

DECLARATION

This thesis is being submitted to Cardiff University in partial fulfillment of the requirements for the degree of PhD.

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Date: 19.10.10



(Sotos Shiakides)

CANDIDATE

To my father
and the memory of my mother

S.S.

ACKNOWLEDGEMENT

I would like to express my deep appreciation and gratitude to my supervisor Professor Christopher Norris for his generous and wise advice as well as for his unfailing encouragement and support while I was working on this dissertation.

**PRACTICAL RATIONALITY
WITH A PRACTICAL INTENT:
A REVISED HABERMASIAN CONCEPTION
CENTRED ON COMMUNICATIVE NEGOTIATION**

A Cardiff University Doctoral Dissertation

by

Sotos Shiakides

MARCH 2010

UMI Number: U516663

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U516663

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

SUMMARY

This dissertation explores how and to what extent rationality can provide guidance for human action – individual and collective, private and public – particularly under the conditions of modernity. The theme is approached through an examination of Habermas's various pertinent contributions.

In both the pre-communicative and the communicative phases of his work, Habermas has been concerned to show the one-sided and distorted nature of modern, instrumental rationalization, and to invigorate the “unfinished project of enlightenment”. Thus he has endeavoured to develop and philosophically justify a complementary, communicative practical rationality. Communicative rationality would be employed for rationalizing social interaction, whereas instrumental rationality would be restricted to human beings' interchange with nature, concerned with securing their means of material reproduction.

One exception to this supposed exhaustive dichotomy is the case of human interaction that is entwined so closely with the means of material reproduction that if it were conducted communicatively, rather than instrumentally, it would be so inefficient as to threaten material reproduction itself. Fields of social interaction that, for the sake of material reproductive efficiency, need to be coordinated instrumentally – as “subsystems of strategic action” – are the economy and state administration.

This study identifies and mainly concentrates on a lacuna common to Habermas's pre-communicative and communicative conceptual frameworks, one that results in an effective veto on the possibility that human agents can communicatively pursue their own ends by means of other agents. In this way, the vast range of social interactions in which the ends of human beings must inevitably be satisfied through the active involvement or consent of other human beings – not only in the economic and state-administrative areas but also in “lifeworld” domains such as family life, education and politics – are *a priori* ceded to strategic action, contrary to Habermas's own intentions.

To remedy this problem, my study sets out to develop and provide philosophical justification for a revised conception of practical rationality centred on the notion of communicative negotiation. This conception, which remains broadly within the framework of Habermas's communicative paradigm, makes the communicative rationalization of all areas of social interaction *in principle* possible.

Finally, the study tentatively explores ways of extending the *practical* applicability of the revised conception to all fields of modern social activity, including those closely entwined with the modes of instrumental action concerned with society's material reproduction.

TABLE OF CONTENTS

| | Page |
|---|------|
| GENERAL INTRODUCTION | 7 |
| 1. PART ONE: HABERMAS'S EARLY WORK ON ACTION AND PRACTICAL RATIONALITY | 25 |
| 1.1 INTRODUCTION..... | 25 |
| 1.2. AN EXPOSITION OF HABERMAS'S EARLY VIEWS ON ACTION AND PRACTICAL RATIONALITY..... | 28 |
| 1.2.1 <i>Techne vs Praxis</i> | 28 |
| 1.2.2 The modern triumph of <i>techne</i> and demise of <i>praxis</i> | 39 |
| 1.2.3 Habermas's early conception of praxial rationality | 58 |
| 1.3 CRITIQUE OF HABERMAS'S EARLY VIEWS ON ACTION AND PRACTICAL RATIONALITY..... | 66 |
| 1.3.1 Critical perspectives concerning the dichotomy between <i>techne</i> and <i>praxis</i> | 69 |
| 1.3.1.1 McCarthy's critique..... | 69 |
| 1.3.1.2 Honneth's critique..... | 79 |
| 1.3.1.3 Giddens's critique..... | 83 |
| 1.3.2 Critical perspectives on Habermas's ideas on the rationality of <i>techne</i> and <i>praxis</i> | 86 |
| 1.3.2.1 Critical perspectives on technical rationality..... | 87 |
| 1.3.2.2 Critical perspectives on praxial rationality..... | 93 |

| | