregating individual interests or preferences. They argue that the results underscore the inadequacy of purely aggregative notions of democracy in ways that can sharpen and ultimately sustain the case for deliberation.

In this chapter, I shall contrast two normative theories of collective choice: contract and discourse ethics.[1] The discussion is centered on the normative problem of "agreement" which is the central concern of both contract theories and discourse ethics. In the first section, I explicate the central tenets of discourse ethics. And in the second section, I compare and contrast discourse and contract as two modes of agreement. And finally in the third, I argue, following Benhabib, that dialogue and contract are no way different if interpreted as procedural ideals. The significance of discourse ethics is realised only when interpreted as a moral-transformatory process as opposed to a mere legal-juridical idea.

3.1   The Idea of Justification of Discourse Ethics

3.2   Discourse ethics and Contract: Modes of Agreement

3.3   Interpreting Discourse Ethics

3.1   The Idea of Justification of Discourse Ethics

The program of discourse ethics forms the anticipatory-utopian dimension of Habermas' critical social theory. The most ambitious and original feature of Habermas's whole work is his attempt to recast the study of society in a theory of communication. As Benhabib (1986: 225) argues, the task of critical social theory is not to develop Kantian imperatives, but to show the potential for rationality and emancipation implicit in the present. Communicative ethics constitutes what is called the normative aspect of critique. It is aimed at projecting the counterfactual historical possibilities of conceptualising interaction differently. The program of communicative ethics belongs to the anticipatory-utopian aspect.

I shall outline the fundamental features of Habermas' theory of communicative competence and provide a clarification of the normative grounds on which the practice of justifying and/or criticising social norms and institutions rests, at least within a democratic regime. This will make sense only if Habermas' single aim is kept in mind. He wants to develop a model that will show how rationality is manifested in ordinary social interaction, that is to say, in ordinary communication between speaking and acting subjects.

The two key notions that are central to Habermas' (1979: 26) theory are: one, the notion of the ideal speech situation, and the second, his concept of communicative competence. Both are closely intertwined. Communicative action is

intrinsically dialogical. The starting point for an analysis of the pragmatics of speech is the situation of a speaker and a hearer who are oriented to mutual reciprocal understanding; a speaker and hearer who have the capacity to take an affirmative or negative stance when a validity claim is challenged.

In the context of Habermas' theory of universal pragmatics (Habermas 1979: 1-69), the concepts of an "ideal speech situation" and "unconstrained dialogue" define the formal properties of argumentations named 'discourses.' According to this theory, anyone engaging in communication, performing a speech act, raises four validity claims and presupposes that they can be vindicated or justified when challenged. In discourses, the background consensus which we 'always already' naively assume is constituted by the recognition of these four validity claims: the comprehensibility of our utterances, the truth of their propositional component, the correctness or appropriateness of their performative aspect, and the truthfulness or authenticity of the speaking subject (Habermas 1984: 23).

Habermas names discourses through which 'truth' and 'normative' claims are thematised "theoretical" and "practical" ones (Habermas 1984: 19).[2] The aim of discourses is to generate a "rationally motivated consensus" on controversial claims. The concept of the ideal speech situation is introduced in this context. The ideal speech situation specifies the formal properties that discursive argumentations would have to possess if the consensus thus attained were to be distinguished from a mere compromise or agreement of convenience. The ideal speech situation is a 'meta-norm' that applies to theoretical as well as to practical reason. It serves to delineate those aspects of an argumentation process which would lead to a "rationally motivated" as opposed to a false or apparent consensus.

The four conditions of the ideal speech situation are: first, each participant must have an equal chance to initiate and to continue communication; second, each

must have an equal chance to make assertions, recommendations and explanation and to challenge justifications. Together they are called "symmetry conditions". Third, all must have an equal chance as actors to express their wishes, feelings, and intentions; and fourth, the speakers must act 'as if' in contexts of action there is an equal distribution of chances to order and resist orders, to promise and to refuse, to be accountable for one's conduct and to demand accountability from others. This can be called the "reciprocity condition". While the symmetry stipulation of the ideal speech situation refers to speech acts alone and to conditions governing their employment, the reciprocity condition refers to existing action contexts, and requires a suspension of situations of untruthfulness and duplicity on the one hand, and of inequality and subordination on the other.

Thus, the ideal speech situation is a 'counterfactual' ideal of organising social action around talk. It describes a set of rules which participants in a discourse would have to follow (symmetry condition), and a set of relations (the reciprocity condition) which would have to obtain between them, if we were to say of the agreement they reach that it was rationally motivated, dependent on the force of the better argument alone. One of the preconditions of the 'ideal speech situation' is that "the speakers should deceive neither themselves nor others about their intentions."

Habermas' discourse ethics is a defense of 'moral universalism'. He claims that not only those who choose to participate in an argumentation situation, but all those capable of speech and action, as competent subjects, dispose of a certain intuitive know-how. The task of universal pragmatics is to render this know-how into a know-that, and to reconstruct the explicit rules governing the implicit knowledge of competent speakers and actors. Such a universal-pragmatic reconstruction of communicative action reveals the four validity claims to be counterfactually assumed.

Such a reconstruction also shows that once these validity claims are no longer taken for granted, consensus can only be established via discourses aimed at the redemption of truth and rightness claims. Whereas the earlier justification strategy[3] sought to establish a deductive relation between conditions of argumentative speech and the norms of a communicative ethic, it is now argued that a subject who engages in argumentation presupposes a certain moral principle which he or she can only deny at the risk of a performative self-contradiction (Habermas 1990: 88-9). The following points provide a broad sketch of the essentials of Habermas' program of communicative ethics (Benhabib 1986: 295-6).

First, the concept of an 'ideal speech situation' serves to summarise the rules of argumentation to yield a rational consensus. Habermas insists on his original claim that in argumentative exchange, the structures of a speech situation, especially immunised against repression and inequality, show themselves.

Second, the motivating principle of the discourses lies in the fact that the self-explanatory character of the life-world requires clarification and mutual interpretation. Thus the rational kernel of discourses lies in questioning, puzzling, explaining, interpreting, negotiating and clarifying the everyday interactions in the social life-world.

Third, the task of universal pragmatics is to reconstruct the presuppositions of communicative action and argumentation. Reconstructions continue the task of transcendental philosophy with altered means. Discourse ethics is a "reconstructive science." The program implies a special division of labour between moral theory and moral psychology. Discursive ethics conceives of moral theory as a theory of moral argumentation. In doing so, it finds indirect support from the findings of

developmental psychology, which provide rational reconstructions of the pretheoretical knowledge of competent subjects.

Fourth, the basic principle of a discourse ethic is that only those norms can claim validity which are agreed upon by all the concerned participants.

Fifth, the moral principle which belongs to the presuppositions of argumentation in a discourse ethic is that everyone who participates in a practical discourse must also implicitly and indeed at the cost of getting involved in a performative contradiction, recognise that normative validity claims can only be justified if the following is respected: the consequences and side effects which would foreseeably result from the universal implementation of a controversial norm, and as they would affect the satisfaction of the interests of each single individual, would be accepted by all without compulsion. Habermas names this the principle of universalisability.

Sixth, to complete this account of the program of a communicative ethics, we must identify the object domain of such a theory. The phenomena which need explanation and clarification are the normative (obligating and binding) validity of norms and of validity claims which arise in relation to norm-regulated (or regulated) speech acts.

Like Kant, Habermas describes the moral point of view as a position of impartiality, distinct from other personal and self-interested perspectives. The central task of moral theory thus becomes that of clarifying and justifying what it means to adopt a moral point of view or impartial standpoint. The further

similarities between Habermas' project and Kant's moral theory can be described under four headings (Habermas 1990: 196-7):

First, discourse ethics is a *universalistic* moral theory. Its basic principle (U) is not restricted to a particular time period or culture, nor does it merely represent the moral beliefs of, say, liberal, white, males. Habermas attempts to establish its universality by deriving this principle from the quasi-transcendental presuppositions of argumentation: Anyone who seriously enters into argument with others presupposes certain pragmatic rules from which the principle of universalisability can be inferred. Habermas also argues that the rules and structure of argumentation are not a contingent feature of advanced capitalist societies, but belong to the evolutionary development of species-wide competencies. In this sense, the claim to universality of discourse ethics is defended by means of a revised transcendental argument. Still it rejects the stronger claim to *a priori* status; it is fallibilistic and depends upon the empirical validity of its reconstruction of specific human competencies.

Second, discourse ethics is *formalistic* moral theory. Like Kant's categorical imperative, the principle of universalisability is introduced as a procedure for testing norms. It does not by itself generate any substantive moral norms; rather, it specifies an argumentative procedure that any norm must satisfy if it is to be morally acceptable. Since Hegel's critique of Kant, the distinction between a formal principle and normative content has been difficult to defend and I will consider it in more detail below. Still, in another sense, Hegel's critique missed its target since the categorical imperative was not intended by Kant to generate moral norms, but rather to specify a test procedure for existing maxims of action. Similarly, discourse ethics assumes the existence of a social world of norms; the validity of

these norms can only be established by way of real discourses actually taking place. To expect that substantial moral norms can be immediately inferred from a clarification of the moral point of view (in the way that Rawls hopes to derive his two principles) is, according to Habermas, to ask for too much from a moral theory.

Third, discourse ethics is a *cognitivist* moral theory. This position is shared by Kant, Rawls and Habermas. Cognitivism maintains that in moral discourse we can critically assess the validity of norms by reason and argument. Voluntarism, by contrast, is the view that moral norms and judgments are based in preferences or expressions of the will that are beyond the reach of rational assessment. Of course, preferences can be assessed in terms of their compatibility or incompatibility with other preferences, but our basic preferences cannot be rationally criticised; they can at most be decided upon. According to Hobbesian contractarians, like Buchanan, morality is the search for a compromise or mutual advantage between competing preferences; preferences themselves, however, are generally accepted as 'given'. Habermas attempts to preserve Kantian connection between reason and morality by linking the acceptability of norms to the validity claims. Norms no less than assertions can be contested, and the validity of both depends upon their capacity to be redeemed through discursive argumentation.

Baynes (1992: 1-5) interprets Habermas' discourse ethics as mediating between the extremes of foundationalism and relativism by developing what can be called a Kantian constructivist defense of the grounds of social criticism. Constructivism parts with stronger foundationalist projects which assume that the grounds of social criticism must consist of self evident principles that are absolute and immune to revision. It also parts with those foundationalist strategies that seek to establish an objective standpoint by appealing to a morally neutral notion of

rational self interest. Kantian constructivism attempts to account for the objectivity of our normative assessments by relating ideals and principles employed in our critical practices to an expressly normative conception of practical reason or, what can be argued, to a conception of ourselves as free and equal moral persons. There is no further claim that these principles or ideals correspond to a prior moral order that exists independently of this conception of practical reason or that these principles cannot be revised in the light of subsequent criticism and renewed argument. Rather, the justification of this normative ground is ultimately reflexive or recursive in the sense that there can be no higher appeal to something beyond the idea of that to which free and equal persons can rationally agree.

Habermas' claim is that an analysis and reconstruction of the conditions of mutual understanding can provide a normative foundation for social criticism. In this sense, he pursues by different means Kant's attempt to develop normative principles for criticism from a notion of practical communicative reason.

Finally, discourse ethics is a *deontological* moral theory. The basic moral principle must be specified in a way that does not presuppose a specific conception of the good life since that would amount to a disregard for the ethical pluralism that characterises modern societies. The distinction is defended in terms of a notion of practical reason or moral autonomy. Discourse ethics differs from Kant in several important respects. Habermas rejects Kant's two-world metaphysics and seeks to reconstruct a Kantian perspective freed from its assumptions. For Habermas, this especially means an attempt to reconstruct the notions of reason and autonomy in such a way that they do not presuppose a distinction between a noumenal and a phenomenal world, or between a transcendental and an empirical ego. Reason no longer stands in sharp opposition to needs and interests; rather, reason is defined procedurally in terms of the structure of argumentation and process of

communication and the question becomes what interpretations of needs can best withstand discursive vindication. Similarly, autonomy no longer signals the exclusion or repression of interests and desires, but rather the capacity to reflect critically upon them and the willingness to redeem them discursively if their interpretation is contested.

Discourse ethics abandons Kant's monological version of the categorical imperative in favour of an intersubjective or communicative version of the principle of universalisability. Kant's categorical imperative - especially the "kingdom of ends" formulation - already has an intersubjective dimension to it. It seems clear that Kant is only able to equate what one person can consistently and rationally will with what everyone could consistently and rationally agree to because of his two world metaphysics. Only because interests, desires, and inclinations are set over against reason and purged from the "Kingdom of Ends" can Kant assume a harmony between the individual and the collective rational will. In discourse ethics, by contrast, such simulated 'thought experiments' should ideally be replaced by actual practical discourses.

Kant formulated the rational procedure to be employed by judgment in logic, science, aesthetics and ethics. One could interpret categorical imperative either as a procedure for generating obligatory moral principles or for testing existing ones. According to Kant, the general rules to be followed in thinking consistently are:

1. To think for oneself

2. in communication with men, to imagine oneself in the place of every other person.

3. always to think in agreement with one self.

Most contemporary neo-Kantian ethical theories take their bearings from this formula, and define "universalisability" as a procedure that would entail considering the viewpoint of all. The "moral point of view" is identified with that perspective which would be acceptable for all. Again, most contemporary ethical theories differ from one another in specifying the procedure which best articulates the moral point of view. John Rawls maintains that his theory of justice as fairness and the procedural device of a collective bargaining game behind a "veil of ignorance" best articulate Kant's view of autonomy.

From the standpoint of communicative ethics, the contemporary theories merely repeat the traditional Kantian mistake. Kant thought that, through solitary reflection, a single rational self could come to define a standpoint which would be acceptable to all qua rational agents. The Kantian moral self was a pure rational agent, identical to all others in this respect. Having abstracted from the differences among concrete moral selves, the Kantian 'thought experiment' in moral theory would proceed by asking what each could, without self-contradiction, consider to be a universal law for all. In attempting to define the consent of the moral point of view, contemporary moral philosophers repeat the same mistake: they assume that the solitary moral thinker can define a relevant moral content for all. In communicative ethics, by contrast, the requirement to articulate the moral point of view amounts to the formula: only those norms can claim validity which meet (or could meet) with the consensus of all concerned as participants in a practical discourse.

From the standpoint of Habermas, Kantian ethics is monological, for it proceeds from the standpoint of the rational person, defined in such a way that differences among concrete selves become quite irrelevant. Communicative ethics defends a dialogical model of moral reasoning, according to which real actors engage in actual processes of deliberation on moral questions. It seems surprising

115

that a discourse or communicative ethic need appeal to the universalisability principle. For, as has been correctly observed, Habermas' discourse model represents a procedural reinterpretation of Kant's categorical imperative: rather than ascribing as valid to all others any maxim that I can will to be a universal law, I must submit my maxim to all others for purposes of discursively testing its claim to universality. The emphasis shifts from what each can will without contradiction to be a general law, to what all can will in agreement to be a universal norm.

The Status and Derivation of the Principle of Universalisability (Habermas 1990: 51-76)

As a rule of argumentation that is constitutive for a practical discourse, the principle of Universalisability (U) may be regarded as a communicative or intersubjective reconstruction of Kant's categorical imperative. It stipulates that a social norm is morally justified only if the foreseeable consequences that would follow from its general observance could be accepted by everyone affected as in their own interest. Presuppositions of argumentation refer to those pragmatic conditions that speakers implicitly assume whenever they enter into serious argumentation. These presuppositions have a transcendental status in the sense that they are unavoidable: anyone who denies them and yet wants to argue seriously involves himself in some form of a performative contradiction. Their unavoidability is thus demonstrated by showing that skeptical critics must presuppose them in the very attempt to deny their existence. Further these presuppositions are of a pragmatic nature. It is not the more narrowly conceived conventional meaning of terms that yields these presuppositions, but rather the connection between the meaning of utterances and the pragmatic conditions of their validity as this was developed in the theory of formal pragmatics.

Habermas' rules represent those formal pragmatic conditions necessary to understanding or communicative agreement. Habermas (1990: 89) offers the following three rules:

1. Every speaker with the competence to speak and act is allowed to take part in a discourse.

2. a) Everyone is allowed to question any assertion whatever.

   b) Everyone is allowed to introduce any assertion whatever into the discourse.

   c) Everyone is allowed to express his attitudes, desires, and needs.

3. No speaker may be prevented, by internal or external coercion, from exercising his rights as laid down in 1 and 2.

Taken together, these rules of argumentation can be seen to represent an "ideal speech situation" or "ideal communication community" and, as a representation of the unavoidable presuppositions of argumentation, this ideal is actually assumed whenever anyone argues seriously. However, this ideal speech situation also functions as a regulative idea that can be used to criticise actual discourses. Habermas defends both of these aspects of the ideal speech situation at once. On the one hand, it is not merely a counterfactual ideal, but a real presupposition of all argumentation that, to a greater or lesser extent, is approximated in discursive practices and embodied in social institutions within the life world. We actually make such idealising suppositions in our discursive practices to the extent that we are open to relevant information, listen to who are concerned, and are ready to be persuaded only by the force of the better argument.

On the other hand, no single discourse can completely fulfil the conditions of an ideal speech situation. Although discourses involve validity claims that transcend the spatial and temporal horizons of their own embodiment, the practice of justifying a claim must always draw upon resources (knowledge, interpretation of needs etc.) *from within the lifeworld* (Habermas 1990: 135). In this sense discourses are islands in a sea of praxis. Moreover, any agreement reached today about the validity of a claim does not rule out the possibility that tomorrow-in the light of a new information, changed social conditions, or reinterpreted need - it will be necessary to enter again into a discourse about the claim's validity. Thus the *transcendent character that belongs to the notion of ideal speech situation does not rule out the immanence (and fallibility) of our discursive practices* (Baynes 1990: 114).

To turn now to the second step in the justification of a principle of universalisability, Habermas does not claim that the principle specified in discourse ethics can be derived solely from an analysis of the presuppositions of argumentation. To arrive at a distinctively moral principle of universalisability, these presuppositions must also be combined with the idea of what it means to justify a norm of action. The rules of argumentation, of course, possess a normative content, but they do not yet constitute a specifically moral principle, where this is understood as defining a category of moral oughtness or obligation.

As Baynes (1990: 114) argues, Habermas insists upon this point for two reasons: First, the rules of argumentation state the necessary conditions for those who want to engage in a particular social practice, namely argumentation. In this sense they are analogous to the rules of a game: they are constitutive for a practice, but it is not impossible to imagine rejecting the practice as a whole. There also does not seem to be any reason why the normative character of these constitutive rules must be interpreted in moral terms.

Second, Habermas associates moral phenomena with communicative action in general, rather than with its more demanding form as argumentation. Our moral intuitions center around the basic ideas of individual well-being and compassion or sympathy for others that may themselves have arisen from a sense of the fragility and vulnerability of our common life. In any case, Habermas seeks to clarify both of these ideas with reference to the supposition of mutual reciprocity that is already contained in communicative interactions and that is so crucial for the formation of both individual and group identities. In this sense the often contrasted notion of autonomy and solidarity have a common root in our communicative interactions, and it is from this model of communicative action that the distinctively moral character of norms and obligations arise. There is thus no attempt to derive a moral obligation from the constraints of rational argumentation alone. Rather the claim is that if a general notion of what it means to justify one's action to others- the notion, that is, of what it means to regulate conflicts of interest in light of the norm of reciprocity implicit in the idea of communicative action- is combined with the more demanding rules gleaned from the analysis of the presuppositions of argumentation, then one can derive a rule of argument constitutive for a practical discourse, namely, Principle U.

Habermas' basic intention in developing the theory of communicative competence was to provide normative-theoretical foundations for social inquiry. Habermas' moral theory is a post conventional one in the sense that there is a clear effort to define moral values and principles which have validity and application apart from the authority of the groups or persons holding these principles, and apart from individual's own identification with these groups.

Benhabib (1986: 300) says that, the more the standpoint of a practical discourse is articulated in theoretical terms, the less is it possible to distinguish between communicative ethics and other rival accounts of the moral point of view

on procedural grounds. According to the theory of practical discourse as well, a certain procedure is presented to moral agents as the "privileged description" corresponding to the moral point of view. This generates a 'dialectic of form and content' in communicative ethics.

Discursive argumentation is the new procedure replacing the universalisability test in Kantian ethics. More precisely, universalisability itself is interpreted discursively, as what "all can will in agreement to be a universal norm". So the question arises as to why then Habermas appeals to 'universalisability principle'?

In his essay *Discourse Ethics: Notes on a Program of Philosophical Justification* (Habermas 1990: 56-76), Habermas claims that the principle of universalisability belongs among the pragmatic presuppositions of argumentative speech, which can only be denied at the risk of performative self-contradiction. It is maintained that the Universalisability principle (U) is not a meta-norm or a substantive norm, but a rule of argumentation which belongs to the very logic of practical discourses. Every valid norm has to fulfill the following condition:

> All affected can accept the consequences and the side-effects its general observance can be anticipated to have for the satisfaction of everyone's interests (Habermas 1990: 65).

The argument which leads to the justification of U is in turn summarised as follows:

> Every person who accepts the universal and necessary communicative presuppositions of argumentative speech, and who knows what it means to justify a norm of action must implicitly presuppose as valid the principle of universalisation (Habermas 1990: 86).

3.2 Discourse ethics or Contract: Modes of Agreement

Given this interpretation of discourse ethics, it is interesting to contrast with contractarian model articulated by Buchanan and Tullock (1962). Buchanan and Tullock develop what is called contractarian public choice. It is aimed at devising procedures for arriving at an agreement.

Buchanan and Vanberg (1989) contrast two interpretations of the role of agreement in politics: a 'social contract notion' and a 'dialogue notion'. It is argued that the two notions can be seen as complementing each other if one explicitly separates two components in human choice that in rational choice theory are often inseparably blended in the concept of preferences - an interest component and a theory component. Buchanan argues that the contractarian agreement notion primarily focuses on the interest-component and the dialogue notion on the theory-component in constitutional choice. For the present purposes, I shall concentrate exclusively on Buchanan's version of contractarianism, the reason being that his version of contractarianism employs a model of individual as *homo economicus* as opposed to Rawls' Kantian notion of a person.

The notion of agreement is the fundamental principle by which the legitimacy of a community's basic constitutional order is assured and it is widely shared across a broad range of otherwise quite different intellectual traditions. But it takes different meanings in different theories and there exist systematically different interpretations as to what kind of agreement actually carries with it legitimising force.

The distinction between the two organised components of choice, the interest component and the theory component is of particular importance for the study of constitutional choice in rational choice theories. In the rational choice tradition, the standard interpretation of choice behaviour is in terms of preferences

and constraints. Preferences are considered to be reflected in a utility function over which a rational actor maximises within certain constraints. As commonly understood, the concept of preference is purely about subjective values. It refers to an actor's evaluations of potential objects of choice. However, the concept typically has more than just an evaluative dimension. It is typically used in a way that blends *evaluative* and *cognitive* components, or to put differently blends person's evaluations of - or *interests* in - potential outcomes of choice and his *theories* about the world, in particular his theories about what these outcomes are likely to be.

Buchanan and Vanberg (1989: 50) argue although it may be convenient to use the notion of preferences in a way that does not explicitly separate the genuinely "evaluative" components from the "cognitive" components, for some analytical purposes it may be useful to separate interests and theories as distinct elements in the choice process. How a person chooses among potential alternatives is not only a matter of 'what he wants' but also of 'what he believes', and for some kinds of choices an actor's beliefs or theories may play a most crucial role. Buchanan and Vanberg (1989: 51) suggests that the second element is particularly important for constitutional choices, that is, for choices among rules. It follows that constitutional analysis can profit from explicitly distinguishing between the evaluative and the cognitive dimension of constitutional choice.

In any choice, there is an interest-component and a theory-component, from the choice among ice-cream flavours to the choice of a dress, though, of course the relative importance of the two components may dramatically vary over different categories.

> On the constitutional level, it should be obvious that people's theories about the working properties of alternative rules and rule-systems, and not just their interests in expected outcomes, are of crucial relevance to their choice behaviour (Buchanan and Vanberg 1989: 51).

Constitutional choices are concerned with the choice of rules for a community or group. By their very nature the rules that are to be chosen are 'public' in the sense that they define the terms under which actions and transactions within the respective community may be carried out. Rules constrain the actions of everybody in the community, though not necessarily everybody in the same way. In addition to their publicness, constitutional choices are clearly of an instrumental nature. The term 'constitutional preferences' refers to a person's preferences over alternative rules or sets of rules. Buchanan and Vanberg argue that

> constitutional preferences can be analytically decomposed into two components: constitutional theories and constitutional interests. A person's constitutional theories are about matters of fact. They are his predictions about what the factual outcomes of alternative rules will be. These predictions may be arranged in a true or false, correct or incorrect scalar. His constitutional interests, on the other hand, are his own, subjective evaluations of expected outcomes, evaluations to which attributes like true or false, correct or incorrect cannot be meaningfully applied (Buchanan and Vanberg 1989: 52).

The 'cognitive' and 'evaluative' components of a person's constitutional preferences are critically different in this regard, and so they argue that the question of how constitutional agreement may be reached raises different issues with regard to 'constitutional interests' as opposed to 'constitutional theories'.

The contractarian notion of agreement focuses attention on the interest component in constitutional choice. Social contract theories typically concern themselves with the issue of how agreement on rules can be achieved among persons with potentially conflicting constitutional interests. One characteristic way in how a social contract theory may approach this issue is paradigmatically exemplified by John Rawls' *A Theory of Justice* (1971). In Rawls' construction, the prospect of agreement is secured by defining certain ideal conditions under which

constitutional choices are hypothetically made. The choosers are assumed to be placed behind a 'veil of ignorance', that makes it impossible for them to know anything specific about how they will be personally affected by alternative rules. Ignorant about their prospective specific interests in particular outcomes, they are induced to judge rules 'impartially'. Potential conflict in constitutional interests is not eliminated, but the veil of ignorance transforms potential interpersonal conflicts into intrapersonal ones. It is important to mention a second essential feature of the Rawlsian construction. While the persons behind the veil of ignorance are assumed to be totally ignorant about their prospective specific interests in particular rules, they are, at the same time, assumed to be perfectly knowledgeable about the working properties of alternative rules. So their constitutional theories are supposed to be perfect and non-controversial. Informational problems with regard to the general workings of rules do exist.

Buchanan's approach differs from that of Rawls. A standard objection against the Rawlsian type of contractarianism is that the conceptual reconstruction of some hypothetical agreement under ideal conditions carried little normative and explanatory significance with regard to actual constitutional choices that are made in a world where people are neither totally ignorant about their identifiable constitutional interests nor perfectly knowledgeable as far as their constitutional theories are concerned. Buchanan's interest is in part about how the social contract notion may be employed in the analysis of constitutional choices that occur under much more realistic conditions.

Buchanan claims that his efforts are much less ambitious than that of Rawls. Rawls identifies the principles of justice that he predicts to emerge from his idealised contractual setting. These principles may become the basis of proposals for specific institutional changes, which may then be debated in the pragmatically oriented arena of day to day politics. By contrast Buchanan does not identify any

set of principles that might be used as fundamental. Buchanan advocates in Gray's (1993: 47) words 'indeterminate contractarianism'. His contractarianism is 'indeterministic' insofar as it does not prescribe any a priori principles of justice (Gray 1993: 48-9).

Both the contractarian and the dialogue notions of agreement imply a procedural criterion - as opposed to an outcome criterion - of legitimacy. From both perspectives the legitimacy of basic constitutional principles is judged not against some predefined ideal system but in terms of the process from which these principles emerge. The normative focus is on the characteristics of the process of constitutional choice and not on characteristics of choice of outcomes as such. Furthermore, in both perspectives a good or proper process is defined as one that assures fairness or impartiality in the rules that emerge.

The difference between the two constructions lies in their somewhat different understanding of the procedural characteristics that are to assure fairness. And it lies in their somewhat different interpretation of 'fairness' itself. To the contractarian notion, the individuals' interests are the basic inputs into the constitutional process. The choice process is expected to reflect these interests, whatever that may be. It is not considered a process in which these interests themselves are judged or rated in anyway. Fairness is considered a matter of the constraints under which constitutional choices are made, not as a matter of the quality of interests that enter into such choices. In particular, it is the voluntariness of choice that, from a contractarian perspective, constitutes the essential prerequisite of fairness. Fairness and voluntariness of agreement are, in a sense, the same. Fairness is not defined independently of that upon which persons voluntarily agree. So agreement is viewed in a social contract dimension as defining what is mutually acceptable to voluntarily choosing persons.

By contrast, within the dialogue or discourse theory framework the notions of agreement and fairness tend to carry a characteristically more 'objectivist' or cognitive meaning in the sense of implying more than just the notion of intersecting or coinciding individual interests. Within this framework, constitutional agreement is not simply - as in the contractarian context - a matter of compromise among separate individual interests. Individuals' interests are not simply viewed as the basic inputs that the process of constitutional agreement is supposed properly to reflect. These interests are themselves to be evaluated and possibly transformed in the process of constitutional discourse. Whether this idea is stated in terms of a distinction between 'brute motivations' and 'refined motivations' - where the former are to be transformed into the latter through 'purging of bias and indoctrination'; or whether it is as Habermas (1983: 78) invokes through the notion of an impartial evaluation of the interests of all who are concerned: what is implied is a critical shift towards a 'truth-judgement' interpretation of constitutional agreement. Stated differently, agreement is viewed as a discovery process, a process by which persons not simply reach a compromise but discover what - in some objective sense - is fair or just.

For the contractualists, the only viable procedure is to search for consensus based on voluntary agreement. Truth, in the final analysis, is tested by agreement. In its purest form agreement is unanimous agreement; if one person dissents from a value proposition then, strictly speaking, it has no truth value what so ever. This thorough going subjectivism would seem to put a straight jacket on political and moral argument (as Buchanan himself does not strictly adhere to the principle of unanimity for all issues) since the likelihood of a universal consensus is extremely remote. Yet the concept of agreement cannot be dismissed entirely from the liberal political vocabulary.

Within the contractarian framework, agreement carries normative significance in and by itself. Agreed-on principles are considered legitimate simply because they are the ones that command agreement, not because agreement is indicative of some other quality that distinguishes these principles. Observed agreement may be normatively qualified in terms of its voluntariness, i.e. in terms of the constraints under which the parties involved express their agreement. But, in its contractarian sense, it cannot be meaningfully qualified in terms of a standard that goes beyond agreement itself. The claim to such qualification is apparently inherent in the dialogue construction. According to Habermas (1990), valid or legitimate norms are not simply those on which persons under specified conditions happen to agree, it is those norms that deserve to be intersubjectively acknowledged because they embody some interest that is recognisably common to all persons concerned. And whether they deserve such recognition has to be examined in practical discourse. Constitutional agreement that emerges from ethical discourse has legitimising force not simply because it is agreement, but because it indicates that the agreed-on rules deserve to be classified as equally good for everybody involved.

Buchanan and Vanberg (1989) argue that Habermas's style of reasoning typically leaves considerable room for interpretation, and one might argue that some of his statements can well be read in a way that is much less in contrast to a contractarian conception. The contractarian and the dialogue notion represent alternative and opposing views: agreement as compromise versus agreement as truth-judgement. Buchanan and Vanberg's (1989: 58) interest in Habermas's discourse ethics is not in the authenticity of interpretation. Instead their interest is in exploring the potential for fruitful integration. And for them, the potential clearly exists if the two constructions are interpreted in the context of the distinction

between constitutional interests and constitutional theories as analytically separable components in constitutional preference and constitutional choice.

Constitutional choice involves both individuals' genuine evaluations of alternative rules as well as predictions about the working properties of such rules. Observed constitutional disagreement may reflect disagreement in either one or both of these components. And the process of reaching agreement can, conceptually, be discussed in different terms depending on which one of the two components is concerned. Buchanan and Vanberg are of the view that

> to the extent that disagreement over the rules reflects genuine differences of interests, reaching constitutional agreement is clearly a matter of compromise, of finding terms that are acceptable to everybody, and definitely not a matter of 'discovering the truth'. To the extent, that disagreement is a matter of differences in constitutional theories, the process of reaching agreement clearly is about truth-judgements and it can be properly compared to scientific discourse, to controversy over alternative theories in science (Buchanan and Vanberg 1989: 59).

Though "interest" component and "theory" component are conceptually separable inputs into constitutional preferences, there is no practical way of strictly separating them in actual constitutional choice. One cannot know with any degree of certainty whether, or to what extent, observed disagreement reflects 'merely' differences in constitutional theories, or whether they are based on genuinely different evaluations. In this sense, acknowledgement of the legitimate role of potentially conflicting individual evaluations in constitutional choice clearly excludes an interpretation that assigns, as Habermas does, a truth-judgement function to the constitutional process. Failure to reach agreement in actual constitutional discourse may, but not necessarily, reflect disagreement in theories. Even with perfect agreement in the theoretical dimension, potential for disagreement in evaluations may persist. And respect for such potential

disagreement commands that the limits of 'discourse' and the ultimate role and the need for compromise in constitutional matters be recognised.

So for Buchanan and Vanberg (1989), within the limits specified above, there is an obvious role for 'reason' in constitutional choice. Compared to ordinary market-choice, there seems to be dramatic shift in relevance from the interest component to the theory component when choices among alternative rules are concerned. Persons' preferences over alternatives rules or systems of rules do not simply reflect 'basic values', they are largely a product of their constitutional theories, and, therefore, may be changed through information that impacts on their theories. To the extent that persons' revealed constitutional preferences are informed by their predictions of the working properties of the rules that are under consideration, constitutional agreement can be facilitated by a process that systematically encourages critical examination and discussion of alternative theoretical constructions, separate from and independent of any procedural devices that aim at facilitating agreement in the interest dimension.

The dialogue or discourse notion can be fruitfully interpreted as drawing attention to the importance of the informational dimension choice in constitutional choice, even though advocates might not agree to such limited interpretation. Rational actors to whom efficiency as well as fairness are relevant attributes of a constitutional contract have reasons to be concerned not only about the interest dimension but also about the theory dimension in constitutional choice. Both concerns have certain implications for the kinds of procedural constraints that can be expected to facilitate actual agreement, implications that need not be in perfect accordance.

So far as the interest dimension is concerned, agreement is facilitated by whatever increases persons' uncertainty about the particular effects that alternative rules can be expected to have on them. In the interest of facilitating agreement,

rational actors may, therefore deliberately choose to increase uncertainty. So far as the theory dimension is concerned, prospects of agreement on desirable rules are enhanced not by creating uncertainty but, on the contrary, by raising the level of mutually shared information on the general working properties of alternative rules.

The fact that the public discourse on rules - as political debates in general - is typically carried out by reasoning arguments but can be supposed to be motivated by interests, is sometimes taken as evidence that the political rhetoric is mere camouflage, concealing real interests. And there is a tendency - e.g. in parts of the rentseeking literature - to conclude from this that it is ultimately only interests and the power behind interests that count in the political process, while arguments and reason lack any power 'of their own'. Such interpretation disregards the relevance of the genuine theoretical component in all political and in all constitutional preferences. And it ignores the fact that winning the support for one's own visions and theories is an important part of the political process. The very fact that political discourse is carried out in terms of reasoning argument rather than simple declaration of interests, by itself imposes certain constraints on how one may seek support for one's own proposals.

To conclude, to the extent that persons share a common subset of non-conflicting interests that can be met instrumentally through constitutional rules, the argument over alternative rules reduces to argument over theories. If the interests cannot be factored down into a commonly shared set, while, at the same time, the rules must be public in the sense of imposing constraints on all members of the community, divergent interests must be reconciled. At this level, the function of discussion, dialogue, reason, cannot be to generate agreement on the 'correctness' of alternative theories. Cooperation can replace conflict only if the differing interests, held with varying intensities by persons, can be traded off or compromised, actually or symbolically, in a social contract.

But Buchanan and Vanberg fail to recognise that Habermas himself explicitly notes the critical difference between his discourse notion and a compromise notion of agreement, emphasising the cognitive claims of his own construction (Habermas 1975a: 111-17). He characterises his discourse construction as being critically dependent on the assumption that claims concerning the validity of norms carry a *cognitive* meaning and can be treated like truth-judgements. And he strongly rejects the skeptical premise that the validity of norms cannot be analogous to the validity of truth-judgements.

I shall argue in the next section that what distinguishes Habermas' theory from the contractarians is that discourse ethics cannot be merely seen as procedural theory but is a moral-transformatory theory having practical consequences.

3.3 Interpreting Discourse Ethics

In this section, I shall discuss Benhabib's interpretation of discourse ethics mainly to highlight the significant differences between contract and discourse ethics. In her view, discourse ethics has to be seen as a "moral-transformatory" idea entailing a view of democratic-participatory public life rather than purely a "procedural" idea claiming for "legal-juridical" conception of public life. The significance of the separation of legal-juridical and participatory-democratic conceptions of public life is made precisely to distance discourse ethics from the contractarian approach (Moon 1991: 217-222).

Discourse ethics relates specifically to the moralisation of the social world that occurs when the normativity of existing institutions is brought into question. Discourses are very specialised modes of argumentation. They derive their particular normative force from the fact that individuals are willing to settle a

controversial and conflictual matter without recourse to force, violence, false compromise, or silent acquiescence.

> Discourse ethics requires that controversies over the validity of contested norms be settled through an argumentative process in which the consensus of all concerned decides upon the legitimacy of the controversial norm (Benhabib 1986: 315).

For such a consensus to emerge, Benhabib argues, practical discourse would have to be viewed as a moral-transformative process. If the origins of the concept of a discourse free from domination lie in the search for a democratic public ethos in late-capitalist societies, then it should hardly be surprising that relations of justice would be defined as the privileged object domain of this moral theory. Benhabib argues that there is a significant distinction between legalistic or juridical conception of public life and a democratic-participatory ethos, and the theory of communicative ethics sits uneasily between these two versions. This can be seen, as she says, with a clarification of the concept of 'interests' in communicative ethics.

Discursive argumentation is viewed as being continuous with everyday life-contexts in which real actors contend with one another over the validity of norms. The interests that participants in a discourse bring with them to the argumentation situation are ones that they already have as actors in the life-world. It can be assumed then that in the formula of 'universalisability', the concept of 'interests' is used in a sense that is continuous with such everyday life-world definitions and interpretations. Participation in discourses implies no stipulations regarding the ordinary, everyday understanding of interests.

If participants in discourses bring with them their own interpretation of their own interests, then the question immediately suggests itself: given that the

satisfaction of the interests of each is to be viewed as a legitimate and reasonable criterion in establishing the universality of norms, then is it not the case that universality can only result when a corresponding "compatibility" or even harmony of interests really exists in the life-world?

Consensual agreement reached in discourses would imply the existence of a harmony of interests in the life-world. Now if one were to assume that such a harmony of interests does indeed exist, then it is hard to see why the need for discourses would arise in the first place, since discourses aim to settle validity claims the controversial nature of which implies the presence of conflicting interests in the life-world. Yet if no such harmony exists in the life-world, then it is questionable whether discourses can create consensus around such conflictual interests without the participants in them coming to change their interpretation of their own interests, and perhaps even the life-forms in which they are anchored. If participants in a discourse have conflicting interests, then they could reach consensual agreement either by foregoing some of these interests or by changing the life-forms that generate them.

Discourse ethics is a procedure for distinguishing universalisable from non-universalisable interests (Benhabib 1986: 311). Norms are legitimate only when they correspond to the universalisable or general interest. This claim is also made in the contractarian tradition.

There are three ways of interpreting the concept of general interest.

1. In the contractarian paradigm the general interest minimally means "not taking interest in each other's interests."

2. One can also interpret the "general interest" as describing not merely a set of procedures but an actual social situation in which conflict of interests among individuals disappears. This would imply the reconciliation of the individual and the universal, the private and the public, the empirical and the rational.

3. Finally, one can regard the concept of "general interest" critically in order to reveal the partial and ideological biases of interests claimed to be universal or general. According to this third model, the very concept of a general interest would signify a test procedure.

Since *Legitimation Crisis* (Habermas 1975a), Habermas intends discourse or communicative ethics to be interpreted in this third way as providing a critical test for uncovering non-generalisable interests.

> Since all those affected have, in principle, at least the chance to participate in practical deliberation, the "rationality" of the discursively formed will consists in the fact that the reciprocal behavioural expectations raised to normative status afford validity to a common interest ascertained without deception. The interest is common because the constraint-free consensus permits only what all can want; it is free of deception because even the interpretation of needs in which each individual must be able to recognise what he wants becomes the object of discursive will-formation (Habermas 1975a: 108).

The latter half of this claim could be interpreted in two ways. Either it means that when individuals stop deceiving themselves and others, and discover what their "true" needs are, they will discover them to be identical with those of others, or at least in harmony with them. Or it could mean that through discourses individuals come to realise a certain truth about their needs and interests and change their previously held beliefs about them. The first model is uncomfortably reminiscent of Rousseau's argument that if only each were to look inside one's heart and listen to its voice, he or she would discover the true "general interest."

According to the second interpretation, however, discourses would have to be viewed as moral transformative processes.

This "moral-transformative" moment of practical discourse would serve to distinguish the Habermasian model of the "suppressed generalisable interest" from the position of contractarians. Benhabib argues that without the explicit acknowledgement of this dimension, this model cannot be distinguished from others at all. Habermas has explicitly distanced himself from identifying practical discourse with any kind of collective bargaining or negotiating model.

> A compromise can be justified as a compromise only if both conditions are met; a balance of power among the parties involved and the non-generalisability of the negotiated interests exist. If even one of these general conditions of compromise formation is not fulfilled, we are dealing with a pseudo-compromise. In complex societies pseudo-compromises are an important form of legitimation (Habermas 1975a: 112).

In the light of this argument, the contractarian theories appear as a model of pseudo-compromise on the general interest for they fail to meet the second condition. The contention is that the discourse model would fall into 'legalism' if its moral transformatory aspect, entailing not compromise on interests but the real transformation of certain interests, were underplayed, and the theory were presented as a more refined version of a 'universalisability formula' in neo-Kantian ethics (Benhabib 1986: 313-315).

If Habermas assumes that, in discourses individuals preserve the same need and interest interpretations as they had in ordinary, everyday contexts, then a consensual acceptance of norms reflecting the general interests of each can hardly follow. For this to be the case, either one must assume that a preestablished harmony of interests exists - clearly an unacceptable premise for a critical social theory - or one must interpret this process minimalistically as aiming at the

establishment of the lowest common interest, while leaving substantive conflict of interests untouched. This could consist in the formation of a legal system, a constitution embodying the two principles of justice and like. The content of a universalist ethical position would then be exhausted by a legalistic or juridical construction. But if neither alternatives is accepted, then one must assume that discourses are processes through which new needs and interests emerge which could lead to a consensus among participants.

Benhabib argues that needs and interests cannot be taken as given but must be seen as a result of the processes of socialisation and the life experiences of individuals in particular social formations (Moon 1991: 217). Discourse should not be confined to finding rules which bridge the differences among given constellation of interests and identities but should call into question the conflicting interests, needs and aspirations which lead to conflict. Habermas himself argues that in discourses actors share as a background a lifeworld that forms the context for communication while at the same time providing the resources that the actors need to engage in this process.

The moral transformative conception of discourse holds out the possibility that people can come to discover new needs and interests, whose satisfaction for one person is compatible with their satisfaction for others. For Benhabib such a discourse would be oriented not to the "generalised other" but to the "concrete other" - to specific individuals with particular identities and aspirations with whom we would seek to establish a genuine community (Moon 1991: 217). Benhabib (1986: 340) argues that the standpoint of the "generalised other" requires us to view each and every individual as a rational being entitled to the same rights and duties we would want to ascribe to ourselves. In assuming this perspective, we abstract from the individuality and concrete identity of the other. By contrarst, the standpoint of the "concrete other" requires us to view each and every rational

being as an individual with a concrete history, identity, and affective-emotional constitution.

Although the conception of 'subject' is empirically rooted in the learning process of the cultural evolution of modernity, Habermas, at the same time abstracts from it for normative purposes. The purpose of this abstraction is two fold: one is the individual's reflective ability to question the interpretive framework fixed by the cultural tradition. Secondly, such reflective questioning is accompanied by an ability to articulate one's needs linguistically, by an ability to communicate with others about them.

Thus, in both instances, the assumption is that agents adopt a critical reflective distance toward the lifeworld instead of merely adhering to the norms governing everyday life. Communicative ethics shares with Kantian moral theory the emphasis on the role of reason in ethics, and the necessity of viewing rationality as universally binding.

Benhabib's (1986: 298) contention is that communicative ethics runs the risk of falling into a certain "rationalistic fallacy" of the Kantian sort, in that it ignores the contingent, historical, and affective circumstances which made individuals adopt a universalist-ethical standpoint in the first place. A contemporary reformulation of the Hegelian objection to Kantian moral theory as applied to communicative ethics would be: does the cognitivist bias of communicative ethics also lead to the rationalistic fallacy, namely, to a view of reason as a self-generating faculty, determining both the conditions of its own genesis and application? (Benhabib 1986: 317). Such an ideal of the total self-determination of reason is essential to the idealist concept of reflection. Critical theory, by contrast, began with the insight that the total self-reflection of reason amounts to the impossible demand to transcend the human situatedness of reason, but the requirement that reason engage in self-reflection was also justified by

critical theorists, insofar as they argued that reason which is incapable of becoming aware of its own contingent origins and applications in society would turn into an instrument of domination. The rationalistic fallacy arises when the continuing critical self-reflection upon the conditions of the possibility of reason, conditions which reason never wholly determines, is minimised and reason is viewed as self-generative.

In my view, this objection is not a valid one since as discussed in chapter 2, Habermas himself has moved away from the strict Kantian defense of communicative ethics. The ideal speech situation is procedural but this does not mean totally disconnected from the life-world. Proceduralism does not imply formalism and ahistoricism, as it is usually thought to do. There is no reason why a procedural theory cannot be part of a more general theory of self, historical change and social structure. Agents form their opinions and views in the life-world but at the same time are rationally detached in order to adopt a critical reflexive standpoint.

This question is concretised by turning to the beginnings of discourse. Habermas comments:

> Anyone who does not participate, or is not ready to participate in argumentation stands nevertheless "already" in contexts of communicative action. In doing so, he has already naively recognised the validity claims - however counterfactually raised - that are contained in speech acts and which can be redeemed only the communicatively. Otherwise he would have to detach himself from the communicatively established language game of everyday practice (Habermas 1975a: 159).

In *Discourse Ethics: Towards a Justification Program* (Habermas 1990: 43-116) Habermas utilises this argument in the context of the refutation of skepticism. The moral skeptic can indeed refuse argumentation; but 'through

argumentation he cannot indirectly deny that he shares a sociocultural form of life, that he has been raised in contexts of communicative action, and that he reproduces his life in them. In a word, he can deny morality but not the ethical content of lived relations in which he abides, so to speak, day by day. Otherwise he must take flight in suicide or in a deep psychosis. Skepticism is the refusal to share a common way of life.

Habermas admits that the commitment to consider all individuals as potential participants in discourse presupposes a universalistic commitment to the equality, autonomy, and rationality of individuals.

> Neither the willingness nor the ability to consider moral questions from the hypothetical and disinterested perspective of a participant in practical discourse falls from heaven; they result from interests that are formed only under social conditions, as well as from learning processes and experiences that are open to social groups only in certain situations. The interest in rational discourse is itself one which precedes rational discourse, and it is embedded in the contingency of individual life histories and in collective patterns of memory, learning, and experience. This interest can be enabled or frustrated in the life of the individual; just as available patterns of political culture and traditions in society may encourage or hinder the development of discursive rationality (Habermas 1982: 253).

The purpose of TTCA (Habermas 1984) is precisely to make visible those processes of rationalisation of the life-world which constitute the tradition of the moderns for us, and which first make the ideal of discourse possible. It might be questioned whether the specificity of the cultural, political and institutional configurations which would enable the ideal of discourse embodiment can be captured via a general theory of rationalisation which radically distinguishes between the rational reconstruction of logic of development and the narrative unfolding of historical traditions.

Communicative ethics demands from its participants a willingness and an ability to consider normative questions from a universalist standpoint and to regard every being as an equal regardless of the actual constellation of relations in real life. The necessity of discursive argumentation arises when, through conflict and crises, social and political agents challenge an established background consensus. Discourses arise when the intersubjectivity of ethical life is endangered; but the very project of discursive argumentation presupposes the ongoing validity of a reconciled intersubjectivity (Benhabib 1986: 321). For those who feel that the reconciliation in social life has been achieved at their expense, it might be morally justified to refuse participating. This does not imply resorting to violence or to force, but simply the refusal to engage in dialogue since the mutuality of shared existence is indeed endangered by the existing configuration of power.

There are some situations where the nature of the conflict between the parties is such that there can be no dialogue, for the preconditions of dialogue - namely, the mutual recognition of each other as discursive partners - simply do not obtain. Structural inequalities between the parties, such as those pertaining to wealth, power, or status may be such that reciprocal recognition does not exist. Social conflicts may resist being cast into the discourse model of conflict resolution.

This means that communicative ethics is contingent on both ends: the willingness and capacity of individuals and the culture at large to adopt such an ethical standpoint in the first place as well as the moral sagacity and political insight necessary to concretise the principles of such ethics in action.

Conclusion

In this chapter I contrasted two normative theories of collective choice: contractarian public choice and discourse ethics. From the standpoint of Buchanan, both discourse ethics and contract can be complementary if the former limits itself to the theory-component in choice and the latter to the interest-component of constitutional choice.

Unlike Habermas, Buchanan's view is that moral controversies cannot be decided with 'reason' because the value premises from which we infer moral assumptions are irrational. By contrast, I argued, following Benhabib, that Habermas' theory cannot be merely seen as procedural but more importantly should be seen as a moral-transformatory theory.

Notes

[1] There are many different theories of conflict resolution. Within contract theories, there are Hobbesian contractarians like Buchanan (1962) and Kantian contractarians like Rawls (1971). Secondly, liberals like Ackerman (1980) articulate a liberal model of dialogue. And finally, Habermas provides a critical theory of discursive legitimation. In this chapter, I am concerned with Buchanan and Habermas because the former represents public choice theory and the latter critical theory.

[2] Theoretical discourse is the form of argumentation in which controversial truth claims are thematised. Practical discourse is the form of argumentation in which claims to normative rightness are made thematic. Here persons are rational who can justify their actions with reference to existing normative contexts. This is particularly true of those who, in cases of normative conflict, act judiciously, that is, neither give in to their effects nor pursue their immediate interests but are concerned to judge the dispute from a moral point of view and to settle it in a consensual manner. The medium in which we can hypothetically test whether a norm of action, be it actually recognised or not, can be impartially justified is practical discourse (Habermas 1984: 19).

[3] In his earlier work Knowledge and Human Interests (1970b: postscript) Habermas develops the theory of cognitive interests (or knowledge-constitutive interests). The theory of cognitive interests is concerned with uncovering the conditions of the possibility of knowledge. It is Habermas' contention that the human species organises its experience in terms of *a priori* interests, or cognitive interests. That there is a 'basis of interests' follows from an understanding of humans as both labouring and language-using animals: they must produce from nature what is needed for material existence through the manipulation and control of objects and communicate with others through the use of intersubjectively understood symbols within the context for rule-

governed institutions. Besides these two interests, there is also, according to Habermas a third one which is the reflective character of knowledge. This is an interest in reason, in the human capacity to be self- reflective and self determining, to act rationally. Thus the interests are the technical, practical and emancipatory (empirical-analytic; historical-hermeneutic and critical).

Habermas accords to the category of cognitive interests a some what problematic status as quasi-transcendental. Here he takes psychoanalysis, particularly its insights into the significance of self reflection, as the prototype of a critical science and transposes the model to the realm of social analysis and political practice. In his latter work, Habermas redefines critical theory. Instead of constructing an *a priori* based theory of cognitive interests, Habermas argues that critical theory needs to be more empirical.

Chapter 4

The Problem of Cooperation - Strategic Action and Communicative Action

Introduction

While Buchanan and Tullock (1962) and Riker (1982) were concerned with the normative task of aggregating preferences, the problem addressed by both Downs (1957) and Olson (1965) was to use microeconomic tools to explore the provision of public goods and collective action. In this chapter I shall discuss the problem of cooperation in the prisoner's dilemma. Rational choice theorists have explored the problems of cooperation in a game theoretic framework.

As discussed in chapter 1, public choice theories rely on a 'thin' version of instrumental rationality based on the notion of *homo economicus*. This model of rationality conceptualises the structure of social interaction as 'parametric' in that agents have 'fixed' or 'given' preferences. Game theory is essentially a critique of it. In section 4.1, I discuss the prisoner's dilemma problem and the problem of indeterminacy in strategic interaction in game theory and then move on, in section 4.2, to discuss the possible alternatives. And Finally, in section 4.3, I shall show the meeting points for productive conversation between rational choice and critical theory.

4.1   The Problem of Indeterminacy - Limits of Strategic Rationality

4.2   Rational Cooperation - Interests and Social Norms

4.3   Norms and Interests - Strategic Action and Communicative Action

4.1 The Problem of Indeterminacy - Limits of Strategic Rationality

The paradox which public choice theories exposed is the conflict of individual and social rationality. The disanalogy arises in the process of preference aggregation (voting: majority rule) and in the problem of collective action (free-rider's problem) or prisoner's dilemma. The significant contribution of game theorists comes through demonstrating the limits of a particular form of individualism in social science: one based exclusively on the model of persons as *preference satisfiers* principally associated with public choice theory literature (Buchanan and Tullock 1962; Downs 1957; Riker 1982; and Olson 1965). The structure of instrumental rationality used in these theories has been explained in chapter 1. In rational choice theories the individual is modelled as a *homo economicus* who is essentially concerned with his preferences acting in a predominantly parametric environment.

Critiques challenged the methodological assumptions of rational choice theories. In *Pathologies of Rational Choice Theories* (1994), Green and Shapiro provide a critical appraisal that would assess rational choice scholarship on its own terms. They argue that practitioners of rational choice theory operate within the confines of an esoteric technical vocabulary that is seldom understood by anyone else, while and on the other hand, critics tend to ignore the rational choice paradigm without understanding it fully.

Green and Shapiro (1994: 33) identify a list of 'methodological pathologies' in rational choice theories and locate much of the responsibility for this explanatory inadequacy in a "syndrome of fundamental and recurrent methodological failings rooted in its universalist aspirations". They are as follows:

1. Post-hoc theory development

2. Search for confirming evidence

3. Projecting evidence from theory

4. Arbitrary domain restriction

The first of these pathologies is 'post hoc theory development'. Green and Shapiro contend that rational choice theories are "method driven" rather than "problem driven" and that, as a result, rational choice theorists look only for evidence that is consistent with their theory - sometimes to the point of fitting the data to the theory. Rather than formulating bold predictions that are falsifiable by empirical evidence, rational choice theorists tend first to look at empirical evidence, then design a rational choice model that fits it. On their view, rational choice theorists allegedly engage in "arbitrary domain restriction". This means that rational choice theory is applicable wherever the theory seems to work (Green and Shapiro 1994: 45).

Game theorists have been quick enough to recognise these difficulties and attempted to redefine the notion of rationality in explaining social and political interaction.

Let me briefly recapitulate the problem of collective action. The problem of collective action arises when social action is likely to be distorted by the private interests of the agents who are to carry them out. This problem is associated with public goods and explored by Olson (1965) and Downs (1957). The collective action problems are the most pervasive problems in everyday life situations. They fascinate social scientists because they involve interaction where the individual pursuit of what seems rational produced a collectively self-defeating result, in regard to e.g. disarmament, joining a trade union, corruption and why do we stand when we can all sit. The non-coincidence of self-interest behaviour and socially co-operative outcomes has been a key factor in the way game theory has evolved.

Likewise, many times social interactions can also have unintended consequences (Elster 1989b: 94-95). Jean Paul Sartre referred this to 'counterfinality'. For example, when farmers try to get more land by felling trees, they can end up losing land because large-scale deforestation leads to soil erosion. When everyone gets to his feet to get a better view of the game, no one succeeds and all get tired from standing up. But unintended consequences can also make everybody better off. This is Adam Smith's 'invisible hand': the pursuit of self interest serves the common interest. For example, people paint their houses in order to protect them from bad weather, but in doing so they may also offer others benefit of a pleasant sight. The analytical structure of the collective action problem is similar to that of the prisoner's dilemma.

Let me start with the prisoner's dilemma. There are innumerable examples in social life of this perverse tendency of individual rationality generating collective disaster. For example, it is better for all commuters if all go by bus than if all go by car, but for each it is always better to go by car. It is better for everybody if nobody litters in the park, but individuals have no incentive to abstain form littering. Likewise, voting, reporting one's income correctly, supporting public radio stations and joining a revolutionary movement conform to a similar pattern.

The prisoner's dilemma was discovered around 1950 by Merrill Flood and Melvin Desher. Its greatest strength is that it makes the strategic aspects of such interaction explicit. In its simplest form, the prisoner's dilemma involves two players in interaction, each facing two possible strategies. Since each player can choose independently of the other, their two pairs of strategies produce four possible outcomes.

	Cooperate	Defect
Cooperate	1,1	-1,3

Defect	3,-1	0,0

Each outcome is a payoff pair such as the lower left outcome (3, -1) in which the first payoff goes to Row and the second to Column. In a prisoner's dilemma situation, there are two players, Row and Column, each facing two strategy choices, cooperate and defect. If both cooperate, they receive positive payoffs of 1 each. If both defect, they receive payoffs of 0 each. If one cooperates and the other defects, the cooperator receives an even worse pay-off, -1, while the defector does very well with a positive pay-off of 3. There is therefore a strong incentive to defect. Thus defection is the *dominant strategy* for each player: if Row's choice does not influence Column's, then Row is better off defecting *no matter what* Column does. As is clear from the representation, Row gets more in each column from choosing defection than from choosing cooperation. To sum up, both Olson's free-rider problem and the prisoner' dilemma suggest a discouraging perspective on the problems of cooperation and coordination. However, the most dismal aspects of both problems reflect the static nature of the analysis and the fact that it is a one-shot game, i.e., when the prisoner's dilemma game is played only once, it is a dominant strategy for players to defect and therefore not to achieve what would be an efficient outcome with respect to the aggregate well being of the players.

The so-called prisoner's dilemma has been the mainstay of game theory. However, in recent years game theorists have found that defection is not necessarily the dominant strategy if the situation is repeated over and over again. Many social interactions are repeated either with the same person or with people who are drawn from the social group. Since one-off encounters typically occur only between strangers, the analysis of repeated games promises to extend the scope of game theory considerably.

I intend to explicate the abstract reasoning of strategic interaction by way of discussing some of the central ideas in game theory. Game theory is rapidly becoming established as one of the cornerstones of the social sciences. Its claims go beyond the domain of economics representing an opportunity to unify the social sciences by providing a foundation for a rational theory of society. Some argue that

> Game theory may be viewed as a sort of umbrella or unified theory for the rational side of social science....(it) does not use different, ad hoc constructs... it develops methodologies that apply in principle to all interactive situations (Aumann and Hart 1992: 4).

> If one accepts that interaction is the essence of social life, then game theory provides solid micro foundations for the study of social structure and social change (Elster 1983b: 77).

In discussing game theory, I shall confine myself exclusively to four themes: first is the distinction between 'parametric' and 'strategic' actions. Second is the problem of 'indeterminacy'. Third is the importance of the notion of belief-formation in strategic settings. And finally, the notion of preference formation. All four themes illuminate the limitations of rational choice theory.

Game theorists model individuals not as isolated and atomistic but as strategic agents engaged in forming beliefs about each other. The fundamental idea is that game theorists attribute to actors a unique human capacity for strategic behaviour (Elster 1979: 2, 9-18). They populate their models with actors who are strategically competent, who are capable of engaging in *strategic* rather than merely *instrumental or parametric* action. In a *parametric* decision the agent faces *external constraints* that are in some sense *given or parametric*. First he estimates them all and decides what to do. A strategic situation, on the contrary, is characterised by interdependence of decisions. Each agent has to anticipate what others are likely to do, which may require an estimate of what they anticipate that

he will do. His decision enters as *part-determinant* of the constraints that shape his decision. Likewise, game theorists also presume that social actors can 1. understand that their environment partly consists of other 'intentional' beings and 2. recognise those others as 'equally' rational. A game is defined as any interaction between agents that is governed by a set of outcomes for each possible combination of moves.

Because, for each agent, the context of interaction is constituted partly by other intentional agents, it cannot be treated as a parameter. As a result, the apparently straightforward idea of rationality no longer provides solid theoretical moorings. In strategic settings the concept of rationality itself calls out for scrutiny.

Thus a theory of games promises to apply to any social interaction. The key assumptions of game theory are: common knowledge of rationality (CKR) and actors' knowledge the rules of the game. The assumption of common knowledge of rationality implies that players are in agreement about the rules of the game that is being played. Moreover, agents regard each other as rational in the interacting setting. Game theorists demonstrate that strategic rationality can coordinate social and political interaction. The theory of games attempts to determine what *'rational'* can mean when an individual is confronted with the problem of optimal behaviour in games and equivalent situations.

The analytical rigour of game theory is employed to model social interaction in everyday life. As said before, game theorists conceptualise social interaction as consisting of other intentional agents and recognise those others as *equally rational.* This assumption of mutual recognition of rational individuals has enriched the scope of the explanation in social interaction. In a strategic interaction individuals need to take other rational beings whose actions affect each other into consideration. So in an interactive setting 'belief formation' assumes an important role since agents have to form expectations regarding other agents. This means,

your beliefs about others affect their behaviour when those beliefs lead them to expect that you will act in particular ways.

So the question is what is the source of beliefs? (Elster 1983a: 4-6; 16-17) Where do they come from? To put it differently, most game theorists agree on one aspect of the belief formation in the social world: how to update beliefs in the presence of new information? In the absence of evidence, how do agents form probability assessments governing events like the behaviour of others? One answer is that we can use previous experience to generate expectations i.e. experience with other people provides a basis for belief formation.

Beliefs are treated as purely subjective assessments. They are only revealed *ex post* by the choices people make. But the disadvantage of this is that it might license almost any kind of action and so could render the model close to vacuous. So in order to overcome this, game theorists supplement the assumption of instrumental rationality with the assumption of common knowledge of rationality (CKR). The purpose of the latter is to place constraints on people's subjective expectations regarding the actions of others.

Thus expectations regarding what others will do are likely to influence what it is instrumentally rational for you to do. Thus fixing the beliefs that rational agents hold about each other is likely to provide the key to the analysis of rational action in games. If you want to form an expectation about what somebody does, what could be more natural than to model what determines their behaviour and then use the model to predict what they will do in the circumstances that interest you? You could assume that the person is an idiot or a robot or whatever, but most of the time you will be playing games with people who are instrumentally rational like yourself and so it will make sense to model your opponent as instrumentally rational. This is the idea that is built into the game theory to cover how players form expectations.

In game theory, a solution consists of a definite prediction regarding what strategically rational players will do in the situation being modelled. The conception of equilibrium in game theory is influenced greatly by the work of John Nash. A Nash equilibrium occurs if there is a potentially self-reinforcing agreement whereby each actor does what is best for him/her given what others do. *It can be understood intuitively as an agreement from which no party has an incentive to defect* (Johnson 1993: 79-80). Harsanyi defines it more exactly: A given strategy of a certain player is called a best reply to the other players' strategies if it maximises this player's payoff so long as the other players' strategies are kept constant. So a given combination of strategies is called an equilibrium point if every player's strategy is a best reply to all other players' strategies. Thus when people can enter into binding agreements with others, "an equilibrium corresponds to an outcome in which no coalition has the incentive or the means for unilaterally insuring an improvement in the welfare of all of its members".

The concept of Nash equilibrium employs a powerful idea of *'mutual rationality'*. This idea is essentially concerned with *'respecting the mental capacities'* of every individual participating in the game. The presupposition of mutual rationality enriches its search for equilibria in a strategic game. Thus game theorists argue that given CKR and consistently aligned beliefs (CAB) Nash equilibrium is possible. For game theorists, the formulation of social interaction in this way can overcome the disputes surrounding the problems of indeterminacy. The recognition of mutual rationality makes the prediction of individual strategies possible in any game and arrived at a Nash equilibrium.

The Nash equilibrium is attractive because as time goes by and agents adjust their expectations of what others will do in the light of experience, then they will seem naturally drawn to a Nash equilibrium because it is the resting place for beliefs (Heap and Varoufakis 1995: 192). Since many social interactions are

151

repeated either with the same person or with people who are drawn from the same social group and since one-off encounters typically occur only between strangers, the analysis of repeated games promises to extend the scope of game theory considerably. In dynamic settings players learn things about their opponents from the way they have behaved in the past. Such learning can be exploited by players behaving in particular ways to develop 'reputations' for playing the game in particular ways (Heap and Varoufakis 1995: 167).

The strength of the concept of equilibrium lies in its predictive power. What must be balanced is choices of actions i.e., intentions, which are thus analogous to physical forces. Social equilibria occur when actors choose in the most advantageous way, given the choices of others, and reach an outcome they would not wish to depart from. That is, they would not wish to have chosen differently because the outcome reached is the best they can achieve under the circumstances.

If a single equilibrium point exists for a given configuration of actors' preferences and a set of institutional rules, then it is possible to derive predictive hypotheses about what actual agents will do, assuming that people behave rationally. If there are many possible equilibria, then rational choice models become more indeterminate; if there are no equilibria, then the political world threatens to be chaotic and inherently unpredictable in its basic structure. Thus predictions made using Nash equilibrium concept can be either non-existent or indeterminate. That is why so much of the theoretical rational choice literature revolves around trying to identify the necessary and sufficient conditions for the existence of equilibria.

The thinking behind the Nash equilibrium is quite brilliant. It cuts through the knot of webs of beliefs and arrives at a simple respect of everyone's mental capacities. Nash argues that because players are rational and respect each other's

rationality, they are naturally drawn to the Nash equilibrium since the latter is the only one that respects equally every one's rationality. Internal consistency of each player, as assumed by rational choice theorists, is not enough when the game is played under CKR, mutual respect requires that beliefs should also be mutually consistent. It means that no instrumentally rational person can expect another similarly rational person who has the same information to develop different thought processes (Heap and Varoufakis 1995: 25). If a person knows your plans he would not want to change the beliefs he holds about you and which support his own planned actions. Players must also form the same probabilistic assessment of what is likely to happen when they go to work with the same information. When this is made clear, it will touch on the matters of *uncertainty* (Elster 1989a: 4-10). Nash is undeniably at the heart of game theory and the existence of multiple Nash equilibria in many games has set an agenda of refining Nash.

The point of the refinement project is to reduce the number of Nash equilibria where possible so that prediction of Nash is not rendered vacuous by the presence of multiple equilibria. Game theorists explored in various ways to overcome the problem of multiple Nash equilibria but were unsuccessful. Let me bring in together some of the arguments which have surfaced over the Nash Equilibrium. Firstly, it is not clear that the consistent alignment of beliefs (CAB), which is necessary for Nash, can be justified by appeals to the assumptions of rationality CKR. Something else seems to be required and the best game theory has come up with so far is the Harsanyi doctrine (and its defense by Robert Aumann). This has the effect of making rational players believe that there is a unique rational way to play a game because rational players must draw the same inferences from the same information. Once this is conceded then it follows from the assumptions of instrumental rationality and CKR that the way to play must constitute a Nash equilibrium. It is the status of the Harsanyi-Aumann argument which is in dispute.

According to Harsanyi, the argument is that when two rational individuals have the same information, they must draw the same inferences and come, independently, to the same conclusion (Heap and Varoufakis 1995: 58). This would still leave it open for different agents to entertain different expectations since it only requires that rational agents draw the same inferences from the same information but they need *not* enjoy the *same information*. Robert Aumann, in defence of Harsanyi, discounts this possibility stating that two rational players could not 'agree to disagree' in such a manner, since the moment rational agents discover that they are holding inconsistent expectations each has a reason to revise their beliefs until they converge and become consistent. Thus the Harsanyi-Aumann argument implies that rational agents, when faced by the same information with respect to the game should hold the same beliefs about how the game will be played by rational agents.

Here again it is difficult to see Aumann's defence without a problem. Such a defence of CAB is plausible only when the idea of an explicit 'dialogue' in real (i.e. historical) time is considered. But Aumann does not appeal to any notion of dialogue and without such a process there cannot be any agreement. Following Habermas who thought that agreement can be arrived at through dialogue, we assume that an opposition of incompatible positions will give way to consensus acceptable to both sides once differences are resolved in communication. This would seem to create a problem for Aumann's argument at least as far as one shot games are concerned, that is, interactions which occur between the game once and then you might discover *ex post* that you must have been holding some divergent expectations.

To sum up, whatever the case, the basic method of game theory is to argue that individuals try to predict what others will do in reply to their own actions. Thus game theory provides a particular although not uncontentious perspective on

social action: individuals are both optimisers and perfectly rational. Whether this view of social interaction is descriptive of how people actually *do* behave, or is prescriptive of how rational people *would,* or even *should* behave, is interesting. Despite its tremendous success in applying the model of calculating rationality in social interactions predicting human behaviour in economic, political, social realms, it still has serious failings (Lyons: 94). Although Nash equilibrium is championed by game theorists as the most effective tool to analyse situations of cooperative outcome, it still needs to confront the question of 'equilibrium selection'.

An equilibrium point is only a necessary - not a sufficient - condition for the existence of a solution. This is because a game might admit either of no equilibrium or several (Elster 1989a: 7-17). The problem is best illustrated in Elster's work. His works (Elster 1983a) and (Elster 1989a) are an internal critique of instrumental reason. Elster argues that rational choice theories rely on intentional explanations. Briefly put, an intentional explanation has the following structure. Elster argues that the thick theory of rational action can be defined as the choice of actions which *best* satisfies a person's preferences (Elster 1986a: 1-18). He describes rational action as involving three *'optimising'* operations:

1. finding the best action, for given beliefs and desires;
2. forming the best grounded belief, for given evidence;
3. collecting the right amount of evidence, for given desires and beliefs.

Ideally, then, a rational choice explanation of an action would satisfy three sets of requirements. First, there are *optimality* conditions which underlies the principle of utility maximisation. Second, there is a set of *consistency* conditions. Finally, there is a set of *causal* conditions which means the action must not only be rationalised by the desire and the belief; it must also be caused by them and

moreover, caused 'in the right way'. In all this, the elementary unit of explanation is that individual action is guided by some intention. Rationality, then, is understood as a variety of intentionality. For something to be rational, it has to be within the scope of conscious, deliberate action or reflection. Intentional explanation forms the basis of rational choice explanation and it is also the feature that distinguishes the social sciences from the natural sciences. Explanation in terms of *optimisation* remains the paradigm case of intentional explanation in social sciences. To put it in a nutshell, instrumental rationality equates the rational action with the choice of the means most likely to satisfy a given set.

Now given this sketch of rational action, Elster argues that it can go wrong at all three different levels and in each case the failure may be due to either *'indeterminacy'* or *'irrationality'* (Elster 1989a: 5). To the extent that rationality fails to give unambiguous prescriptions, it is indeterminate and to the extent that people fail to follow its prescriptions - that they behave irrationally - the theory is inadequate. What distinguishes Elster's critique from game theorists is that the latter only recognise the problem of 'indeterminacy' at the level of actions. Elucidating the triadic relation of beliefs and desires; evidence and action, Elster gives a powerful exposition of the nature of preference formation.

The problem of indeterminacy in rational choice theory arises in two forms: one is the problem of *multiple optima* and the other is *non-existence* of rational choice. This problem is best illustrated in non co-operative game theory. And game theory provides scant assistance in coming to terms with the indeterminacy it generates. The implication is that there are a number of equilibria on which players might potentially converge. With respect to belief formation, Elster argues that 'uncertainty' and 'strategic interaction' create problems. Uncertainty is defined as 'radical ignorance', the lack of ability to assign numerical probabilities to the possible outcomes associated with the various options. Next consider strategic

interaction as an obstacle to rational belief formation. Rational choice requires beliefs about choices to be made by other people. These beliefs, to be rational, must take account of the fact that these others are similarly forming beliefs about oneself and about each other. In the absence of enforcement and commitment devices, there is no way in which a player can form a rational belief about what the other will do. Problems of indeterminacy also arise with respect to the optimal amount of information one should collect before forming an opinion.

The central argument is that rationality itself requires us to recognise the limitation of our rational powers, and that the belief in the omnipotence of reason is just another form of irrationality. Similarly, the notion of preference formation includes beliefs, desires and actions. Individual rationality based on mere 'given' preferences is a fragile foundation for rational choice theory. Preferences can be distorted and individual rationality itself can run into problems. There are any number ways for preferences to be distorted. For example, weakness of will and excess of will, cognitive dissonance reduction, adaptive preference formation, preference reversals, intrapersonal problems of preference aggregation and other forms of cognitive compartmentalisation like self-deception and wishful thinking (Elster 1983a; 89a 89b).

By conceptualising interaction in strategic terms, rational choice theorists appear to capture some of the elemental features of social interactions. In other words, how much can be said about the outcome of games will tell us much about how much of the social world can be explained in instrumentally rational, individualist terms. As discussed earlier, in prisoner's dilemma game the stark prediction is 'mutual defection'. The existence of this type of interaction together with the inference that both will arm has provided one of the strongest arguments for the creation of the state. The aim of rational choice theorists is to study the ability of human beings to devise co-operative solutions to resolve problems in a

decentralised settings without the intervention of a coercive state. The investigation of this problem - how is spontaneous order possible - is sometimes referred to as the Hayekian programme. Attempts to implement the Hayek programme include Taylor (1987); Axelrod (1984).

So in an iterated prisoner's dilemma game, one that is repeated, there is no dominant strategy. In a famous tournament, Robert Axelrod found that the winning strategy under these conditions of continuous repeated play is a strategy of tit-for-tat, one in which a player responds in kind to the action of the other player. This is the central theme of Axelrod's celebrated work *The Evolution of Cooperation* (1984).

One of the main reasons for looking at repeated games is to explore the intuition of linking repetition with co-operation. In repeated games one alternative is to adopt a 'tit-for-tat' strategy rationally. Recall that the problem in one-shot games arose because the obvious remedy of making an agreement to co-operate failed in the absence of an enforcement mechanism. The point of repetition is that it allows the players themselves to enforce an agreement. Players are able to do this by being able to threaten to punish their opponents in future plays of the game if they transgress now. The tit-for-tat strategy (Axelrod tournament) embodies precisely this type of behaviour. It offers implicitly to cooperate by co-operating first and it enforces co-operation by threatening to punish an opponent who defects on co-operation by defecting until that person co-operates again. To put it differently, playing this strategy allows your opponent to develop a reputation for co-operation simply by playing cooperatively in the previous play of the game. Co-operation is possible if the interaction is provided sufficiently often to make the long term benefits outweigh the short-term gains. Axelrod invited professional game theorists to enter programs for playing a computer based repeated, round robin, version of the prisoner's dilemma game. Under the tournament rules, each

entrant (program) was paired with another once in a random ordering, and in each of these contests the game was repeated 200 times. In fact 14 people responded and the round robin tournament was actually played five times to produce an average score for each program. Tit-for-tat, submitted by Anatol Rapoport won the tournament. So in the indefinitely iterated prisoner's dilemma game, each player can choose among various reaction functions, each of which tells him what to do in any given game as a function of the history of choices of both players. One such function is "always defect". Another is "conditional co-operation", tit-for-tat is defined as follows: Always co-operate in the first game. In each subsequent game, co-operate if and only if the other player co-operated in the previous game. Thus under certain conditions, the use of tit-for-tat by both players is an equilibrium of the game. Here co-operation is maintained by the shared knowledge that it will be permanently unravelled by a single defection. Elster says that intuitively, the requirement of unanimity may to be too strong and that conditional co-operation of this kind can be sustained by a norm of fairness (Elster 1989c: 44). Moreover, tit-for-tat is not fully conditional, since in addition to the conditional injunction to cooperate if others cooperate previously, it also includes the *'unconditional instruction'* to cooperate in the first round.

Axelrod (1984: 5) chooses the United States Senate as the typical case of the emergence of cooperation. Here each senator has an incentive to appear effective to his or her constituents, even at the expense of conflicting with other senators, who are trying to appear effective to their constituents. There are many opportunities for mutually rewarding activities by two senators. These mutually rewarding actions have led to the creation of an elaborate set of norms, or folkways, in the Senate. Among the most important of these is the norm of reciprocity - a folkway which involves helping out a colleague and getting repaid in kind. It includes vote trading but extends to so many types of mutually rewarding

activities that it is not an exaggeration to say that reciprocity is a way of life in the Senate.

The end story is that cooperation, once established on the basis of reciprocity, can protect itself from invasion by less cooperative strategies.

To sum up, it is plausible to treat the pervasive indeterminacy generated in game theoretic models as a negative result (Johnson 1993: 80; Heap and Varoufakis 1995: 2). More specifically it is possible to interpret the problem of indeterminacy as identifying the limits of strategic rationality as a mechanism for coordinating social and political interaction. The point here is not that actual social and political practices generate indeterminate outcomes but that in the austere world captured by *formal* game theoretic models, unembellished strategic rationality does so (Bohman 1991; Elster 1989: 85). The unintended upshot of this effort is to demonstrate systematically what critical theorists suspect, namely, that in many settings strategic rationality alone does not suffice to sustain social relations. In fact, normal game theoretic results show that in dynamic settings, strategic interaction generates widespread indeterminacy in the form of multiple equilibria and attendant coordination problems (Johnson 1993: 80).

Although tit-for-tat is one response, there are other responses. Elster appeals to *norms* for explaining action (Elster 1989c). Tit-for-tat can be interpreted as a norm but it is a norm which presupposes strategic talents among the players. Thus rational prudence still underlies the tit-for-tat strategy. Elster, on the other hand, takes a broader interpretation of norms. Critical theorists might advance a distinctive interpretation of game theoretic results by identifying what beyond strategic rationality is at work in coordinating social and political interaction. Critical theorists identify the mechanism - the binding force of validity claims raised in communicative action. This is the juncture where a conversation between critical theory and game rational choice theory can begin. However, game theorists might

counter that insofar as their theory begins to specify its own limits, they need not rely on critical theory to perform that task (Johnson 1993: 80).

In his theory of communicative action, Habermas seeks to reconstruct "institutionally unbound" speech acts. That means, utterances derive force as mechanisms for coordinating social interaction solely from universal validity claims rather than from the contingent normative or institutional context in which they are advanced (Habermas 1979: 38-40; 1984: 295). So too, game theory reconstructs institutionally unbound strategic actions. It aspires to identify equilibrium outcomes sustained solely by the choices of strategically rational actors. It standardly treats institutions, culture as incidental and inessential detail and relegates them to the boundary conditions that tie the modelled structure to its unmodeled environment.

In the next section, I shall discuss Elster on social norms and then move on to discuss the comprehensiveness of discourse ethics.

4.2 Rational Co-operation: Interests and Social Norms

The problem of indeterminacy led some rational choice theorists to move away from the strict postulates of the theory. Theorists like Elster have appealed to 'norms' to supplement 'interests'. Elster's analysis differs from other rational choice theorists in the sense he appeals to a broader typology of human motivations.

The literature on the problem of cooperation can be divided into two parts: centralised and decentralised solutions. Centralised solutions include State and selective incentives. Both State and selective incentives argument presuppose the presence of a centralised coercive agent (Olson 1965). Decentralised solutions include altruism (Margolis 1982), spontaneous solutions like Tit-for-Tat (Axelrod 1984), anarchical order (Taylor 1987) and social norms (Elster 1989c). Excepting Elster, in all these solutions there is a tacit acceptance of "rational prudence" as a

motivational principle for resolving cooperative problems. All these works in one way or another are interested in the political-modelling process in studying the conditions under which cooperation can be sustained. I argue that mere rational self-interest cannot be a helpful guide in modelling cooperation. Even game theorists like rational choice theorists stipulate that no explanation of successful collective action should ever appeal to more than 'rational prudence'. This is for the reason that they fear that a broader concept of motivation would rob the theory of any explanatory power. Typical cases of cooperation in the absence of central authority include international politics or trade barriers between two industrial nations.

Collective action can be defined as the choice by all or most individuals of the course of action that, when chosen by all or most individuals, leads to the collectively best outcome. This course of action can also be referred to as co-operative behaviour. Despite the fact of free-rider problems and prisoner's dilemma situations, it is an empirical fact that people do cooperate on many occasions. Public choice theories fail to explain much of the cooperative behaviour that occurs in the real world. They limit the scope of theory to suit the theoretical postulates, thereby arbitrarily restricting the domain of inquiry. A better way of approaching the problem of cooperation is to pose the question differently. Why do people co-operate? In order to explain this problem rational choice theorists need to go beyond the assumption of self interest and appeal to a richer and more diverse account of motivational assumptions.

The task of social sciences is not just to explain the failure of co-operation but also to explain 1. why we are not in a state of nature and 2. how societies do have a modicum of order. Elster discusses the conditions for order in the social world by posing the question: what is it that glues societies together and prevents them from disintegrating into chaos and war? (Elster 1989c: 1). He argues that

much of social and public choice literature seems to be out of touch with the real world, in which there is a great deal of 'honesty' and sense of 'duty' (Elster 1989a: 179). Elster disputes the view of public choice theories that civilised society depends on having institutions that make it in people's rational self-interest to speak the truth, keep their promises and help others - not on people's having good motivations (Elster 1989b: 52). The most common criticism against public choice theories is that they fail to explain high baseline rates of cooperative behaviour in many spheres of social and political life like voting, voluntary associations, environmental movements etc. Elster argues that decentralised solutions are more fundamental than centralised ones, since compliance with central directives is itself a collective action problem.

In discussing the problem of cooperation, Elster employs two conceptual tools. One is rational choice theory and the other is the theory of social norms. Elster firmly believes that social norms provide an important kind of motivation for action that is irreducible to rationality or indeed any other form of optimising mechanism. Elster recognises the limitations of modelling collective action on purely instrumental grounds. The problem of collective action which was discussed extensively in chapter 1 implies that the rational self-interest of individuals may lead them to behave in ways that are collectively disastrous. Elster considers spontaneous mechanisms for coordination and cooperation. Voting, cleaning up the environment and abstaining from polluting it are classical collective action problems.

Elster appeals to social norms such as the "norm of fairness" and the "norm of everyday Kantianism" (Elster 1989b: 113-23, 131-34) to supplement rational choice. And within everyday Kantianism, Elster identifies a specific type of irrationality called 'magical thinking' which plays an important role in motivating people to cooperate (Elster 1989c: 186). For norms to be social, they must be

shared by other people and partly sustained by their approval and disapproval. Some norms, like norms against cannibalism or incest, are shared by all members of society. The other respect in which norms are social is that other people are important for enforcing them, by expressing their approval or disapproval. These sanctions can be strong. In addition to being supported by the attitudes of other people, norms are sustained by the feelings of embarrassment, anxiety, guilt, shame etc. that a person suffers at the prospect of violating them or at the prospect of being caught violating them. Elster gives an 'individualist' defense of social norms. A norm is a propensity to feel shame and to anticipate sanctions by others at the thought of behaving in a certain, forbidden way. He argues that social norms have a *grip on the mind* (Elster 1989b: 61) that is due to the strong emotions their violations can trigger. He believes that the *emotive* aspect of norms is a more fundamental feature than the more frequently cited *cognitive* aspects. If norms can coordinate expectations, it is only because the violation of norms is known to trigger strong emotions, in the violator himself and in other people. The operation of norms is to a large extent blind, compulsive, mechanical or even unconscious.

The force of norms - the feature that makes manipulation and interpretation worthwhile - is that they do have a grip on the mind; otherwise, there would be nothing to manipulate. Thus Elster defines social norms mainly by their intrinsic nature. Elster also distinguishes social norms from other related phenomena like moral norms; legal norms; convention equilibria; private norms; habits and compulsive neuroses; tradition; cognitive phenomena. And, finally, norms have independent motivating power and are irreducible to optimisation. The task is to identify the precise mix of motivations - selfish and normative, rational and irrational - that make interaction possible.

Elster argues that there exist distinct norms that may induce people to co-operate. These include the "norm of fairness" and "everyday Kantianism". Both are

non-selfish motivations. Kantianism says individuals want to do what would be best if all did it. It is not sensitive to what others do. The norm of fairness says individuals do not want to take a free-ride on the co-operation of others, but neither do they want to cooperate when few others do.

Norm of everyday Kantianism (Elster 1985: 148): This says that one should cooperate if and only if universal co-operation is better for everybody than universal defection. There are two reasons why the principle is called everyday Kantianism rather than simply Kantianism. First, Kant's own formulation of the principle does not refer to what would happen if everyone acted in a certain way. Rather it asks whether one can 'will' that every one will act in that way. One reason one cannot 'will' that all will act in a certain way might appear to be that it would be worse for all if all did so. This is not Kant's view. Rather he argued that one cannot 'will' x if the notion of all doing x harbours a logical or pragmatical contradiction. If breaking promises, for instance, were to be made into universal principle, the concept of promising would lose its meaning. Second, the arguments made below presuppose a naive form of Kantianism which excludes the use of mixed strategies. Some of the paradoxes of everyday Kantianism would disappear if the agents were allowed to randomise between cooperating and not cooperating.

The everyday Kantian does not consider the costs to himself of cooperating. Like people motivated by fairness, everyday Kantians are usually outcome insensitive with respect to benefits but not with respect to costs. Roughly speaking, they proceed in two steps. First, they use something like the categorical imperative to decide where their duty lies. Then before acting, they consider whether the costs are prohibitive, which, on a given occasion, they may well be. The trade-off will differ across people. Some Kantians pay virtually no attention to costs, while in others the voice of duty is reduced to a whisper that is easily offset

by considerations of cost. For that reason, even everyday Kantians might be sensitive to the number of other cooperators, namely if these increased the costs of co-operation above the point where they offset the call of duty. In cases of low cost or constant cost cooperation (like voting), this complicating factor can largely be ignored, but in other cases it could be important.

Furthermore, in practice the everyday Kantian is somewhat sensitive to benefits. Although he does not consider the likely impact of his cooperation except by asking, "what if everyone did that?" he does consider the impact of universal cooperation. It seems plausible that the strength of his feeling of 'duty' depends on the difference between universal cooperation and universal noncooperation. The smaller the difference, the lower the voice of conscience and the more likely it is to be offset by considerations of cost. Once again, this dependence will vary across people. However, everyday Kantianism is not without problems. Consider a situation in which unilateral cooperation is harmful to other people. Everyday Kantianism prescribes the cooperative strategy in such cases, regardless of the disastrous consequences that might ensue if few others follow suit. Unilateral disarmament could under certain circumstances increase the risk of war. Unilateral acts of heroism or sacrifice can give authorities or employers an excuse to crack down on non participants as well as participants.

Two questions that embody the principle of everyday Kantianism and that have strong emotional appeal are: If not me, who? and But what if everyone did that? Both questions suggest that there are only two possible states of the world, one in which everyone cooperates and one in which no one does so. It hints that it is up to me which of these states will come about. If I am in a sufficiently confused state of mind, I may indeed be persuaded to believe, to act as if I believed, that everything turns upon my behaviour. Psychologically, if not logically, there is a short step from the thought 'if I don't do it, why should anyone?' to the thought 'if I

don't do it, nobody will'. Elster elaborates on this point, which is of fundamental importance.

To put it differently, everyday Kantianism rests on a form of 'magical thinking' - a specific type of 'irrationality' that plays an important role in many decisions to cooperate. Magical thinking is the confusion of 'causal' and 'diagnostic' efficacy, or the belief that is as if one could change the 'cause' by acting on the 'symptoms'. This phenomenon is illustrated in an experiment by George Quattrone and Amos Tversky in which two groups of subjects were told that people with a certain kind of heart have a longer life expectancy and greater tolerance to cold water after exercise than people with a different type of heart. When later asked to keep their arm in cold water after an exercise task, they endured it for a longer time than they had done before they were exposed to the information. The subjects acted as if by modifying the symptoms they could, magically, alter the cause. In a similar fashion, the form of magical thinking also works in the prisoner's dilemma. Consider two persons in the prisoner's dilemma who have to make their decisions independently of each other. If they are sufficiently alike, each of them may reason in the following manner. If I cooperate, there is a good chance that the other will cooperate too. Being like me, he will act like me. Let me cooperate to bring it about that he does too. Once again, the behaviour that is diagnostic of the other's cooperation is chosen as if it could have causal efficacy.

Norm of Fairness (Elster 1989c: 187): The norm of fairness[1] tells an individual to cooperate if and only if everybody else, at least a substantial number of others, cooperate. Although all members of a group may share this norm, they may have different thresholds of cooperation. For some people, the norm takes effect with a relatively small number of other cooperators. Others may require nearly universal cooperation before they join. Among the other cooperators whose presence

triggers the norm of fairness for a given person, some may themselves be motivated by the same norm, with a smaller number of other cooperators required. Among the latter, some may also be motivated by the norm of fairness, but as we descend the chain we shall eventually meet some people who cooperate for other reasons. Cooperation could never arise in a population in which everybody was motivated by the norm of fairness. It is interesting to see why.

Although the norm of fairness is not in itself consequentialist, it can coexist with consequentialist considerations. People motivated by fairness may be sensitive to costs of cooperation even if they do not consider the benefits. A person's own conscience and social pressure tells him to cooperate when most other people do.

The norm of 'fairness' makes cooperation conditional upon the actual cooperative behaviour of others, not on their anticipated cooperation. Elster is concerned with conditional cooperation generated by 'conformism' rather than by rational prudence. In this respect it differs crucially from conditional cooperation in iterated games. In these games, all group members can be conditional cooperators who converge on cooperation because everybody expects everybody to do so. But the norm of 'fairness' does not work in the same way. Consider families in a peasant community who are led, by individual rationality, to have more children than is collectively rational. If there were a social norm against large families, all would benefit from it. Could such a norm - that is, a behavioural package that involved having smaller families punishing defectors, punishing nonpunishers and so on - be represented as the solution? Elster thinks not. Social norms cannot be sustained by instrumental rationality of this kind. To be effective, they must be internalised, so that violating them in the presence of others is felt to be shameful and wrong, not simply a mistake or a lapse from rationality. The emotion of shame is not within the scope of rational willing.

The obligation created by the norm of fairness does not derive from outcomes. The norm tells us to defect when there are few cooperators. There might be social norm against cooperating in such cases. The general norm against sticking one's neck out could easily lead to sanctions against unilateral cooperators. Conversely the norm of fairness tells us to cooperate when many others do so. In such cases, the norm of fairness might prevent them from taking a free ride. This problem is not that serious. There is too much cooperation, but not because the excessive cooperators actually harm anyone. Their contribution simply is not worth the effort. More serious problems arise when the last act of cooperation harms everybody, contributors and non contributors. Here is an example. Joining the army in wartime is a more substantial example. Those who stay home to work in vital industries may feel that they are violating the norm of fairness. In some situations that lack of coercive institution, the norm of fairness could drive co-operation beyond the optimal point.

It might even happen that universal cooperation is inferior to universal noncooperation. For example, consider the problem of cleaning up the lawn after a fete. If everybody joins in, the lawn will be trampled to destruction. It is better for all to have a lawn that is green but littered than to have it destroyed as a result of "misguided solidarity". Or consider again the problem of organising war. Suppose that the country in question initiated that war, so that universal noncooperation meant peace rather than defeat. It could then be worse for all if all joined the war effort than if nobody did, assuming that universal cooperation is so inefficient as to bring about defeat.

Thus collective action is defined by the feature that contributions have diffuse benefits and precise costs. In the standard theory, this provides individuals with a reason to abstain from cooperating. Turning the argument on its head, one can see that it is precisely because contributions are easily identifiable that they can

169

become the object of a social norm to cooperate. The fact that an additional contribution may actually bring about a slight decrease in the benefits from collective action has little motivating power.

All societies and groups face collective action problems. In all, at least some of the dilemmas are overcome and cooperation is achieved, frequently by coercion. Elster considers non coercive, voluntary cooperation. But in any given case we will observe that the individuals who make a voluntary contribution have different motives and not just instrumental ones. A successful campaign, strike, lobbying effort or election cannot be traced to a single homogeneous motive that animates all the contributors. Different motivations, building up on one another, can add up to a high rate of participation. Elster (1989c: 203) classifies broadly three motivational types in social interaction:

1. Selfish outcome-oriented rational individuals care exclusively about the output of collective action. Non cooperation is their dominant strategy.

2. Everyday Kantians cooperate under all circumstances. Strictly speaking this need not be true. If the costs of cooperation depend on the number of cooperators, even Kantians may make their cooperation conditional on that of others. Since Kantians are insensitive to the costs of cooperation, cooperation is a dominant strategy.

3. People motivated by the norm of fairness would not take a free ride on the cooperation of others, but neither do they want to cooperate when few others do. Here cooperation is viewed as *conditional* depending on the number of other cooperators.

To illustrate his case of collective action from a mixed motive perspective Elster (1989c: 206) takes the example of 'voting'. Unlike Downs' (1957: 38-50) analysis of voting (as argued in chapter 1), Elster focuses on the interplay between 'selfishness', 'fairness' and 'everyday Kantianism' (Elster 1989c: 206). Voting is seen as an expressive act. Consider the decision to vote in a large anonymous election. Self-interested rationality dictates abstention at least in the absence of selective incentives. Other attempts to demonstrate with high turn-outs of selfish voters have invariably failed. Hence most writers invoke the notion of 'civic duty' or see voting as an expressive act. But these concepts are rarely given specific content. At the individual level, turn-out is explained by individual properties of voters: age, sex, marital status, income, occupation, and education and so on. In addition turn-out is also explained by three variables:

1. direct costs of voting,
2. expected closeness of the election and
3. perceived differences among the major alternatives.

The importance of the these variables shows that 'civic duty' does not exclude some sensitivity to outcomes.

The first step towards a better understanding of turn-out rates must be to refine the notion of 'civic duty'. In the present framework, the notion can be understood in two ways, corresponding to 'Kantian duty' and 'fairness'. The motivations of duty and fairness are not outcome oriented as far as the benefits of voting are concerned, although they can be sensitive to costs. The fact that the costs of voting are a major part of the explanation of low turn-outs is consistent with an interpretation of civic duty in terms of fairness or everyday Kantianism. The influence of the other two external variables is consistent with everyday

Kantianism but not with the norm of fairness. Because of its roots in magical thinking, everyday Kantianism becomes stronger the smaller the reference group to which it is applied. The more other people are like oneself, the more plausibly (in a psychological rather than logical sense) one can infer that they will behave like oneself. Therefore, if the election is expected to be close, the more plausible is the notion that my voting will be pivotal. Also the smaller the perceived difference between the outcomes, the weaker is the voice of conscience. There are no similar reasons that the norm of fairness should be sensitive to the expected closeness or to the perceived difference.

Now the question is to ascertain which proportions of the electorate were motivated by which varieties of civic duty. Part of the explanation of voter turn-out also lies in changes in individual attributes. Demographic trends increase the number of young unmarried voters who are relatively unlikely to vote. Part of the explanation may lie in changes in the external variables. The effort and efficacy of the major parties in mobilising their supporters obviously also matter. Also, alienation from the political system may be a factor of some importance (Elster 1989c: 208).

It can also be argued that the steady decrease in voter turn-out can also be due to increasing political apathy and alienation from the political system. But the notion of alienation must be disambiguated. It could mean that people are less likely to vote because they feel less committed to a political system that fails to remove poverty and racial conflict. And political alienation could also be understood as a self-reinforcing erosion of civic duty, quite independent of substantive issues and failures.

Unlike rational choice theory, critical theory's humans are never social isolates. Their rationality has cognitive, moral and subjective aspects not just instrumental ones. Rationality means the competence to decide when it is

appropriate to act instrumentally, or in conformity with social norms, or dramaturgically, as an expressive subject, the ability to judge these qualities in other. This communicative competence is a matter of intersubjectivity, not just isolated subjectivity. But the perplexing problem critical theorists face is how to institutionalise discourses.

No doubt strategic calculation is one explanatory variable to explain action, but there will be many others, ranging from traditions, norms and cultures to differences of people' capacities and the contingencies of historical circumstances. Failure of theories to be empirically informed can result in irrelevant theorising and the mushrooming of controversies driven by little more than the theoretical conjectures out of which they emerged. Moreover rational choice theories presume uniform strategic capacity of all the members. The theoretical prediction is highly contingent on actors' strategic talents. Rational choice predictions run afoul when faced with asymmetrical game-playing capacities of the actors. Even tit-for-tat assumes that actors have uniform strategic talents.

4.3   Norms and Interests: Communicative and Strategic Action

The relevance of Habermas' work in the present context lies in his insistence that strategic rationality cannot by itself successfully coordinate social and political interaction (White 1988: 25). Taking strategic rationality as a counterpoint, Habermas depicts communicative reason - embodied in validity claims to truth, rightness, and sincerity that are implicitly and necessarily raised in human speech - as an unavoidable pragmatic presupposition of language use and hence of social interaction. He casts his theoretical enterprise as a research program intended to contribute to an empirically oriented critical social theory (Habermas 1984: xxix, 274). In this way he aspires to take seriously the theoretical contributions of the

social sciences, suitably disencumbered of what he considers misplaced positivist pretensions.

Habermas as we have discussed anchors his entire theoretical project in a categorical distinction between two sorts of social interaction. I do not want to repeat the contrast which has been extensively examined in chapter 2. Here I shall merely elucidate some of the problems it encompasses. Habermas seems to subscribe to a minimalist conception of strategic action. When first discussing strategic action, he suggests that it incorporates into the individual decision making "the anticipation of decisions on the part of at least one additional goal oriented actor (Habermas 1984: 94-95). Yet he also almost immediately begins to include 'egoism' and 'atomism' as intrinsic features of strategic action. These features are largely artefacts of his own theoretical apparatus and are not essential to the concept of strategic actors. He also observes that game theory employs the concept of self-interest and that individuals are solely motivated by egocentric calculations of utility (Habermas 1984: 88, 94-95, 101). And finally, he also depicts strategic action as atomistic and the actors as isolated; solitary and atomistic (Habermas 1984: 10, 85, 273-74).

Communicative action operates in the medium of language and is oriented toward reaching understanding. It coordinates interaction via consent or rational agreement. Parties to communicative action aim to cooperatively negotiate shared understandings of the nature of their interactions. Communicative action derives its force from the potential for rational agreement embodied in validity claims. This guarantee represents the "telos of mutual understanding" that Habermas claims is inherent in human communication. Moreover he insists the process of raising the validity claims is not a contingent aspect of language use but constitutes the unavoidable presupposition of communicative interaction. Consequently to deny that his or her utterances raise validity claims means ending up in a "performative

contradiction". Given this brief sketch it seems that validity claims are central to Habermas' idea of communicative action. The question arises of what sort of argument Habermas advances for the existence of validity claims His theory is plausible to the extent that he can demonstrate how rational agreement or consent, operating in the medium of language, coordinates social interaction. And agreement or consent in turn emerge because speakers mutually recognise the binding force of the validity claims raised in speech acts.

The difficulty is that Habermas' analysis of speech acts is not itself an argument for validity claims. Instead, it presumes that validity claims exist and function to co-ordinate social interaction: validity claims give the illocutionary act a rationally motivating force (Habermas 1979: 65). Habermas presumes - but does not show - that there is a telos of understanding built into the validity basis of human speech.

Given their apparently divergent approaches to the question of rationality, it is useful to view critical theory and rational choice as competing research traditions. The substantive point of contact between rational choice and critical theory is that game theorists typically concentrate on interactions between rational actors within a prespecified context. Habermas on the other hand focuses on a conceptually prior level of analysis. He focuses on the efforts of agents to define the context within which their ongoing interactions will transpire. Habermas, as we have seen, characterises strategic interactions as coordinated by influence, arbitrary choice or complementarity of interest. He says that if the interactions are not coordinated through achieving understanding the only alternative is force. But precisely because strategic actors mutually recognise one another as equally rational and goal oriented they can grasp that the pursuit of their individual plans is entangled in a web of social interdependencies.

Habermas seeks to constrain the nongeneralisable interests that provide players with incentives to dissemble or misrepresent. But he also presumes that they are communicatively competent. He insists that they are capable not only of producing comprehensible utterances but of embedding their utterances, however unanticipated they maybe, in a system of criticisable validity claims of truth, rightness, and sincerity (Habermas 1979: 26-27). This competence sustains processes of contesting and redeeming validity claims that impose pragmatic constraints on the ability of players to make utterances that solely express their nongeneralisable interests. Here Habermas identifies what game theorists lack - a mechanism that might compellingly account for the binding force of language in strategic interaction.

Habermas dissociates the force of rationally motivated agreement from both the standard model of social norms and strategic action. For him, social norms operate as prior constraints on action (Habermas 1984: 35, 287, 296-97). Habermas largely conforms to the way in which rational choice theorists distinguish between normative behaviour and rational action (Elster 1989c: 97-107). An alternative way of conceptualising the indeterminacy problem is to look to Habermas for solutions. It is interesting to see whether problems of collective action can be unlocked by appealing to notions like morality, social norms and discourse. The strength of discourse ethics lies in its validity claims. What constitutes a just and fair equilibrium solution can be decided dialogically. Since in dialogue competing reasons are exchanged in an attempt to reach an agreement, the morality of preferences can be challenged in the public sphere.

The problem of cooperation can appeal to a Kantian sense of rationality which requires that we should act upon rules which can be acted upon by everyone. In this context, the best reply to another's action rule is one which generalises to form a Nash equilibrium when the best is understood in an

instrumentally rational fashion. Of course, there many be other demands which Kantian reason makes, but taken in isolation, the 'universalisability' condition might provide an alternative foundation for Nash.

Kant's practical reason demands that we should undertake those actions which when generalised yield best outcomes. It does not matter whether others perform the same calculation and actually undertake the same action as you. The morality is deontological and it is rational for the agent to be guided by a categorical imperative. Consequently in the free rider problem, the application of the categorical imperative will instruct Kantian agents to follow the cooperative action, thus enabling rationality to solve the problem when there are sufficient numbers of Kantian agents. This is perhaps the most radical departure from the conventional instrumental understanding of what is entailed by rationality because, while accepting the pay-offs, it suggests that agents should act in a different way upon them. Kantian morality refers to any kind of impersonal evaluation - explicated in the notion of impartiality or categorical imperative. The notion of rationality is no longer understood in the means-ends framework as the selection of means most likely to satisfy given ends.

Of course there is a tricky question concerning whether these rather weaker moral motivations (acting on what is 'fair' or what is 'right') mark a deep breach with the instrumental model of action. It might be argued that such ethical concerns can be represented in this model by introducing the concept of ethical preferences. Thus the influence transforms the pay-offs in the game. So even though people still act instrumentally, the game ceases to be a prisoner's dilemma after the transformation. There is some evidence that the prisoner's dilemma can be unlocked when individuals are suitably morally motivated. However, as discussed in chapter 3, the strength of Habermas' theory of discursive rationality is that it goes beyond Kant in the sense it subjects all moral claims to dialogue.

For Habermas, Kantian morality is rooted in a monological model of subject-object dualism. Habermas contends that issues concerning 'fairness' will have to be resolved dialogically. The justification of Habermas' procedure of moral argumentation as universally valid takes as its starting point the unavoidable presuppositions of communicative action. Habermas insists that an impartial assessment of claims can only proceed dialogically, in terms of a real argumentation where the individuals concerned reach an agreement co-operatively. As discussed earlier (Chapter 2), according to Habermas communicative action is action oriented towards reaching understanding. The interaction will be coordinated by quite distinct mechanisms - by consent in the case of communicative action and by influence, arbitrary choice, or complementarity of interest in the case of strategic action. Language forms the medium of interaction. Communicative ethics can direct us only to a particular way of thinking about fair procedures for adjudicating normative claims.

It is often argued by liberal critiques like Ackerman (1989: 7-8) that the postconventional ethics like Habermas's which incorporate very strong criteria of fairness, is unrealistic to the point of being somehow illegitimate. 'Strong' here means simply that a given agent may find that there is a substantial gap between what the criteria require and what he/she initially perceives to be in his/her interest. The proposed criteria of rules of discourse are so strong as to constitute an illegitimate imposition on individual identity by placing unwarranted constraints on what one can rightly hold onto as one's basic interests or undertake a as part of one's ground project in life.

The specific interpretation Habermas gives to the postconventional, post-metaphysical criterion of reciprocity or fairness must be understood in relation to his notions of 'generalisable interests' and real discourse. Arguments based on interests common to the group - "generalisable interests" in the language of critical

theory or "public goods" in the language of public choice - are normally more acceptable than arguments based on self-interest. The point is that discussion can invoke different kinds of 'motivations' than one finds in isolated individual behaviour. From the standpoint of Habermas, in cases of normative conflict, actors must act judiciously, that is, they should not pursue their immediate interests but must be concerned to judge the dispute from a moral point of view and settle in a consensual manner (Habermas 1984: 19).

One important issue which needs clarification is what exactly is required of individuals when they agree to follow the discursively interpreted demands of *'fairness'* (White 1988: 75). When one tries to justify a normative claim, one is obligated to show that the interests underlying it are 'generalisable' rather than merely 'particular'. In some cases this demonstration and an ensuing agreement might come easily. For example, traffic rules and laws against murder can be seen as resting on generalisable interests in the safety and sanctity of persons. But of course most questions in ethics and politics are not so amenable to simple solution. What does communicative ethics require agents to do when agreement is not so easily reached? The rules of discourse require that agents sincerely reflect upon the different 'need' interpretations which underlie their respective, but conflicting, concepts of what are generalisable. This means that they must exhibit a sort of ongoing critical flexibility: a willingness to reconsider and possibly modify their need interpretations, when they appear to manifest weaker claims to universality than alternative ones (White 1988: 75). At this point the only thing which it is necessary to emphasise is that the result of such discursive reflections on needs is not necessarily any consensus.

However the criterion of *flexibility* provides no guarantee that agents will come up with generalisable interests. If this is so, what guidance does communicative ethics offer when agents appear to have exhibited such flexibility

and yet still fail to come to a consensus on interests? Compromise appears to be the only possible solution. According to Habermas (1975a: 112) when interests continue to conflict - that is, they do not prove susceptible to generalisation - resort must be made to compromise. This assumption needs some clarification.

Habermas's position here is usually misunderstood in one way or another. Many social theorists view his reference to compromise with suspicion given his apparently contradictory claim that he can show us a way out of an impenetrable pluralism of conflicting values, needs and interests. This claim has led critics to infer that Habermas, although he speaks of compromise, actually has a deep contempt for it, seeing it as an "eliminable imperfection" resulting from the flaws of existing individuals who have not yet developed the genuine human needs which members of an emancipated society would have. But critical theorists like White (1988: 75) argue that there is no such hidden agenda in communicative ethics on the question of genuine needs. So this interpretation of Habermas is simply wrong.

The question is how can a discursive perspective get us beyond an ethics of simple contractarianism, which relies only on the strategy of bargaining between conflicting interests? Such an alternative may seem to be the only direction in which to go if one abandons the strong minimal universal criteria for dealing with normative disputes.

Rejecting strong criteria for fair agreements, rational choice theorists contend that we must be more modest and simply seek compromises which prove acceptable to whatever different actual points of view and interests are involved. This position is articulated by the contractarian theorists like Rawls (1985: 223-251) and Buchanan (1989: 49-62). Although this argument is attractive, it is nevertheless achieved at the cost of a tacit endorsement of whatever structures of inequality might exist in a given society. Such an ethics will assign an unqualified legitimacy to compromises which solidify initially unequal bargaining positions. It

ignored the operation of power in social and political relations. This is because it has no way of distinguishing compromises which are acceptable but illegitimate from ones which are acceptable and legitimate. An example of the former would be an agreement which a disadvantaged person accepts on prudential grounds, even though that acceptance occurs under conditions of constraint. The problem with a perspective like this is that it cannot adequately differentiate "validity claims from power claims" (White 1988: 76). It fails to give useful standards for unravelling hidden power relationships in a apparently consensual arrangement. This is a crucial ethical aspect of a critical approach to the theory and practice of politics.

From a Habermasian perspective the basic guidelines for compromise construction must themselves be justified in discursive terms. This alone can supply a standard for separating 'legitimate' from 'illegitimate' compromises (Habermas 1975a: 112). In particular, the discursive emphasis on procedural equality, participation, non-deception and non-manipulation provide criteria in relation to which compromises must be called to account. These discursive constraints on compromise do not give us a precise formula or method for unambiguously separating legitimate from illegitimate compromises. In other words, although one can say that a given compromise must not rest on one-sided manipulation or advantages derived from an unequal bargaining position, these prescriptions must always be interpreted and applied by actors operating within a particular cultural tradition. Thus, for example, one cannot categorically assert that, because normative claims must be evaluated from a perspective of equality, actors cannot ever accept claims which entail social relationships permitting different sorts of inequalities. What the constrained indeterminism does allow us to do, however, is to shift the burden of proof in normative argumentation. And in shifting the burden of proof, communicative ethics shifts the responsibility for abandoning reason in favour of force. In some cases where inequality between agents exists, a

framework like rational choice will confront the disadvantaged agent with the choice of accepting a legitimate bargain which confirms his inequality or breaking off reasoned dialogue and using force - that is, a choice which renounces the claim to moral legitimacy.

According to Habermas arguments about compromise come into play in situations where sincere reflection by participants in discourse does not lead to their discovering generalisable interests. The criterion of sincere reflection on needs actually entails an ethical orientation. The rules of discourse require a reflective elucidation of the need interpretations underlying disputed norms. And each agent takes up this critical, reflective attitude toward the norms proposed by another agent, he/she forces the other to be self-critically reflective about his/her own needs and their universality. As needs are examined in the undistorted, dialogical light of discourse, agents have the possibility of reaching more truthful interpretations of their own particular needs as well as those which can be communicatively shared. In other words, Habermas' model of the subject has moral implications. Habermas' emphasis on a minimal moral orientation gives priority to the questions to be asked in situations in conflict.

It is interesting to compare discourse ethics and other neo-Kantian approaches like Rawls's. According to the Rawlsian view, the principles of justice most compatible with the value of the autonomous personality can be specified via a model of collective choice constrained by certain procedural rules. These are deemed to be reasonable limitations on choice, defining the circumstances of justice. Although procedural, this Rawlsian model of legitimacy is not dialogic. The choice of the principles of justice is not thought to result from any kind of dialogue among the parties to the original position. However contract has its relevance in the realm of the political. This can be seen in Habermas' own agreement with the Rawlsian notion of 'overlapping consensus'.

In one of his interviews, Habermas (1995: 20) expresses sympathy with the contractarian construction in so far as it identifies a case for grounding universalistic norms. Habermas cites Rawls' conception of "overlapping consensus" (Rawls 1985: 224) as a good case for the nonpartisan adjudication of international conflicts. It relies on the case that universal world religions converge in some core condition of moral intuition that can be interpreted as equal respect for all, equal respect for the need to promote the integrity of each individual person. For Habermas this assumption forms the normative basis for consensus within world perspectives and religious interpretations of deep moral feelings which could support the norms of peaceful coexistence. Although this is compromise it is a compromise which is based on universal norms. The only difference between Rawls and Habermas is that the former wants differences to be resolved through mutual identification of universal norms which can form a normative basis for consensus putting aside differences[2] and the latter sees consensus on norms to be arrived at in dialogue.

The idea of egalitarian reciprocity is at the heart of Habermas' theory. As argued in chapter 3, interpreting discourse ethics as a moral transformatory idea, Benhabib (1989: 148) argues that discourse ethics can also be part of a more general theory of self, historical change, and social structures. Her contention is that proceduralism does not imply formalism and ahistoricism. As Benhabib argues, Habermas himself notes that norms are generated not by philosophy but by real life (Habermas 1990: 204). Thus the validity of disputed norms emerges in everyday communicative practice. Discourse ethics is susceptible to the problems of indeterminacy. Critical theorists argue that

> the purpose of discourse theory is not to draw a blueprint for a well-ordered society. Rather, the purpose is to develop a model of public dialogue such as to demystify existing power relations (Benhabib 1989: 154).

This involves: identifying those issues which are prevented from becoming public because of existing power constellations; identifying those groups that have not had access to means of public expression and advocating their inclusion in the discourse of legitimacy; distinguishing agreement and pseudo-compromises based on the intractability of power relations; saying what is in the public interest as opposed to the universalisation of what is only the interest of a particular group (Benhabib 1989: 154).

On this interpretation, critical theory need not take a stand on all issues in everyday social and political life. It is not indeterministic in so far as it addresses the more substantive issues like power and domination (White 1988: 86). But at the same time there are weaknesses of communicative action. Critical theorists need to provide a more precise idea of the conditions under which language can be expected to coordinate social and political interaction. For example, communicative action remains susceptible to pressures of nongeneralisable interests. This is the clearest in pure-conflict, zero-sum interactions where communication has no force. More fundamentally, suggesting that the validity basis of language works to coordinate social interaction is not the same as establishing how it does so. One of the perplexing problems is how to institutionalise discourse ethics. Rational choice theorists argue that Habermas' ideal speech situation is so idealistic that it is far removed from reality which is a messy one and that social choice problems are not that simple to be resolved by mere dialogue.

Although mutually hostile intellectual traditions, rational choice and critical theory are engaged in determining what rationality can mean in the realm of social and political interaction. Habermas' enterprise can help rational choice theory move beyond instrumental rationality. As argued by James Johnson (1993: 74), game theory extends and deepens the critical theorists' basic intuition that unembellished strategic rationality cannot adequately sustain social and political interaction.

Critical theorists, on the other hand fail to recognise the multiplicity of theories present in rational choice paradigm. As James Johnson (1991: 113-114) argues, the problem with the external critics of rational choice theories such as 'interpretive' and 'critical theories' is that they subscribe to the orthodox view of rational choice as "positive" science. This could be for two reasons: one is that they rarely demonstrate close familiarity with economic or rational choice models and consequently have a fuzzy and uncritical picture of this theoretical programme. The other reason is that they tend to adopt the standard economist's assumption of "given" preferences (and subsequent neglect of the literature on preference formation) as the platform to press their point. By contrast, rational choice theorists readily admit that the assumption of 'exogenously' determined preferences is a lacuna in their technical edifice. Consequently, critical theorists' accusation that rational choice accounts provide a monocausal explanation is not a valid one in the light of the internal revisions. Critical theorists dismiss Elster's works as another attempt to build a naturalist social science (White 1989: 11). Critical theorists need ·to abandon their positivistic interpretation of rational choice theory.

Conclusion

I conclude with a general observation that rational choice theorists on their own terms have begun realising the weakness of strategic action. Both critical theory and rational choice theory can gain from engagement. As Dryzek (1992: 397) says

> by deploying an expanded communicative conception of rationality, critical theory can help move public choice beyond several impasses. In turn critical theory benefits from this encounter by gaining content for its currently rather abstract critique of politics.

In dealing with these problems of aggregation game theorists approach it from a pragmatic angle in seeking to find solutions in realistic settings. Critical theory on the other hand idealises social interaction and finds itself miserable when it comes to the problem of institutionalising discourses. To put it differently, rational choice theories have explored the structures of social interaction predominantly in the institutional setting of a democratic polity. Their main concern is with institutional design. On the other hand, for most of its history critical theory has ignored issues of institutional analysis and design of the sort that have preoccupied rational choice. And this is for two reasons: The first is revolutionary heritage. The second is, critical theory has operated for too long on an abstract and metatheoretical plane.

Much of the social choice and public choice literature seems completely out of touch with the reality wherein a great deal of honesty and norm governed behaviour is present and not simply opportunistic self-interest behaviour. In the next chapter I draw attention to the political arguments for democracy. I sympathise with both critical theorists and internal critiques of rational choice like Elster's that the task of politics is not just to devise institutions that can harness opportunistic self-interest to socially useful purposes. But an equally important task is to create institutions that embody a valid conception of justice.

Notes

[1]This idea is used more in hermeneutic or interpretive sense. Interpretive theorists argue that all social encounters rely on a web of shared beliefs (derived from rules) which enable to be remarkably powerful in sending non-verbal messages. The norms are shaped in tradition, history, language and culture. So they are culture-specific and history specific and thus differ from society to society. The power of these norms lies in their dictation of the everyday actions of individuals in a society. In this sense norms are noninstrumental and the choices of individuals are determined by them. Interpretive theory is exemplified by thinkers like Charles Taylor, Clifford Geertz. The theory focuses inquiry on the "web of language, symbol and institutions" that are the tangible vehicles of cultural significance. And they self consciously reject rational choice theorist's ambition to explain and predict.

² This idea can also be found in Ackerman (1989) and Holmes (1988). Ackerman presents a pragmatic-political as opposed to moral justification of the idea of a public dialogue based on the conversational constraint of neutrality. The question Ackerman poses is how different primary groups can resolve the problem of coexistence in a *reasonable* way. For him, the way out is the path of conversational restraint. It means "When you and I learn that we disagree about one or another dimension of the moral truth, we should not search for some common value that will trump this disagreement, nor should we try to translate our moral disagreement into some putatively neutral framework; nor should we seek to transcend our disagreement by talking about how some hypothetical creature would resolve it" (Ackerman 1989: 16-7). What Ackerman suggests is that we should simply say *nothing at all* about this disagreement and try to solve our problem by invoking premises that we do agree upon.

Having constrained the conversation in this way, we may instead use dialogue for pragmatically productive purposes: to identify normative premises all political participants find reasonable (Ackerman 1989: 16-7).

Chapter 5

Conceptualising the Political : Non-cognitive Vs Moral-cognitive

Introduction

In this chapter, I shall turn to the political implications of the two theoretical paradigms. The best way of characterising the politics of the two traditions is that public choice theorists, especially Buchanan and Tullock (1962) and Riker (1982) are committed to classical liberalism's tradition of limited government, constitutional constraints and the free market as opposed to populist notion of democracy. Riker (1982) is critical of populist theories that, inspired by Rousseau, see the outcomes produced by democratic practices as corresponding to something like the popular will. Public choice theorists see 'liberalism' as the only alternative to populism. Critical theory in its Habermasian form claims to represent discursive democracy.

From the standpoint of critical theory, the choice between liberal and populist interpretations of democracy is miscast and does not exhaust the constitution of democracy. In other words, Schumpeter and Rousseau do not exhaust the options available to democratic theory. The analytical results of public choice and social choice theories underscore the inadequacy of purely aggregative notions of democracy in ways that can sharpen the case for deliberation. In the eyes of public choice theorists like Buchanan (1991: 217), however, discursive democracy is another ideology committed to esoteric participatory democracy i.e. 'just another prisoner of romantic illusion'. For him, any attempt to go beyond the behavioural postulate of *homo economicus* is considered as idealistic, romantic and illusory (Buchanan 1986).

One significant difference between rational choice and critical theory is that the former discredited the notion of 'popular will' whereas in the latter the crucial question is what might underpin the development and recognition of a truly general interest. In other words, the aim is to examine the possible normative justification of democratic institutions.

In section 5.1, I shall discuss the competing conceptions of politics within rational choice theories (mainly public choice and Jon Elster) and in section 5.2, I turn to discuss Habermas' moral-cognitivist interpretation of democracy.

5.1 Rational Choice and Democratic Politics

5.2 Conceptualising Politics: Towards a Moral-Cognitivist View

5.1 Rational Choice and Democratic Politics

Concept of Preference-formation

The concept of preference-formation has been the central concept of rational choice theories. Rational Choice findings about the logic of majority rule and collective action problem have initiated new reflections on the normative foundations of democracy (Buchanan and Tullock 1962: 85, Green and Shapiro 1994: 4). Buchanan and Tullock develop a normative theory of collective choice grounding in contractarianism (chapter 3). And similarly rational choice theorists have come up with solutions regarding collective action problems (chapter 4). Here, in this section, I shall examine diverse views within rational choice theory on democratic politics.

Public choice theories see the political process as instrumental rather than an end in itself. Buchanan and Tullock (1962) and Riker (1982) see the decisive political act as a private rather than a public action viz. the individual and the secret vote. With these usually goes the idea that the goal of politics is the optimal compromise between given, and irreducibly opposed, private interests. Using his recent writings on preference formation, internal critics like Elster (1979; 1983a; 1983b; 1989a) argue that public choice theories have an undeveloped conception of preference formation. Preferences are seen as static, fixed and unchanging in the political realm. On the contrary, for Elster (1986b: 112), the purpose of politics is not simply to implement preferences, but following Habermas, to select them, in a forum in which various distortions may be discussed and brought to light - not by a political elite but by the public binding itself, outlawing certain preferences or ensuring the availability of a range of choices. The distinction between market and politics thus depends in part on a belief that preferences should not be taken as exogenous variables. This view is closely associated with a conception of individual freedom that understands the term to refer to selection rather than implementation of ends. The associated understanding of political freedom understands the term to refer to a deliberative process in which the citizenry in general select social ends (Sunstein 1988: 348).

As we have seen, rational choice theorists are critical of aggregation mechanisms which attempt to generate a general will. But there are significant differences in their responses. Elster (1986b: 111) argues that the notion of consumer sovereignty is acceptable in the market because the consumer chooses between courses of action that differ only in the way they affect him. In political choice situations, the citizen is asked to express his preference over states that also differ in the way in which they affect other people.

Elster (1983a: 33) argues that the social choice mechanism is capable of resolving the market failures that would result from unbridled consumer sovereignty but as a way of redistributing welfare it is hopelessly inadequate. If people affected each other only by tripping over each other's feet, or by dumping their garbage into one another's backyards, a social choice mechanism might cope. However, the task of politics is not only to eliminate inefficiency, but also to create justice - a goal to which the aggregation of prepolitical preferences is a quite incongruous means (Elster 1991: 120).

So the internal critics like Elster argue that principles of the market must differ from principles of the forum. The political process is not merely a means of serving persons' interests by satisfying their preferences. It is also a process which changes people's preferences. People are socialised and democracy helps to create a new human being, more tolerant, less selfish, better educated, and capable of cherishing the new values of the era of the enlightenment. Politics is the study of the ways of transcending the prisoner's dilemma problems (Elster 1976: 248-49).

Rational choice theories have shown in a series of impossibility theorems that there is no specific method for preference aggregation. Consequently, on the basis of the thin theory of individual rationality, they articulated a thin theory of collective rationality. The task of politics is seen as the question of devising rules and institutions for co-ordinating individual preferences. Buchanan and Tullock (1962) make unanimity a condition for political decisions. They argue for unanimity on the libertarian grounds that anything short of it will violate the rights of the minority, unlike the consensus theorists who believe that unanimity will emerge in rational discussion.

In this approach to rationality, preferences are assumed to be 'given'. The substantive rationality of the agents is never made into an issue, nor is the morality of their preferences. As Elster (1989a: 5-8) says, preferences can be distorted by

191

multiplicity of mechanisms.[1] Consequently, preferences as a basis for deriving collective mechanisms are a fragile foundation for social choice. One reason for not drawing practical political conclusion of limited government from the incoherency of "majority rule" is that one can question the adequacy of the model of preferences that is built into the approach when one is considering matters of political choice. People offer reasons for favouring their preferred alternatives.

The logic of advancing political preference must be understood in the light of these underlying reasons, and not simply as claims to have one's preferences satisfied whatever they happen to be. So the alternative model of collective choice would be not a process of preference aggregation, which there is a mapping from a set of individual orderings to a social ordering, but a process of dialogue in which reasons are exchanged between participants in a process that is perceived to be a joint search for consensus.

Elster (1983a; 86b) sympathises with Habermas' discourse ethics seeing the expression of preferences as an action. According to Elster the preferences people act upon may not be a guide to what they really prefer because of 'weakness of will' or 'excess of will'. And moreover an individual should not be able, by misrepresenting his preferences, to bring about an outcome which is better according to his true preferences than the one that would be brought about if he expressed these true preferences (Elster 1989a: 178). So even if we require that the social preferences be collectively rational with respect to the expressed preferences, they might not be so with respect to the real preferences (Elster 1983a: 32).

Elster argues that the political system could be geared to the task of changing preferences rather than aggregating them. Both in theory and practice, the idea is that the central concern of politics should be the *transformation of preferences* rather than their aggregation (Elster 1986a: 112). On this view the

core of the political process is public and rational discussion about the common good, not the isolated act of voting according to the private preferences. Elster relies on Habermas' notion of discourse ethics for this idea. Against the prepolitical nature of preferences, Elster associates with Habermas's writings for which the input of the social choice mechanism is not raw, quite selfish or irrational preferences that operate in the market, but informed and other regarding preferences (Elster 1986b: 119).

Knight and Johnson (1994: 279) insist that aggregation needs to be supplemented by institutional arrangements that embody and enhance democratic deliberation. In their view, aggregative arrangements lack the 'moral resources' required to generate and sustain legitimate collective solutions to politically contentious issues. The primary concern here seems to be that electoral outcomes are susceptible to influence by various arbitrary, exogenous, social, cultural or economic asymmetries (Knight and Johnson 1994: 278).

Elster's writings on preference formation convey a disillusionment with instrumental rationality. He considers several non-instrumental grounds for explaining action. With respect to individual choice, *social norms* can supplement rationality in explaining action and with respect to political choices, Elster argues, *justice* offers a guide to reform. Elster's contention is that in the present state, social sciences cannot aspire to be having a theory of social equilibrium. Social sciences as of now can at best be satisfied with uncovering mechanisms i.e. causal mechanisms (Elster 1989c: 250) in explaining social reality. In other words, it is important to recognise the inherent by contingent nature of social reality. Instead of arguing that society is to be understood on the model of the unitary actor, one might argue that the individual should be understood on the model of the fragmented polity. As said before, intrapersonal problems of preference aggregation, self-deception and other forms of cognitive compartmentalisation,

weakness of will, distort in complex ways in determining individual actions. This means, although Elster is suspicious of any grand planning, he is defensive of micro-level short term planning.

Distancing himself from the Marxists, in his essay *Self-realisation in work and politics: the Marxist conception of the good life* (Elster 1989c: 127-158), Elster expresses sympathy with Habermas' discourse-based model of politics wherein preferences get challenged and transformed in dialogue. However, Elster is apprehensive that it might be rejected as utopian (Elster 1986b: 103-29). Even so, Elster views politics as a vehicle for self-realisation and not just preference aggregation (Elster 1989d: 147).

Elster distances himself from the Marxists who did not believe that there would be room or need for conflictual politics in communism; *a fortiori* Marx could not promote politics as a channel for individual self-realisation (Elster 1989d: 147). Later Marxists have thought differently, notably Jurgen Habermas. Habermas suggests that Marx simply ignored the development of normative structures of interaction and concentrated exclusively on production paradigm. Elster is sympathetic to the view that the development of moral competence through rational discussion is a form of self-realisation that ought to be valued as highly as self-realisation at the "work place" (Elster 1989d: 147). Elster is sympathetic to analytical Marxists like Roemer and Cohen (Kymlicka 1991: 177-184) who stress the importance of the "access to social goods" and the influence it has on to the individuals.

Spontaneous Order or Social Constructivism

The second issue which remains contentious is the ideological divide between spontaneous order and social constructivism. One of the concerns of the political

theory of liberal individualism is to pass judgement on the legitimacy of particular institutions. Institutions, in this view, are to be regarded as legitimate insofar as all individuals who are governed by them would have broadly 'agreed' to their creation. Naturally much will turn on how agreement is to be judged because people in desperate situations will often agree to the most desperate outcomes. Thus there are disputes over what constitutes the appropriate *reference point* for judging whether people would have agreed to such and such an arrangement. This is an insight which has a special relevance for the discussion in the political theory of liberal individualism concerning the conscious creation of institutions through agreement. If the test of legitimacy is 'would individuals agree to such and such arrangement?' then we need a model which tells us what individuals will agree to when they interact. It seems natural for contract theorists to model the 'negotiation' as bargaining and interpret the outcome as the terms of agreement. Hence the need to know the likely outcome in order to have a *'standard'* for judging whether the institutions in question might have been agreed to.

Responses to the conflict of individual and social choice are varied. One response coming from conservatives like Oakeshott is that the *frailty of human reasoning* excludes conscious, deliberate reform altogether. In their view attempts to change society in a specific direction embody pseudo rationalism, the failure of reason to define and respect its own boundaries. In Elster's view, this is an ultraskeptical conclusion. Morality and social norms also count more than mere enlightened self-interest (Elster 1989b: 158). Before proceeding to discuss Elster's view, it is interesting to see the contrast between two different versions of the 'political' within rational choice theory. One strand advocates "spontaneous order" and the other "social constructivism" or "political rationalism". The former draws support from New Right libertarians like Hayek. The contrast is best illustrated in their differing views of the State.

Questions regarding the role of State have preoccupied political theorists for a long time. Many political theorists believe that something like State can resolve the particular difficulty of overcoming the prisoner's dilemma and provide public goods. In the case of multiple equilibria the State can through suitable action guide the outcomes towards one equilibrium rather than another. Thus the problem of equilibrium selection is solved by bringing it within the ambit of conscious political decision making. Likewise, with prisoner's dilemma/free rider problems, the State can provide the services of enforcement and guide the outcomes towards the cooperative equilibrium.

There have been two responses to this argument of State intervention and the thinking behind it has largely shaped the mental world of the twentieth century politics. Buchanan and Riker argue that there are problems of political failure that subvert the ideal of democratic decision making and which can match the market failures that the State is attempting to rectify. As argued in chapter 1, public choice theorists (Riker 1982; Buchanan and Tullock 1962) suspect the idea of something like the "general will" or "common good" in whose name the State might be acting. These are essentially negative arguments coming from the political right against 'political rationalism' or 'social constructivism'. The positive arguments against political rationalism turns on the idea that these interventions are not even necessary. The failure to intervene does not spell chaos, chronic indecision, fluctuations and outcomes in which everyone is worse off than they could have been. Instead spontaneous order will be thrown up in the form of a tit-for-tat strategy.

From the standpoint of tit-for-tat strategy, the argument for the State seems to be weakened because it appears that a group can overcome the free-rider problem without recourse to the State for contract enforcement. So long as the group can punish free-riders by excluding them from benefits of cooperation, then

there is the possibility of spontaneous public good provision through the generalisation of the tit-for-tat strategy. Although this is possible in small groups, it is impossible to do away with the state in the contemporary large societies wherein the state plays a larger role.

However, there are many practices and institutions which are surrogates for the State. Since something like the State as contract enforcer might well arise spontaneously through the playing of free rider games repeatedly, it need not require any *a priori* grand design. There need be no constitutional conventions. The result counts strongly for what Hayek refers to as the 'English pragmatism' as opposed to the 'European continental rationalism'. Hayek prefers the English tradition because he doubts that the formation of State is part of a process which liberates the social agent and that there is the knowledge to inform some central design so that it can perform the task of resolving free-riding better than spontaneously generated solutions like tit-for-tat (Heap and Varoufakis 1995: 195-96). The implication is that reason has its own limits.

Public choice theory represents an older normative approach to the state which has gained well established recognition in the economic profession. It consists of analysing the process of political decision making under the democratic system and of designing rules that ensure that the best preferred outcome will emerge from this process. Particularly relevant to the present discussion is the analysis of political parties as competing oligopolists trading in votes: according to this logic, each party lays down its political platform with a view to attracting as many votes as political market in which politicians behave like merchants in approaching voters. Influenced by the economistic logic, the public choice approach evades the issue of power: the state appears just as a channel through which the preferences of the public concerning goods get expressed. The state has therefore no *sui generis* power as a voice of the social interest. Neglect of power

phenomena can be traced not to any inherently conservative ideological stance, but to the working logic of their scientific paradigm. The theorists working in this paradigm view the world as an arena where agents freely contract or refuse to do so and whenever explicit transactions are not identifiable, implicit deals are posited to save the model: all human interactions are reduced to contracts.

Although there is apparently no general consensus about what power means exactly in rational choice theories, a useful starting point is Elster's distinction between force and coercion, taking coercion to imply the presence of an intentional agent or coercer, while force need not imply more than the presence of constraints that leave no room for choice (Elster 1985: 211-2). Force is an abstract mechanism through which the impersonal logic of an economic system limits the actors' choices and affects their life situation. By power, Elster means that a powerful agent, for his own benefit, drives another agent to take a course of action that he would not have followed in the absence of the first agent's intervention. It is however important to distinguish between two basic forms of power according to whether the power to act enhances or harms the influenced agent's interests. Power is considered exploitative in the latter case while it is akin to a leadership process in the former. The point is that public choice theories were essentially concerned with analysing the free play of market forces and their rules.

Elster's views falls more in line with social constructivist thoughts. By this, I mean the attempt to account for the objectivity of our normative assessments by relating the ideals and principles employed in our critical practices to an expressly normative conception of free and equal moral persons. The justification of this normative ground is ultimately reflexive or recursive in the sense that there can be no higher appeal to something beyond the idea of that to which free and equal persons can rationally agree (Habermas 1990). For social constructivists one can turn social outcomes into matters of social choice through rationalist intervention.

However, it is important to cultivate a greater sense of irony and toleration of diversity in view of the contingency that surrounds our social institutions and practices.

Arguing against public choice and social choice theorists who conceive the task of politics as devising institutions that can harness opportunistic self-interest to socially useful purposes (Buchanan and Tullock 1962: 21), Elster says that one of the aims is also to create a just society. He argues that justice provides an alternative guide to political action (Elster 1991: 116). Thus Elster's political rationalist solutions are grounded in 'justice' as an alternative motivation for political reforms.

His argument is that the main political reforms of the past century have not been supported mainly by instrumental considerations. Rather they have been carried by social movements anchored in a conception of justice. However it is important to note that Elster's aim is not to propose a theory of justice but to show the significance of it as a norm as opposed to self interest. This norm of justice forms an alternative guide to political action. Elster argues that given the fragility[2] of instrumental thinking in politics, the chosen conception of justice must focus on the inherent rights of the individuals to equal shares in decision making and in material welfare.

Elster (1991: 136) criticises Buchanan's version of "commutative justice" or "justice as mutual advantage". The arguments from commutative justice rests on a vision of society as a joint stock company, with the citizens cooperating for mutual advantage. For example, the implication of Buchanan's theory will be that although taxpayers may be willing to have some of their taxes spent on non-taxpayers, they usually insist on taking part in the decision to spend the money that way, and on excluding non taxpayers from the decision: 'no representation without taxation.' The denial of the right to vote to non-taxpayers rests on a very narrow conception

of justice. It is a vision of the well-ordered society as emerging from a bargain among self-interested individuals, in which those who have nothing to contribute and hence no bargaining power cannot expect to receive anything either, except from charity. Universal adult suffrage rests on a simpler and more compelling conception, transcending both instrumental considerations and commutative justice. Society is indeed a joint venture but the bond among its members is not simply one of mutual advantage, but also one of mutual respect and tolerance. If the first step in the development of democracy was the idea that no group of persons can be assumed to be inherently superior to others, the second was that no group can be assumed to be inherently inferior.

To illustrate how a bargaining strategy works, let us assume that a 'reform' is proposed and justified on consequentialist grounds. For each citizen the reform has dual effects: burdens and benefits. For each proposal there is some group which will think it unfair, compared to some other proposal under which it would be better off. The very plurality of co-operative arrangements prevents any one of them being chosen. This is a bargaining problem in which the outcome may very well be the breakdown of negotiations and the perpetuation of the inefficient statusquo which has the advantage.

Bargaining theorists like Thomas Schelling take the view that, in inherently conflictual situations the only way to reach agreement may be to hit upon a 'naturally salient outcome', a 'focal point' or in other word a 'Schelling point'. So according to this form of agreement what is naturally salient depends on perceptual psychology, social conventions and shared history, not on formal features that can be stated in rational choice theorists. Typical focal points in bargaining are "allow no exceptions", "divide equally", "do as we did last year" or "do nothing". The parties are assumed to have a common interest in arriving at some agreement but a conflict of interest over which agreement should be accepted (Elster 1989c: 50).

Thus focal points in democratic politics do not always embody substantive equality, but they represent formal equality in the sense of impartiality or fairness.

Agreement in a bargaining strategy is facilitated if one outcome is especially salient. For example, when two countries are bargaining over their common frontier, their task is greatly facilitated if there is a river that goes through the border region. When a country emerging from colonial rule has to choose an official language, the language of the colonial power may be the only one that is acceptable to tribes with widely different dialects.

Following Habermas, Elster (1986a: 118) offers a different solution as far as agreement in politics is concerned. Bargaining must be distinguished from attempts to reach agreement by rational discussion. In the latter, manipulation or misrepresentation for a position is not allowed. In rational discussion, the only thing supposed to count is the power of the better argument. It is easy to argue, as public choice theorists do, that ideals of fairness, equality and justice are mere window dressings of self-interest. Elster argues, however, that some norms of justice are more robust than others.

In modern democratic societies the perceived justice of social institutions and policies is a condition for their long-term viability. The importance of the 'justice motive in social behaviour' equals its elusiveness. Like all norms, those of justice and fairness are extremely context dependent in the way they are interpreted and applied. Likewise, the norm of equality can be implemented in very different ways, depending on the reference group - equality with whom? - and the dimension along which people are required to be equal- equality of what?

For major institutional changes an argument from positive justice is needed. Elster (1988b: 324) argues that values like justice, liberty and democracy have been the major proximate causes of social change over the last few centuries. To be sure, the efficacy of these values must in turn be explained by 'material

201

conditions', but this is different from saying that they are themselves material in content.

If a reform is perceived as fundamentally 'just', people will be motivated to endure the costs of transition and the extensive trial and error procedures that may be required before a viable implementation is found. Elster substantiates his case by giving two examples: one is *universal suffrage* and the other is *the rise of welfare state*.

Constitutionalism and Democracy

Another aspect of democratic politics which is of considerable importance is the significance of constitutional constraints. The focus on constitutions is because they are the embodiments of legitimacy in most democratic societies (Elster 1988: 304). It is a truism that constitutional constraints make it more difficult for the assembly or the society to change its mind on important questions. Regarding majority rule, groups no less than individuals are subject to fits of passion, self-deception and hysteria which may create a temporary majority of decisions which will be later regretted (Ackerman 1989; Holmes 1988; Elster 1988a).

The purpose of the constitutional constraints is two fold. One is to guard against the 'irrationalities' by 'self-binding' and 'precommiting' to rules (Hubin 1986: 82-94). The other is that constitutional change is brought about by way of 'justice'. Here again, justice forms a bedrock for evaluating institutional changes and constitutional choices. Elster is influenced by the arguments made by egalitarian liberals like Holmes (1988: 19-58) and Ackerman (1989: 5-22) as against the libertarian liberals like Buchanan. Buchanan's constitutional contractualist position is feared for its conservative implications since it preserves the status quo, making change impossible. The very logic of the self-interested, utility maximising

behaviour of individual creates a moral presumption in defense of how economic and political goods or resources are distributed in the status quo. What individuals possess by way of economic and political goods or resources is what they should possess. In turn this justifies the incentive and political force to preserve the status quo.

A conservative position about collective action starts from the assumed interest of the individual over protecting his existing assets and freedom of action. The very set of assumptions which drive the theory of rational choice creates a strong moral justification and political force for the current distribution of economic and political goods and resources (Petracca 1991: 186). Coupled with its narrow and self-regarding empirical conception of human nature, rational choice theory can be used to justify a government system fraught with political apathy and acquiescence and great inequalities in the distribution of economic goods, political power and privilege (Petracca 1991: 186).

A constitution gains legitimacy from what has been called "the normative force of the factual": compared to the alternatives, which are contending with the status quo and with each other, it has the privilege of existing. Elster (1988b: 326) argues against the view that one may rationally count on being able to achieve full democracy by the temporary abolition of democracy on the grounds that the current generation has rights that set limits to the sacrifices one can impose on it for the benefit of later generations. He argues that on the positive side only democratic processes are feasible and legitimate paths to socialism. His view is that the process must be legitimate in order to be feasible. If the general tendency of socialist reforms is perceived to be fundamentally just, people will have the motivation to ensure the extensive trial and error procedures that will be needed to implement them. So the task of politics is not just to devise institutions that can harness opportunistic self-interest to socially useful purposes but also to create

institutions that embody a valid conception of justice. In other words, is constitutionalism only a tool deployed in the self-interest of the property-holding class? Or are constitutional guarantees for property in the interest of everybody? Against Buchanan's contractarian principle of unanimity and commutative justice, Elster defends the claim that constitutional change can be justified only on grounds of distributive justice (Sunstein 1988: 349).

Drawing the contrast between democracy and constitutionalism, Elster argues democracy is defined as simple majority rule based on the principle of one person one vote. Constitutionalism refers to limits on majority decisions, more specifically to limits that are in some sense self imposed. The limits can take a variety of forms. They can be procedural or substantial. One central rights-protecting element in modern constitutions is the principle of legality: to be punishable, an act must be explicitly forbidden by a law which was in force at the time it was committed. The effect of the principle of legality is to exclude arbitrary punishment. So constitutions serve two overlapping functions: they protect individual rights, and they form an obstacle to certain political changes which would have been carried out had the majority had its way. The latter function is served in several ways: by declaring certain changes as unconstitutional; by making the process of change so complicated and demanding that few proposals will be able to clear the hurdles. The central argument for constitutional constraints on democracy is that without such constraints democracy itself becomes weaker, not stronger.

It is important to recognise these divergent views within rational choice theories for the reason that it can open up possibilities of theoretical conversation with other research traditions. Elster's criticisms sharpen the case for the Habermasian enterprise of discursive democracy.

Interestingly, attempts at a critical and productive engagement with critical theory came from rational choice theorists like Jon Elster (1986a) and also Buchanan and Vanberg (1989) who have shown a great deal of hermeneutic sensitivity in their attempts to explore Habermas' work. As Johnson (1991: 197) says,

> although Habermas does not directly engage in debate with rational choice theorists, other theorists challenge rational choice theory under the banner of communicative action.

From the standpoint of rational choice, the whole problem with Habermas's enterprise is that it is utopian and unrealistic (Buchanan and Vanberg 1989: 49-62). Even Elster (1986b: 103-129) fears that discourse ethics might be dismissed as utopian precisely for its idealist nature. As one can see, this criticism is not just made by rational choice theorists but also shared by liberals like Ackerman (1989) and Holmes (1988).

Unlike Habermas and to some extent Elster, Buchanan (1986: 44-52) argues that there is a potential for 'tyranny' if the enterprise of politics is interpreted as being analogous to that of science as a 'truth' discovery process. On his view, Habermasian dialogue precisely commits this mistake. For Buchanan the truth-discovery claim is only possible in the domain of science. Contrasting *Belief, Science and Truth* and *Interests, Politics and Order* Buchanan (1986 44-52) argues that in its common representation, science is a process in which conflicts about truth are resolved. In it the existence of reality is itself independent of any belief about it: that is to say objective reality exists independent of subjective world views. So the scientific enterprise is necessarily teleological, even when the provisional nature of any established truth is recognised. By comparison and contrast, politics is a process in which conflicts among individual interests are

settled. In this enterprise there is *no* independently existing "interest" analogue to truth, towards which an interaction process converges. The end-states emergent in the two processes remain categorically different. Buchanan argues that when politics is wrongly interpreted as being analogous to science, as a truth-discovery process, coercion may find moral legitimisation for those who claim enlightenment. By contrast, when politics is rightly interpreted as a process for settling conflicts among interests, which are acknowledged to be individually derived, those who seek to impose preferred solutions do so without claim to moral superiority.

However, Buchanan is mistaken if he thinks that the Habermasian enterprise is aiming to discover any transcendental absolute truth. Moreover, in Habermas' case, since the method of discovering truth is by universal consensus it is implausible to see the legitimisation for tyranny in politics.

It remains to be seen how far Habermas' ideal of communicative ethic is viable for a democratic public ethos. The concern for a democratic public sphere stem from his critique of scientism and technocratism which have subverted the scope and domain of an informed public realm. So the ideal of discourse ethics is a critique of this sort of political culture. As Knight and Johnson (1994: 290) argue, what is central in this view for present purposes is that the political order consists of rules and procedures that revolve around the requirement that collective decisions be criticised and defended with reasons. Habermas's view is that the conceptual pressures built into the structure of language induce pragmatic contradiction among those who seek consistently to defend narrowly self-interested policies. Habermas's discourse ethics is an intersubjective or "public" version of Kant's categorical imperative. Insofar as a discourse ethics requires that the validity of social norms be tested in actual discourses rather than through monologically conducted thought experiments, his construal of the principle of publicity emphasises more than Kant the need to encourage and maintain a wide

array of institutions that together constitute an active and robust public sphere. The concept of publicity (Habermas 1991: 196-211) assumes a prominent role in Habermas' work. Thus in Habermas' theory of communicative action, the idea of agreement is located in the structure of communication. The central task of critical social theory is to identify ways in which this ideal can be institutionalised within an actually functioning public sphere.

The aim of discourse ethics is to provide normative grounds for social criticism and the justification of this normative ground is ultimately 'reflexive' or 'recursive' in the sense that there can be no higher appeal to something beyond the idea of that to which free and equal persons can rationally agree. Habermas' claim is that an analysis and reconstruction of the conditions of mutual understanding can provide a normative foundation for social criticism.

5.2 Conceptualising Politics: Towards a Moral-Cognitivist View

As argued in chapter 3, following Benhabib (1989: 149-50), the purpose of the discourse theory of legitimacy can be seen as not to develop a blueprint for social order but to suggest a critical vantage point from which to judge power relations in our societies. Consequently the political utility of this model is to uncover the hidden relations of power and ideology in the social and political realm rather than offer any design for social and political institutions. The representatives of this position include Dryzek (1992); White (1988; 1987); Bohman (1989); Baynes (1990); and Benhabib (1986; 1989).

The political implications of Habermas' programme are explicitly argued in his earlier works *Theory and Practice* (1973: 41-81) and *The Structural Transformation of the Public Sphere* (1991: 181-235). Habermas begins his analyses by tracing the conceptual history of the category of the 'political'.

Aristotle saw politics as being continuous with ethics, the doctrine of the good and just life. As such, it referred to the sphere of human action, *praxis*, and was directed to achieving and maintaining an order of virtuous conduct among the citizens of the polis. Habermas argues that the practical intention of politics, as well as the nature of its subject-matter, determined its cognitive status: politics could not assume the form of a rigorous science, of *episteme* but had to rest content with establishing rules and norms. The capacity thereby cultivated was *phronesis*, a prudent understanding of variable situations with a view to what was to be done.

> With the rise of modern science, the classical conception of politics was drastically altered. Theory came to mean the logically integrated systems of quantitatively expressed, lawlike statements characteristic of the most advanced sciences. Given their description of the relevant initial conditions, such theories could be used to predict future states of a system; providing the relevant factors were manipulable, they could be used to produce states of affairs. Adopting this ideal of knowledge of politics, Hobbes outlined a program that took human behaviour as the material for a science of man, society, and the state. On the basis of correct understanding of the laws of human nature it would be possible to establish once and for all conditions for a proper ordering of human life. The sphere of the practical was absorbed into the sphere of the technical; the practical problem of the virtuous life of the citizens of the polis was transformed into the technical-administrative problem of regulating social intercourse so as to ensure the order and well-being of the citizens of the state. To this extent both scientific-technical and economic-administrative apparatus became dominant in the social integration (McCarthy 1979: viii).

Habermas' insistence is that *praxis* cannot be reduced to *techne*,[3] nor rationality to instrumental rationality. Political emancipation cannot be equated with technical progress. And rationalisation in the dimension of social interaction is not necessarily instrumental. Habermas attempts to base this form of social integration in communicative action which is a moral practical insight of

determining the intersubjective norms. Habermas' aim, influenced by Weber, is to recognise the significance of values and norms in shaping the world. Politics, in its discursive mode, is a creative process of self-realisation responsible for redefinition of worldviews. Habermas pointed to the difficulties of combining praxis and techne. Rationally organised work can do no more than generate products. It does not produce truth or intersubjectively valid norms. Habermas aim' is precisely to maintain this distinction between economy and politics, or labour and interaction or production and language. Social consciousness is independent of the economic realm.

Since discourse ethics is cognitivist in the sense that it presupposes certain abilities and capacities among the participants, one can draw moral and political conclusions from such program. Habermas's discussion of democracy scattered throughout his writings is cast primarily in terms of the sociological concept of legitimation (Habermas 1975a; Bronner 1994: 284), rather than in terms of normative political theory or philosophy. But because Habermas interprets legitimation in a normative and cognitivist manner, it should not be too hard to reconstruct the normative political theory and the relation between 'knowledge' and 'politics'.

Habermas' understanding of politics takes a 'cognitivist turn' by the fact that he considers the recognition of the legitimacy of a political order to be a matter of 'epistemic judgment'. One could as well say that he is an epistemic populist (Bronner 1994: 284). On the cognitivist view, a group must collectively manifest in its members' social relationships certain abilities of speaking and acting necessary in order to establish and reproduce a form of life with a democratic political order.

Habermas' endeavour is to establish a 'minimal cognitivist' ethics and politics on which institutions and practices can be based. This claim can only be defended if one transforms two basic categories of the modern moral and

epistemological tradition, 'autonomy' and 'competence', in terms of critical social theory (Benhabib 1986: 282). On the one hand Habermas argues that enlightenment should be criticised only when it is drained of its cognitive and normative content, when its institutions are no longer a means of achieving the emancipatory goals of social learning and justice, but he doesn't offer any political realisation of his ideal except for an abstract discursive democracy. Habermas sees an intimate relationship between 'knowledge' and 'politics' as well as between 'epistemic justification' and 'social criticism'. Habermas's cognitivist interpretation of democracy is an attempt to revise and improve the enlightenment by means of critical social theory, in order to let it better fulfil its emancipatory intentions. Most often critiques charge Habermas' theory as utopian and unrealistic. But if one carefully analyses Habermas' idea of the public sphere, discourse assumes prominence in the institutional realm of civil society constituted of associations.

Since Habermas transforms epistemology from a foundational discipline into a normative theory of social learning (Habermas 1984; 1987; 1979), the concepts of reason and knowledge which emerge from this theory can be applied to politics, in order to explicate the moral and cognitive presuppositions of a just democratic order. The normative force of these epistemic contents can then become the basis for a critique of ideology and power in the modern state. Thus, when both are normatively interpreted, 'knowledge' and 'politics' are interrelated: under certain general and specifiable conditions, politics can become a process of social learning directed to human emancipation. Such a rich and multifaceted concept of emancipation is the goal of Habermas's attempt to rescue, by means of a thoroughly cognitivist interpretation of democracy, the enlightenment's connection of knowledge and politics.

Just as the consequences of one-sided conceptions of enlightenment reason have become clear in this century, so too have the consequences of the non-

cognitivist reactions against the enlightenment. Shorn of the enlightenment's moral and epistemic commitments, democracy takes one of the two forms: either it becomes a way of aggregating individual preferences or it becomes a mere legal procedure, reducible to something like the counting of votes in majority rule or the rule of law in liberalism.[4] In either case, it can become irrational: on such impoverished interpretations, the theory of democracy typically ends up in the familiar paradoxes of the tyranny of the majority. These results lead some to reject democracy entirely as one more myth of the enlightenment; for others it is to be restricted so as to eliminate its irrational tendencies, as in Riker's recent proposal to limit the sovereignty of majorities to getting rid of ruling elites at periodic intervals.

Bohman (1989: 266-7) argues that given the obvious failures of concepts of the 'will' and its preferences, only one option seems to remain open: democracy must once again be given an epistemic or cognitive dimension. Its results and institutional structures must be judged by epistemic, rational standards appropriate to its practical domain. This is precisely the task of Habermas' interpretation of democracy as a process of communication and social learning.

The connection between competencies and abilities on the one hand and autonomy and emancipation on the other has a lot to do with what reaching a democratic decision actually entails. In truly democratic order, participants must be able to determine the rules, values, ends, means and institutional procedures of their own association with each other, rather than accept as given those that have been handed down by tradition or convention. The cognitivist interpretation of democracy specifies the capacities and competencies presupposed in a democratic order. For Habermas, democracy is based on abilities of social cognition, self reflection and moral reasoning, the same cognitive abilities on which the social scientist depends as a participant in dialogue. Unlike the social scientist, whose abilities are manifested in adequate theories, the political achievement of these

cognitive abilities is manifested in the participation and successful reproduction of a social structure and type of communication which Habermas calls practical discourse. That is, democracy is the political expression of a set of general abilities required for complex interactions, including the ability to remove oneself from ongoing social life and to make public, shared judgments of epistemic adequacy based on a common set of norms and interests. The point of a cognitivist interpretation of democracy is to specify and analyse the formal and epistemic conditions for the possibility of participating in such a practice of discourse.

Habermas focuses his attempt to retain the normative rather than the foundational status of epistemological notions like truth and reason in an empirical research program he calls reconstructive theory. The aim is to develop a modest, fallibilistic, empirical and yet universal account of the rationality manifested in human activities. Normative social theory as envisaged by Habermas retains the fundamental task of epistemological reflection in supplying standards and criteria of rational adequacy and justifiability. For Habermas, the various competences represent aspects of an emerging form of reason in different, irreducible domains-what he calls discursive rationality. As the highest exercise of the abilities which are only acquired socially in interaction and collective learning, it may be called communicative competence.

From the standpoint of Habermas, democracy is not itself a particular competence, but a method of *institutionalising learning* in such a way as to politically protect and enhance it. Democracy is therefore a practical hypothesis about how a collective will may be formed in public processes of deliberation. In its institutions, a collective will ties the development of social systems to control through a politically effective institutionalisation of discourse. If this is the effect to be achieved, the proper goal is not merely to secure individual liberty, but to establish intersubjective social structures free from domination, violence and self-

deception; for this to occur, the topic under discussion must be conscious, rational decisions about the moral structure of society itself. As evident in most modern constitutions, the structure of decision-making itself must be reflective and open-ended; the scope of the issues which this process considers cannot be limited *a priori*, but is itself an issue for revisable, public deliberation.

Habermas's discussion tends to be more on the level of practical discourse in general than of the political structure of democracy. In his idea of the public sphere, the notion of publicity itself takes on a normative rather than a merely functional significance and one permitting the formation of public opinion in the normative sense of a 'general interest' and not in the factual sense of the result of an opinion poll.

Habermas identifies a number of basic features of post conventional moral reasoning which must enter into a similar description of collective practical judgments in democracy. These features are the hallmarks of any modern cognitivist ethics, whose universal norms extend to the political sphere as to any other domain of action. Once these abilities for moral-practical reflection are described, they can be applied to democratic processes and political obligations in two crucial ways: first, Habermas' social reinterpretation of the categorical imperative supplies a distinct understanding of how universality is achieved in collective deliberation; and second, a comparison of the critical, reflective abilities which constitute the basis of both democratic participation and social scientific analysis shows how the independence of standards of rational consensus may be maintained through self reflective critique of ideology. Social scientific theoretical knowledge will have an important place in democratic order. As both independent of and yet continuous with participants' knowledge of the social order, it can have a public, enlightening role in uncovering systematic self-deceptions and ideological and distorting influences on democratic structures of communication.

Unlike the early modern emphasis on science, Habermas finds "the rational content of modernity" in its new possibilities for moral and political identity. In modernity, moral and political reasons must have the following epistemic characteristics in order to motivate highly competent actors who have undergone a basic learning process that devalues all traditional moral and religious identities. Their abilities of moral reasoning extend to the following dimensions: reflexivity and universality

There are numerous necessary but not sufficient conditions for agreements with these epistemic and formal characteristics. Some are related to the formal conditions of the process of communication itself. To analyse these communicative conditions, Habermas has constructed a counterfactual model ideal speech situation. The role of this model has been both misunderstood and overestimated in discussions of Habermas's view of politics. It serves the analytic purpose of thematising issues of the structure of communication related to participation in the processes of reaching agreement. All democratic institutions must more or less fulfil these conditions, or else the agreements reached could be problematised as forms of pseudo or forced consensus. For example, if an agreement were reached under conditions in which all speakers are not given equal chances to speak, equal chances to utter any of the different types of utterances, or equal opportunities to adopt any role in interaction, then the consensus so reached could have been formed by extra-discursive means. Habermas' point is to provide a counterfactual normative standard for the ideal conditions of reaching agreements in communication; such a standard serves as an independent test of any de facto agreement. These formal norms of communication and interaction in processes of reaching agreement are more inclusive and specific to the actual structure of democratic decisions than the usual interpretation of simple procedures, such a majority rule, or one person, one vote.

214

If these interdependent formal and cognitive conditions are met, then the result of the collective process of deliberation and argumentation will express a socially effective rational will based on a communicatively shared and established general interest:

> the discursively formed will can be called rational, because the formal properties of discourse and of the deliberative situation sufficiently guarantee that a consensus can arise through appropriately interpreted, generalisable interests (Habermas 1975a: 108).

Unlike Kant's view of moral reflection, in collective deliberation no needs are *prima facie* excluded until after collective deliberation is concluded and the generalisable interests are formed. But this general will cannot be determined in advance; it is only in interaction in institutionalised discourse that such interests are formed and discovered. If this were not so, no procedure could rescue democratic discussion from the intractable problems involved in ordering already fixed individual preferences.

> Discursively redeemable norms or generalisable interests have a non-conventional core; they are neither merely empirically found nor simply posited; rather they are, in a non-contingent way, formed or discovered (Habermas 1975b: 177).

Democracy becomes the political form of this communicative process, where everyone affected by the decision must be party to the agreement, be able to recognise the validity of his judgments, and form this judgment and will in a collective process.

Rousseau already recognised that not everything can be decided by the "general will", although he restricted the scope of issues to a greater extent than Habermas's interpretation would warrant. For Habermas the question is not to

distinguish the general will from the will of all; it is rather, to determine what sorts of issues can or cannot be settled in a consensual manner. In this regard, Habermas recognises that there are limitations to practical discourse that are not present in theoretical discourse. In political institutions, participants are actors as well as speakers, so in this sense science does not offer an apt analogy for democracy.

Given the communicative interpretation of democracy as a process of discursive will formation, one can locate a whole range of potentially effective ideologies that might restrict communication even in these relatively transparent and self reflective institutional settings. In the modern world, ideologies become primarily what Habermas (1970b) calls "distorted communication"; that is, they need to be analysed as barriers or structural restrictions on social processes of communication, which in modern conditions become significant primarily in discursive political contexts and in everyday interaction.

The first sort of ideology has to do with the formal level of communication in the political process. Given Habermas' definition of democracy as an institutionalisation of discourse, it can be inferred that the basic formal characteristics of democratic institutions and procedures, such as equality, must now be recast in communicative terms. All the basic formal conditions of communication related to processes of reaching agreement are contained in the counterfactual construction of the ideal speech situation, so that any violation of these conditions is inconsistent with the thoroughgoing democratic participation of all those affected. There may be different types of violations on this level. They may be internal to the process of communication itself, such as in the failure of the institution to create a framework sufficient for all to have an equal chance to be heard and to participate.

As Bronner (1994: 307-310) argues, material social conditions may also influence communication. For example, if participants enter the discourse with

216

large scale pre-existing inequalities of wealth and power, the difficulty in reaching a rational consensus will rise in the same proportion as the degree of inequality. Habermas thinks this is because such material conditions allow the privileged actors to act strategically; in any case they distort the formal requirements of the internal structure of communication in a process of consensus and create fundamentally conflicting aims between economic and political institutions. Capitalism, Habermas has always insisted, is inconsistent with democracy, since it organises society non-democratically. Its structural inequalities delegitimise any existing democratic structures by having them take on functions with regard to economic conditions.

These same 'material conditions' also may affect the 'cognitive conditions' of participation for certain actors, since it is surely the case that the acquisition and exercise of certain politically significant competences requires access to social goods.

In public choice theories, democracy is interpreted as a device for preference aggregation of individual preferences under conditions of non-cognitivism, instrumentally rational behaviour and self-interest. Non-cognitivism means simply that values and preferences are like emotions, beyond the reach of rational argument (Dryzek 1991: 406). It is this self interest or non generalisable interests that does the bulk of the damage. It also means that individuals cannot escape these problems by subjecting their preferences to rational scrutiny and possible adjustment in the interests of determinate collective choices. The distinction between facts and values has been the departure for a non-cognitivist treatment of practical questions. The conviction is that moral controversies cannot be decided with reason because the value premises from which moral questions are inferred are irrational. From the standpoint of critical theory, public choice theories articulate a non-cognitivist theory of politics.

The cognitivist theory of democracy sets out general cognitive conditions for the formation and discovery of collective and individual autonomy, which particular societies may violate or fulfil. Habermas sees more in the rationality of collective decision than simply relying on individual judgments; he does judge any consensus by cognitive standards, lest he fall back into the paradoxes of pure proceduralism which plague such formal-democratic rules like 'one person one vote'. Discursive will formation must issue in a general will, and not merely in a procedurally correct decision; Habermas' contribution is not only to understand the formation of the general will in communicative terms, but also to analyse the cognitive presuppositions for successful participation in processes of institutionalised discourses.

The model of a public sphere envisioned by Habermas is thus by no means limited to proposals for reforming the traditional institutions of parliamentary debate. Rather the public sphere must be broadly conceived as a vast array of institutions in which a wide variety of practical discourses overlap. On this model, the moment of deliberation, so to speak, does not reside primarily with the judiciary nor with the body of elected representatives, but is dispersed throughout a vast communicative network.

Towards a democratic public sphere

In conclusion, I shall elucidate the political function of the Habermas' model of public sphere. The importance of the public sphere lies in its potential as a mode of social integration. Public discourse is a possible mode of coordination of human life, as are state power and market economies. But money and power are non discursive modes of coordination; as Habermas' later theory stresses, they suffer from tendencies toward domination and reification.[5] Thus state and economy

are both rivals of the domination of the public sphere. The public sphere remains above all a critical principle of public life. Habermas aims to recover the enduringly valuable ideal of the public sphere from its historically partial realisation (Calhoun 1992: 4).

The liberal public sphere emerged in the specific historical circumstances of a developing market economy. Habermas defines it as a sphere between civil society and the state, in which critical public discussion of matters of general interest was institutionally guaranteed. In its clash with the bureaucratic practices of the absolutist state, the emergent bourgeoisie gradually replaced a public sphere in which the ruler's power was merely represented before the people with a sphere in which the state authority was publicly monitored through informed and critical discourses by the people.

Habermas traces the complex trajectory of the development of the literary and political self-consciousness of this new class with the rise of the novel and of literary and political journalism and the spread of reading societies, salons and coffee houses. At the same time this also meant the crystallisation of the contradiction between the liberal public sphere's constitutive catalogue of "basic rights of man" and their de facto restriction to a certain class of men. And with the further development of capitalism, the public sphere expanded beyond the bourgeoisie to include groups that were systematically disadvantaged by the workings of the free market and sought state regulation and compensation. The consequent intertwining of state and society in the late nineteenth and twentieth centuries meant the end of the liberal public sphere.

The public sphere of the emergent social welfare state democracies is rather a field of competition among conflicting interests, in which organisations representing diverse constituencies negotiate and compromise among themselves and with government officials, while excluding the public from their proceedings.

219

The press and broadcast media serve less as organs of public information and debate than as technologies for managing consensus and promoting consumer culture.

While the historical structures of the liberal public sphere reflected the particular constellation of interests that gave rise to it, the idea it claimed to embody - that of rationalising public authority under the institutionalised influence of informed discussion and reasoned agreement - remains central to democratic theory. In the post-liberal era, when the classical model of the public sphere is no longer sociopolitically feasible, the question becomes: can the public sphere be effectively reconstituted under radically different socioeconomic, political and cultural conditions?

The public sphere can be conceived above all as the sphere of private people coming together as a public; engaging in a debate over the general rules governing relations in the basically privatised but publicly relevant sphere of commodity exchange and social labour. The medium of this political confrontation was peculiar and without historical precedent: people's public use of their reason.

> Public sphere was constituted by private people. The private sphere comprised civil society in the narrower sense, that is to say, the realm of commodity exchange and of social labour; imbedded in it was the family with its interior domain (Habermas 1991: 30).

The typical illustration of publics would be the arts (theatres, museums, paintings, literary associations, music, poetry, journals, reading societies so on). The publics constituted organised discussion among private people that tended to be ongoing. Their historical effect in the long term was political: to prise the high decisions of the state out of their conventions of secrecy and subject them to the open rational scrutiny of the public sphere of discoursing individuals. The common

institutional criterion was discussion of the problematic areas which until then had not been questioned.

However, Habermas is aware of the inadequacies of the bourgeois public sphere. The bourgeois public sphere was based on the fictitious identity of the two roles assumed by the privatised individuals who came together to form a public: the role of property owners and the role of human beings as pure and simple (Habermas 1991: 56).

Habermas shows the disintegration of the public sphere with the emergence of advanced late capitalism. Thus he argues

> In the hundred years following the heyday of liberalism, during which capitalism gradually "organised," the original relationship of public and private sphere in fact dissolved; the contours of the bourgeois public sphere eroded. But neither the liberal nor socialist model were adequate for the diagnosis of a public sphere. Two tendencies dialectically related to each other indicated a breakdown of the public sphere. While capitalism penetrated more spheres of society, the public sphere lost its political function, namely: that of subjecting affairs that it had made public to the control of a critical public. The principle of the public sphere, that is, critical publicity, seemed to lose its strength in the measure that it expanded as a sphere and even undermined the private realm (Habermas 1991: 140).

The public sphere as a functional element in the political realm was given the normative status of an organ for the self articulation of civil society with a state authority corresponding to its needs. The social precondition for this developed bourgeois public sphere was a market that, tending to be liberalised, made affairs in the sphere of social reproduction as much as possible a matter of private people left to themselves and so finally completed the privatisation of civil society (Habermas 1991: 74).

With codifications of civil law a system of norms was developed securing a private sphere in the strict sense, a sphere in which private people pursued their

affairs with one another free from impositions by state (Habermas 1991: 75). The functions of the public sphere were clearly spelled out in the law. A set of basic rights concerned the sphere of the public engaged in rational critical debate (freedom of press, freedom of opinion and association etc.) and the political function of private people in this public sphere (right of petition, equality of vote etc.). A second set of basic rights concerned the individual's status as a free human being, grounded in the intimate sphere of the patriarchal conjugal family (personal freedom, inviolability of the home etc.). The third set of basic rights concerned the transactions of the private owners of property in the sphere of civil society, equality before the law, protection of private property, etc. (Habermas 1991: 83).

As a consequence of the constitutional definition of the public realm and its functions, publicness became the organisation principle for the procedures of the organs of the state themselves; in this sense one spoke of their publicity (Habermas 1991: 83).

The political public sphere of the social welfare state is marked by two competing tendencies. Insofar as it represents the collapse of the public sphere of civil society, it makes room for a staged and manipulative publicity displayed by organisations over the heads of a mediatised public. On the other hand, to the degree to which it preserves the continuity with the liberal constitutional state, the social welfare state clings to the mandate of a political public sphere according to which the public is to set in motion a critical process of public communication through the very organisations that mediatise it (Habermas 1991: 232).

Thus for Habermas it also led to the disintegration of that specific portion of the private realm within which private people, assembled to constitute a public and to regulate those aspects of their commerce with each other that were of general concern, namely, the public sphere in its liberal form. The downfall of the public sphere, demonstrated by its changing political functions, had its source its

the structural transformation of the relationship between the public sphere and the private realm in general (Habermas 1991: 142- 143). To this repoliticised social sphere the distinction between public and private could not be usefully applied.

The attempt to relieve the public sphere of the intrusion of private interests failed as soon as the conditions under which the privatisation of interests was to be accomplished were themselves drawn into the conflict of organised interests. And the political public sphere tended to adopt the interests of civil society as its own. Interventionism had its origin in the transfer onto a political level of such conflicts of interest as could no longer be settled within the private sphere (Habermas 1991: 142).

Reframing the argument in terms of the rationalised life world, Habermas draws a distinction between instrumental rationality and communicative rationality. Habermas argues that the symbolic life world is colonised by two mechanisms of money and power. The life world is subjugated to alien standards of technical control.

Limitations of the Concept of Discursive Democracy

Discourse theory is not without limitations. Rational choice theorists might argue that the process of dialogic exchange would not necessarily preclude the occurrence of paradoxes of collective choice, since differences of judgment might remain after debate and dialogue, leaving 'voting' as the only practicable way of overcoming disagreement. But from the standpoint of critical theory, voting does not exhaust democratic practice. What might advocates of deliberative democracy learn from public choice theories' critique of voting? Most obviously they might acquire a subtler understanding of the internal difficulties of aggregative arrangements.

Knight and Johnson (1994: 279) argue that many are skeptical of the democratic institutional mechanisms that rely solely or even primarily on electoral mechanisms, i.e. on ways of aggregating individual interests or preferences. Knight and Johnson (1994: 279) complain that aggregation is not adequate to the task of producing normatively binding political outcomes. They insist that aggregation needs to be supplemented and perhaps entirely supplanted by institutional arrangements that embody and enhance democratic deliberation. However Knight and Johnson (1994: 279) point to internal difficulties to the ideal of democratic deliberation: there is the problem of how deliberation might be institutionalised. The problem of collective action, prisoner's dilemma need to be resolved through institutional mechanisms and no amount of interpersonal dialogue is possible in large communities which need to take large decisions. Even deliberation can itself be susceptible to distortions.

As Sen (1986) says social choice problems are not so simple that they can be overcome through dialogue. Idealised deliberation is seen as a critical standard from which to assess existing institutional arrangements. But the prospects for moving existing arrangements toward the standard set by this idealised scheme are debatable and Knight and Johnson (1994: 287) argue that such movement would require a highly conflictual political struggle and moreover the outcome of that struggle is highly uncertain. Again it would depend on the vagaries of strategic advantage and in that sense, could not be justified simply by reference to the process by which it was brought about.

The case of deliberative democracy is not without problems and internal difficulties. It can be charged that Habermas' discourse ethics is realised only in small and relatively homogeneous societies rather than larger democratic societies wherein the only way of political expression is through 'voting'.

So the question is how far deliberative democracy is realisable. Democracy is a procedural norm not a theory of justice. Some see democracy as aggregation of preferences and some others see it as deliberation. Defenders of socialism who see discursive democracy as a means to achieve socialism down play the role of majority rule and democratic procedures (Bronner 1994; Bohman 1989). Instead, they uphold the ideal of the rational consent of all, following ideally extended democratic debate among competent and conscientious citizens who are striving to achieve the common good.

The advocates of the discourse-based conception of socialism are of the view that in the ordinary sense of the term democratic society could be as democratic as you like, yet thoroughly unjust. The ideal is recognisably drawn from the thought of Jurgen Habermas. The ideal of deliberative democracy shades over into the ideal speech situation ethic, the idea that any ethical claim implicitly supposes that it could be redeemed in an ideal deliberative procedure in which no force except the force of the better argument prevails. If sufficient stress is laid on the idea of rational consent to rules directed toward the common good, the possibility of conflict between majority rule procedures and the attainment of substantively just outcomes can be eliminated.

But the very idealised democratic conception at this lofty level of abstraction is then left virtually without links to any recognisable notion of socialism in the sense of public ownership. What has happened is that the ideal of deliberative democracy, so idealised, has become a hypothetical-rationality test, roughly a restatement of the Rawlsian ideal of wide reflective equilibrium. But this is a purely formal notion that has in itself no substantive ethical or political content. Neither would it be wise to try to do as much as we can to create circumstances in which the conditions of free, rational dialogue among all citizens are approximated

as fully as possible. Critics argue that to try to create a free public realm, rational dialogue end up recreating the unholy chaos of the tower of babel.

Advocates of deliberation claim that it somehow mediates or transforms rather than simply minimises or accommodates conflict. And they argue that, in the ideal case at least, it will issue in consensus. This is because deliberation has the effect of transforming the substance of participant's preferences. This might take the form of exposing and revising objectionable preferences or of inducing reflection and consideration of the grounds for holding otherwise unobjectionable preferences. In either case, deliberation involves changing preferences. Deliberation might transform preferences. By contrast, rational choice theorists consider preferences as normatively fundamental and hence immune to revision.

Conclusion

In this chapter, I have tried to show that rational choice theorists articulate a noncognitivist theory of democracy. The objection is that just as the deficiencies of one-sided conception of enlightenment reason have been explicated by Habermas, so too democracy interpreted as mere preference aggregation is devoid of any moral-cognitive content. Habermas' discourse ethics, on the contrary, appeals to not just instrumental motivations but also moral-cognitive ones.

From the standpoint of critical theory, the embodiment of reason in the political realm means the establishment of a republican form of government with guarantee of civil liberties and an institutionally secured public sphere, so that political power could be rationalised through the medium of public discussion to reflect the general will or common good. What is central in this view for present purposes is that political order consists of rules and procedures that revolve around the requirement that collective decisions can be criticised and defended with

reasons (Johnson 1994: 290). Critical theorists recognise that societies are moral and ethical communities and not just comprising of atomistic self-interested individuals.

For Habermas, the discussion of political issues is inseparable from moral questions. In this respect he stayed close to a position that he developed in *The Structural Transformation of the Public Sphere* (1991). Habermas seeks to locate a ground where political conflicts can be addressed rationally within the context of definable norms, which are open to critical scrutiny. What becomes more interesting is that unlike his predecessors' wholesale dismissal of the political rationale of liberalism, Habermas, by contrast, is cautious and recognises that the effectiveness of the ideal of discourse ethics depends on the prior institutional realisation of liberal democratic public sphere. To put it in Bronner's words (1994: 285), Habermas is a brilliant theorist of liberal democracy.

One of the persisting problems with critical theory is that unlike rational choice theory the empirical referents are blurred. In claiming that discourse ethics is merely procedural and formal, the theory can be inexact and ambiguous if its empirical referents are not well specified. At least on this account rational choice theory is better off since it directly addresses empirical problems. Critical theory does not address institutional questions. Habermas' theory is equally indeterministic insofar it evades the question of substantive issues.

Critical theorists see the rational choice enterprise as excessively concerned with analytical techniques at the expense of social values. By contrast, discourse ethics fails to address the question of *institutionalising* deliberative democracy. To say that the public sphere functions as a platform for democratic public ethos does not solve the problem of institutionalising discourse. Although both theories are critical of majoritarian democracy as a procedural ideal of democracy, they offer radically distinct solutions. The strength of rational choice theory lies in resolving

problems at the institutional level and so a reference point is always maintained anchoring on to the procedures of democratic institutions. On the other hand, discursive democracy is so abstract that it fails to realise in any institutional mechanisms. As Dryzek (1992: 397) argues, critical theory lacks any empirical content for politics and ends up being only an ideal.

Notes

[1] Elster's argument is that individual preferences are susceptible to weakness of will or excess of will. Moreover expressed preferences cannot be assumed to represent the real preferences since according to 'concealed preference hypothesis', choices often conceal rather than reveal underlying preferences.

[2] Elster challenges the idea of 'given' preferences. In his works Elster (1983a; 1989a) argues that individual preferences cannot be taken for granted. It is important to see the complexity of the formation of preferences. Moreover individual preferences come under the influence of mechanisms like cognitive dissonance reduction, irrationalities like weakness of will and excess of will etc. He argues that to explain intentions causal explanations need to be invoked.

[3] This distinction between *praxis* and *techne* goes back to Aristotle and is recently invoked by not Habermas but many continental philosophical variants like Gadamer and Arendt. The significance of this conceptual distinction for contemporary theorists lies in its critique of instrumental-technical rationality.

[4] This is clearly seen in public choice and social choice interpretation of democracy. Their inerpretation of democracy is non-cognitivist in the sense that preferences are beyond the reach of rational argument.

[5] In his later work *The Theory of Communicative Action* Vol. II, Habermas develops his thesis of colonisation of lifeworld. Lifeworld is distinguished from both state and market as potential threats for domination and power.

The thesis is a critique of the concept of 'instrumental rationality' in social sciences. For a long time the debates and controversies on this theme centered mainly at the level of philosophy of social sciences. For instance, the dispute over "methods in social science" has been the most contentious issue among different theoretical paradigms (Bernstein 1983; Held 1980). Since my previous research involved the study of interpretive theories and positivism at the level of the methodology of social sciences, my focus in the present study is exclusively concentrated on political theory.

I began my thesis by defining the theoretical dilemma for political theory, namely, the problem of collective choice. Hence, the central problem of defining the *rationality* of collective choice in politics and its *reasonableness* form the subject-matter of the thesis. Two theoretical traditions offered different explanations: one is rational choice theory and the other is critical theory. Rational choice theory modelled collective choice as an interaction involving instrumentally rational individuals. Alternatively, critical theory sees interaction as discursively constituted. Rational choice theorists exemplify the Hobbesian tradition, involving pure self-interested individuals, whereas Habermas' communicative reason is a quasi-Kantian conceptualisation of social interaction.

In the social sciences, economics was the first to incorporate the tools of mathematics and other scientific methods to further the ambitious task of 'scientising' the discipline. The reasonable success of the discipline made other social science disciplines adopt these methods. This trend towards precision and prediction has become a burden on the social sciences in the recent years. This urgency for the scientific status of the discipline is unprecedented in the history of ideas (Przeworski 1985: 379). Consequently one such theoretical paradigm

claiming to be scientific and precise in its investigations of the problems of social and political reality became prominent in the late 1970s and 1980s. This is the rational choice paradigm. Its claim for the 'scientificity' and 'universality' of the discipline led to employing the tools of economic methods for the study of politics. Since the time of its introduction into political science, rational choice theorising tended towards contradictory impulses: an interdisciplinary spirit that seeks to unify social science explanation and a parochial tendency to interpret all social phenomena through the lens of micro economics (Green and Shapiro 1995: 203).

By contrast, the theoretical programme of critical theory emerged as a challenge to instrumental reason. Anchoring his theoretical project of modernity in the distinction between strategic rationality and communicative rationality, Habermas provided a theoretical basis for critical theory in his theory of communicative action. Moving away from the philosophical preoccupations of previous critical theorists he redefines critical theory as a critical social theory. Thus from Habermas's standpoint the theory of communicative action is not seen as 'meta-theory' but the beginning of a social theory concerned to validate its own critical standards (Habermas 1984: xxxix).

The theory of communicative action, building on Kant, offers a broader interpretation of rationality. It connects three spheres of rationality: the cognitive-instrumental, the moral-practical, and the aesthetic-judgmental. Thus rationality conceptualised in language becomes a significant attempt to further the struggle for political liberalism. The theory of communicative action does not posit a self-sufficient subject confronting an object, but begins instead with a symbolically structured notion of everyday life in which reflexivity is constituted.

The thinking of Habermas is essentially devoid of romantic impulses. He is sharply critical of his predecessors who equate science with the domination of nature. Habermas entertains the validity of scientific reason within delineated

boundaries and thus, following Kant, draws a distinction between the subject-centered reason underpinning instrumental rationality and speculative reason or reflexivity. The new pathologies of the life-world fostered by instrumental reason are consequently only one aspect of modern society.

If interaction forms the essence of social life then inevitably we must theorise the different modes of interactions in social and political life. The subject matter of social sciences unlike that of natural sciences is complex in many ways. The complexity of social reality makes the task of defining the specificity, scope and limits of the logic of rationality a difficult task. So any theoretical paradigm has to take into consideration the unpredictable nature of social reality. The study at the level of theoretical traditions also has advantages in so far as it opens up possibilities of not only theoretical insights but also practical questions concerning the legitimacy of democratic politics.

Here in the conclusion I will assess

1.   Firstly, the respective merits and demerits of the two competing conceptions of rationality.

2.   Secondly, internal critiques of rational choice theory like Jon Elster who have extended the critical theorists' basic intuition that strategic rationality cannot adequately sustain social and political interaction by way of a critique of the concept of preference formation.

Critical theory and Rational Choice theory share both overlapping concerns and parallel theoretical weaknesses. Although the works seem to have no immediate connection, there is an underlying common theme running through all of them. The concern is with rationality and the reasonableness of democratic consensus in politics. The issue involves the normative grounds for resolving disputes or conflict of interests. In other words, both are concerned with determining what 'rational' can mean in social and political interaction. Critical

231

theory's expanded conception of rationality can help rational choice theory in resolving the problems of aggregation and collective action dilemmas. On the other hand, critical theory needs to be open to rational choice and revise its protrayal of rational choice as aiming to build an orthodox positivistic social science.

The confrontation is between choice and dialogue; interests and norms. Both analytical and continental theories are concerned with fairness and justness of collective choice. What are the criteria for judging the normative validity of collective choice mechanisms? Is it choice or dialogue? However, both paradigms can be enriched in their scope of explanation only if they open up and engage in conversation with one another. The theoretical investigations can be extended in numerous ways provided there is a sufficient degree of refinement in the theoretical postulates of individual rationality. And since politics is a realm of collective choice par excellence, this gives scope to examine the issues of not just collective choice but also power, domination and justice. Both theories proceed with the notion of the 'individual' as a basis of analysis. Rational choice theories have shown the logical inconsistencies in democratic mechanisms such as voting, public goods, collective choice problems.

Rational choice theory discovered two classes of findings: one is the problem of preference aggregation and the other is collective action problem. Both problems illustrate the conflict between individual rationality and social rationality. Consequently rational choice theorists tried various solutions to overcome this theoretical dilemma. Solutions range from contract to spontaneous order. Unlike rational choice, Habermas' model of collective choice is conceived not as a process of preference aggregation, in which there is a mapping from a set of individual orderings to a social ordering, but as a process of dialogue in which reasons are exchanged between participants in a process that is perceived to be a joint search for consensus. In this model, preferences are contested, challenged, altered,

modified and subjected to criticism. So in Habermas' framework, the aim of politics is not aggregation of preferences but also transformation of preferences through public discourse. Thus Habermas' theory offers a broader typology of motivations, which include normative, expressive, critical and not just instrumental ones.

With regard to second problem (the problem of collective action), critical theory can help resolve collective action dilemma. If politics involves resolving collective choice problems then the question immediately involves the normative grounds for assessing the fairness and justness of collective choice mechanisms. Although in many respects State guarantees an ineliminable solution to collective action problems, there are many spheres of social life wherein the State fails to intervene and in most cases the intervention of the State leads to counterproductive outcomes. Since politics is not only concerned with framing rules for social and political interaction it becomes important to define the fairness of the conditions under which interaction takes place. This is clearly seen in collective action wherein game theorists have failed to come up with unique solutions. Game theory demonstrated how strategic interaction in so stark an environemnt generates indeterminacy. It thus makes conspicuous the need to identify additional, non strategic mechanisms to coordinate social and political interaction. In this respect critical theory identifies just such a mechanism - the binding force of validity claims raised in communicative action. Thus 'language' forms the mechanism for coordinating social and political interaction.

Thus by engaging in conversation with critical theory, rational choice theory could abandon its hardcore programme of *homo economicus* and appeal to a broader typology of motivations in order to better explain and model social interaction. In this conjunction, rational choice gains conceptually in that its individuals become free to choose how to act (be it dramaturgically, communicatively, in conformity with social norms or even instrumentally). Thus by

invoking a wide variety of motivational assumptions, rational choice theory can better explain co-operation in prisoner's dilemma situations.

Interestingly, rational choice theorists have themselves been willing to abandon economic subjectivity. For example, Brennan (1989) introduces ethical motivation in political behaviour. And as discussed in chapter 3, Vanberg and Buchanan (1989) allow discursive reason in constitutional choice. Other internal critics like Green and Shapiro (1995) have shown the pathologies of rational choice theory when it comes to explaining collective action. Most importantly, Jon Elster professes sympathy with the fundamental tenet of Habermas' discourse ethics. Elster invokes social norms to supplement instrumental rationality. This move towards incorporating norms into their theoretical edifice is a crucial break for rational choice theory for a more open recognition of the significance of norms and the role they play in establishing a collective choice. The other significant aspect of Elster is that he provides an interesting critique of the notion of preference formation (Elster 1983a; 1983b; 1989a). In Elster's view the notion of fixed and given preferences is a fragile foundation for social choice. In fact, preferences are not given in, the sense of being directly observable. If they are to serve as inputs in social choice process, they must somehow be expressed by the individuals. Moreover choices often conceal preferences rather than reveal preferences. Thus, in Elster's view, politics does not involve just the aggregation of preferences but also the transformation of preferences through dialogue. Not optimal compromise, but unanimous agreement is the goal of politics on this view.

Critical theorists need to show a greater deal of hermeneutic sensitivity in recognising the mulitiple theories present in rational choice theory. Rational choice is not a monolithic, homogeneous theoretical enterprise. And especially when rational choice theorists have themselves begun exploring the work of Habermas.

Jon Elster's work provides a good example of both giving internal critique and also recognising the strengths of Habermas' discourse ethics.

In Habermas' model of collective choice, the individual is a critical, reflective, moral one, unlike the pure instrumental maximiser in rational choice theories. So Habermas' discourse ethics provides a normative standpoint to assess the fairness of the agreements or consensus. In turn, critical theory gains from this conversation with rational choice paradigm since the latter provides substance for its currently rather abstract critique of politics. The strength of rational choice theories lies in its concern for institutional design. More importantly, although deliberation offers a better ideal it cannot escape the difficulties in its institutional realisation. Since fairness and justice are matters which cannot be evaded in any formulation of social and public policy it becomes imperative to emphasise the normative significance of these notions.

The contrast between rational choice and critical theory can be seen in their respective conceptualisation of the 'political'. Rational choice theorists share the conception of the political process as constituted by atomised individuals engaging in secret voting, whereas critical theorists defend deliberative democracy. One way the public sphere is discursively institutionalised is in its differentiation of value-spheres. Habermas sees the public sphere as being autonomous from both market and state. It is an arena in which the civil liberties and democratic rights of individuals are safeguarded. Of importance here is the normative concept of the public sphere. Thus for Habermas the political point was to secure 'criticism' as a moral imperative extending beyond mere self-interest (Bronner 1994: 287). But as he argued in his thesis of the colonisation of the life-world, the public sphere, rather being constituted by the discursive engagement of participants, is becoming increasingly defined by the same forms of instrumental reason as exhibited in the state and the market.

From the standpoint of critical theorists, rational choice theory is deeply flawed in that it takes instrumental reason as the only form of reason realised in modernity. By contrast, critical theorists conceptualise politics as a moral-practical domain wherein rules and norms governing the decisions are subjected to criticism and revision. Consequently the domain of politics is kept separated from market. Rational choice theories, with their assumption of *homo economicus*, end up subjecting political rules to the rules of market. Treating the polity as no more than a set of isolated individuals interacting to maximise gains, the model is devoid of any moral prescriptions.

To conlcude, theoretical encounters at the level of paradigms needs a careful treatment of issues under investigation. Viewing social science research as constituting paradigms is advantageous since paradigms are not closed, static scientific communities but typically sustain multiple theories working within the paradigm. Theories often coexist with empirical anomalies for long periods of time. Moreover, they are never decisively falsified by the facts. They are rejected only when an alternative and more plausible theory is proposed. Green and Shapiro (1994) in their book contend that little of what has been claimed for rational choice theory is backed up by empirical results. The same applies to critical theory. It is also esoteric and far removed from the empirical problems. The paradox is that the need for empirical inquiry of social reality is shared at a meta-theoretical level without actually confronting it. In this sense critical theory is 'a priori' theory. Often, critical theorists mistake empirical details for empiricism. Thus an opportunity for conversation across the two traditions arises inasmuch as critical theory also benefits from this encounter by gaining content for its currently rather abstract critique of politics and rationality.

Alternatives to rational choice theory range from the normative, to cultural, to institutional, to psychological theories. Given the complexity of social reality, it

is unreasonable to expect that only one general theory can be developed to explain disparate phenomena that rational choice theorists think of as political. This has nothing to do with the poverty of theory but rather with the recalcitrant complexity of the political world.

The interesting point is that critical theorists like Habermas concede the role of instrumental reason but question the universalism which is so deeply ingrained in the mind set of the theorists. In other words, rational choice theorists should recognise the failure of their theories to explain much of empirical reality and open up to an engagement in conversation with other theoretical paradigms for better explanations. On the other hand, unlike his predecessors Habermas does not dismiss instrumental rationality but confines its role to spheres like science and technology and market. Strategic rationality has a role but a limited one. By contrast, politics is a realm of collective choice and so cannot be modelled on the lines of market. Elster (1986b) is also sympathetic to this distinction between market and politics. Political process is not just seen as an arena for self serving interests but also a place to deal with questions such as justice, power and domination.

References

Abrams, Robert. 1980. *Foundations of Political Analysis: An Introduction to the Theory of Collective Choice*. New York: Columbia University Press.

Ackerman, Bruce. 1989. 'Why Dialogue'. In *The Journal of Philosophy*, Vol. 86, No. 1, pp. 5-22.

Alt, James and Kenneth Shepsle. 1992. 'Editor's Introduction'. In James E. Alt and Kenneth A. Shepsle. (ed.), *Perspectives on Positive Political Economy*. pp. 1-5, Cambridge University Press.

Aumann, Robert and Sergiu Hart (eds.) 1992. *Handbook of Game Theory*. Amsterdam: North-Holland.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Barry, Brian. 1978. *Sociologists Economists and Democracy*. Chicago: University of Chicago Press.

Barry, Brian and Russell Hardin. 1982. 'Preface and Epilogue'. In Brian Barry and Russell Hardin (ed.) *Rational Man and Irrational Society*. pp. New Delhi: Sage Publishers.

Baynes, Kenneth. 1992. *The Normative Grounds of Social Criticism*. Albany: SUNY.

Benhabib, Seyla. 1986. *Critique, Norm and Utopia: A Study of the Foundations of Critical Theory*. New York: Columbia University Press.

Benhabib, Seyla. 1989. 'Liberal Dialogue Versus Critical Theory of Discursive Legitimation'. In Nancy L. Rosenblum. (ed.) *Liberalism and the Moral Life*. pp. 143-156, Cambridge, MA: Harvard University Press

Bernstein, Richard. 1983. *Beyond Objectivism and Relativism: Science Hermeneutics and Praxis*. Philadelphia: University of Pennsylvania Press.

Bernstein, Richard. (ed.) 1985. *Habermas and Modernity*. Cambridge, Massachussetts: MIT Press.

Bohman, James. 1991. *New Philosophy of Social Science*. Cambridge: Polity Press.

Bohman, James. 1989. 'Participating in Enlightenment: Habermas' Cognitivist Interpretation of Democracy'. In Marcelo Dascal and Ora Gruengard (ed.) *Knowledge and Politics*, pp. 264-289, Boulder, Cl.: Westview.

Brennan, Geoffrey and James Buchanan. 1985. *The Reason of Rules*. London: Cambridge University Press.

Brennan, Geoffrey. 1989. 'Politics with Romance: Towards a Theory of Democratic Socialism'. In Philip Pettit, (ed.) *The Good Polity: Normative Analysis of the State*, pp. 49-66, Oxford: Basil Blackwell.

Bronner, Stephen. 1994. *Of Critical Theory and Its Critics*. Oxford: Basil Blackwell.

Buchanan, James M. and Gordon Tullock. 1962. *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Buchanan, James M. 1975. *The Limits of Liberty*. Chicago: University of Chicago Press.

Buchanan, James M. 1977. *Freedom in Constitutional Contract*. College Station: Texas A and M Press.

Buchanan, James M. 1978. 'From Private Preferences to Public philosophy: The Development of Public Choice'. In James M. Buchanan (ed) *The Economics of Politics*, pp.1-20 London: Institute of Economic Affairs.

Buchanan, James M. 1986. *Liberty, Market and State*. Wheatsheaf: Harvester Press.

Buchanan, James M. and Victor Vanberg. 1989. 'Interests and Theories in Constitutional Choice'. In *Journal of Theoretical Politics* 1, pp. 49-62.

Buchanan, James M. 1991. 'Politics without Romance: A Sketch of Positive Public Choice and its Normative Implications'. In Alan Hamlin and Philip Pettit (eds) *Contemporary Political Theory*, pp. 216- 228, New York: MacMillan.

Calhoun, Craig. (ed.) 1992. *Habermas and the Public Sphere*. Cambridge, Massachussetts: M I T Press.

Dallmayr, Fred. 1986. 'Life-World and Communicative Action'. In Bikhu Parekh and Thomas Pantham. (ed.) *Political Discourse*, pp. 152-178, New Delhi: Sage Publishers.

Dearlove, John. 1992. 'Neoclassical Politics: Public Choice and Political Understanding'. In *Journal of Theoretical Politics*, pp. 208-237.

Dews, Peter. (ed.) 1986. *Habermas: Autonomy and Solidarity*. London: Verso.

Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper and Row.

Downs, Anthony. 1991. 'Social Values and Democracy'. In Kristen R. Monroe (ed.), *The Economic Approach To Politics,* pp. 143-170, New York: Harper Collins.

Dryzek, John. 1991. *Discursive Democracy.* Cambridge: Cambridge University Press.

Dryzek, John. 1992. 'How Far Is It From Virginia and Rochester to Frankfurt? Public Choice and Critical Theory'. In *British Journal of Political Science,* 22, pp. 397-417.

Elster, Jon. 1976: 'Some Conceptual Problems in Political Theory'. In Brian Barry (ed.) Power and Political Theory, pp 248-49, London: Wiley.

_____ 1983a. *Sour Grapes.* Cambridge: Cambridge University Press.

_____1983b. *Explaining Technical Change.* Cambridge: Cambridge University Press.

_____ 1985. 'Rationality, Morality and Collective Action'. In *Ethics* 96, pp. 136-155.

Elster, Jon. 1986a. 'Introduction'. In Jon Elster (ed.) *Rational Choice,* pp. 1-33, Oxford: Blackwell Publisher.

Elster, Jon. 1986b. 'The market and the forum: three varieties of political theory'. In Jon Elster and Aunand Hylland (ed.), *Foundations of Social Choice Theory,* pp. 103-129, Cambridge: Cambridge University Press.

Elster, Jon. 1988b. 'Arguments for constitutional choice: reflections on the transition to socialism'. In Jon Elster and Rune Slagstad (ed.) *Constitutionalism and Democracy,* pp. 303-326, Cambridge: Cambridge University Press.

_____ 1989a. *Solomonic Judgments.* Cambridge: Cambridge University Press.

Elster, Jon. 1989b. *Nuts and Bolts.* Cambridge: Cambridge University Press.

Elster, Jon. 1989c. *The Cement of Society: A Study of Social Order.* Cambridge: Cambridge University of Press.

Elster, Jon. 1989d. 'Self Realisation in work and politics: the Marxist conception of the good life'. In Jon Elster and K. O. Moene (ed.) *Alternatives to Capitalism,* pp. 127-158, New York: Cambridge University Press.

Elster, Jon. 1991. 'The Possibility of Rational Politics'. In David Held. *Political Theory*, pp. 115-142, Polity Press.

Ferejohn, John and Debra Satz. 1995. 'Unification, Universalism, and Rational Choice Theory'. In *Critical Review*, Vol. 9, Nos. 1-2, Winter-Spring, pp. 71-84.

Fiorina, Morris P. 1995. 'Rational Choice, Empirical Contributions, and the Scientific Enterprise'. In *Critical Review,* Vol. 9, Nos. 1-2, Winter-Spring, pp. 85-94.

Friedman, James. 1995. 'Economic Approaches to Politics'. In *Critical Review,* Vol. 9, Nos. 1-2, Winter-Spring, pp. 1-24.

Green, Donald P. and Ian Shapiro. 1994. *Pathologies of Rational Choice Theory.* New Haven: Yale University Press.

Green, Donald P. and Ian Shapiro. 1995. 'Reply, Pathologies Revisited: Reflections on Our Critics'. In *Critical Review,* Vol. 9, Nos. 1-2, Winter-Spring, pp. 235-279.

Habermas, Jurgen. 1970a. *Towards a Rational Soicety.* trans by. J. Shapiro. Boston: Beacon Press.

_____ 1970b. *Knowledge and Human Interests.* trans by, J. Shapiro. Boston: Beacon Press.

_____ 1973. *Theory and Practice.* Boston: Beacon Press.

_____ 1975a. *Legitimation Crisis.* trans by, Thomas McCarthy. Boston: Beacon Press.

_____ 1975b. A Postscript to Knowledge and Human Interests. *Philosophy of the Social Sciences* 3: 157-189.

_____ 1979. *Communication and the Evolution of Society.* trans by, Thomas McCarthy. London: Heinemann.

_____ 1982. 'A Reply to My Critics'. In John B. Thompson and David Held (ed.) *Habermas: Critical Debates.* Cambridge: MIT Press.

_____ 1984. *The Theory of Communicative Action* Vol. I. trans by, Thomas McCarthy. London: Heinemann Books.

_____ 1987. *The Theory of Communicative Action* Vol. II. trans by, Thomas McCarthy. London: Heinemann Books.

_____ 1991. *The Structural Transformation of Public Sphere.* trans by, Thomas Burger. Great Britain: Polity Press.

_____ 1990. *Moral Consciousness and Communicative Action.* trans by, Christian Lenhardt and Shierry Weber Nicholsen. Cambridge: MIT Press.

_____ 1995. *The Past as Future.* Interviewed by Michael Haller. Great Britain: Polity Press.

Hardin, Garrett. 1968. 'The Tragedy of the Commons'. In *Science* Vol. 162, No. 3859, pp. 1243-1248.

Hardin, Russell. 1993: 'Public Choice Versus Democracy'. In David Copp., Jean Hampton., and John E. Roemer (ed.) *The Idea of Democracy,* pp. 157-172, Cambridge: Cambridge University Press.

Held, David 1980. *Introduction to Critical Theory: Horkheimer to Habermas.* London: Hutchinson.

Heap. Shaun H. 1989. *Rationality in Economics.* Oxford: Oxford University Press.

Heap, Shaun H. et al 1992. *The Theory of Choice: A Critical Guide.* Oxford: Blackwell.

Heap, Shaun H. and Yanis Varoufakis. (ed.) 1995. *Game Theory: A Critical Introduction.* London: Routledge.

Holmes, Stephen. 1988. 'Gag rules or the politics of Omission'. In Jon Elster and Rune Slagstad (ed.) *Constitutionalism and Democracy,* pp. 19-58, Cambridge: Cambridge University Press.

Horkheimer, Max and Theodore Adorno. 1972. *Dialectic of Enlightenment.* New York: Herder and Herder.

Hubin, Donald C. 1986. 'Of Bindings and By-products: Elster on Rationality'. In *Philosophy and Public Affairs,* Vol. 15, No. 1 Winter.

Ingram, David. 1987. *Habermas and The Dialectic of Reason.* New Haven: Yale University Press.

Johnson, James D. 1991a. 'Habermas on Strategic and Communicative Action'. In *Political Theory,* Vol. 19, No. 2, May, pp. 181-201.

Johnson, James D. 1991b. 'Rational Choice as a Reconstructive Theory'. In Kristen R. Monroe (ed.), *The Economic Approach To Politics,* pp. 113-142, New York: Harper Collins.

Johnson, James D. 1993. 'Is Talk Really Cheap? Prompting Conversation Between Critical Theory and Rational Choice'. In *APSR*, Vol. 87, No. 1, March, pp. 74-86.

Knight, Jack and James Johnson. 1994. 'Aggregation and Deliberation: On The Possibility of Democratic Legitimacy'. In *Political Theory*, Vol. 22 No. 2, May, pp 277-296.

Kymlicka, Will. 1991. *Contemporary Political Philosophy*. Oxford: OUP.

Lessnoff, Michael. 1986. *Social Contract*. London: Macmillan Publishers.

Marcuse, Herbert. 1964. *One Dimensional Man*. Boston, Mass.: Beacon Press.

McCarthy, Thomas A. 1978. *The Critical Theory of Jurgen Habermas*. Cambridge: Polity Press.

McCarthy, Thomas A. 1979. 'Translator's Introduction'. In Jurgen Habermas. *Communication and the Evolution of Society*, pp. vii-xxiv, London: Heinemann.

McCarthy, Thomas A. 1984. 'Translator's Introduction'. In Jurgen Habermas. *The Theory of Communicative Action Vol I: Reason and Rationalisation of Society*, pp v-xxxvii, Boston: Beacon Press.

McLean, Iain. 1987. *Public Choice: An Introduction*. Oxford: Blackwell.

Mitchell, William C. 1988. 'Virginia, Rochester and Bloomington: Twenty-Five Years of Public Choice and Political Science'. In *Public Choice*, 56: pp. 101-119.

Moon, Donald J. 1991. 'Constrained Discourse and Public Life'. In *Political Theory*, Vol. 19, No. 2, May, pp. 202-229.

Monroe, Kristen R. 1991. 'Introduction: The Theory of Ratonal Action'. In Kristen R. Monroe (ed.), *The Economic Approach To Politics*, pp. 1-52, New York: Harper Collins.

Mueller, Dennis C. 1989. *Public Choice II*. Cambridge: Cambridge University Press.

Olson, Mancur. 1965. *The Logic of Collective Action*. Cambridge, MA: Harvard University Press.

Olson, Mancur. 1993. 'Dictatorship, Democracy and Development'. In *APSR*, Vol. 87, No. 3 September, pp. 567-576.

Olson, Mancur. 1992. 'Toward a unified view of economics and the other social sciences'. In James E. Alt and Kenneth A. Shepsle (ed.), *Perspectives on Positive Political Economy*, pp. 212-231, Cambridge University Press.

O' Leary, Brian. 1987. 'The Odyssey of Jon Elster'. In *Government and Opposition*, Vol. 22 No. 4, pp. 480-498.

Ordeshook, Peter C. 1992. 'The emerging discipline of political economy'. In James E. Alt and Kenneth A. Shepsle (ed.), *Perspectives on Positive Political Economy*, pp. 9-30, Cambridge University Press.

Outhwaite, William. 1994. *Habermas: A Critical Introduction.* Cambridge: Polity Press.

Petracca, Mark P. 1991. 'The Rational Actor Approach to Politics'. In Kristen R. Monroe (ed.), *The Economic Approach To Politics*, pp. 171-203, New York: Harper Collins.

Przeworski, Adam. 1985. 'Marxism and Rational Choice'. In *Politics and Society* Vol. 14 Num 4, pp. 379-410.

Pusey, Michael. 1987. *Jurgen Habermas.* London: Tavistock Publications.

Rasmussen, David M. 1990. *Reading Habermas.* Oxford: Basil Blackwell.

Rawls, John. 1971. *A Theory of Justice.* Cambrdige, Mass.: Harvard University Press.

Rawls, John. 1985. 'Justice as Fairness: Political and Not Metaphysical'. In *Philosophy and Public Affairs* 14, pp. 223-251.

Riker, William H. 1982. *Liberalism Against Populism.* San Francisco: Freeman.

Riker, William H.1992. 'Political Science and Rational Choice'. In James E. Alt and Kenneth A. Shepsle (ed.) *Perspectives on Positive Political Economy*, pp. 163-181, Cambridge University Press.

Rousseau, Jean Jaques. 1762. *The Social Contract.* edited together with the *Discourses* by G.D.H. Cole. London: Dent, 1973.

Sen, Amartya. 1986. 'Foundations of social choice theory: an epilogue'. In Jon Elster and Aunand Hylland (ed.) *Foundations of Social Choice Theory*, pp. 213-248, Cambridge: Cambridge University Press.

Sunstein Cass R. 1988. 'Constitutions and democracies: an epilogue'. In Jon Elster and Rune Slagstad (ed.) *Constitutionalism and Democracy*, pp. 327-353, Cambridge: Cambridge University Press.

Taylor, Michael. 1987. *The Possibility of Cooperation*. Cambridge: Cambridge University Press.

Weale, Albert. 1984. 'Review Article: Social Choice Versus Populism? An Interpretation of Riker's Political Theory'. In *BJPS*, Vol. 14, Part-3, pp. 369-384.

Weale, Albert. 1992. 'Social Choice'. In Shaun Hargreaves Heap et al. *The Theory of Choice: A Critical Guide*, pp. 199-216, Oxford: Blackwell

White, Stephen K. 1988. *The Recent work of Jurgen Habermas*. Cambridge: Cambridge University Press.

White, Stephen K. 1987. 'Toward a Critical Political Science'. In Terence Ball. (ed.) *Idioms of Inquiry*, Albany: SUNY.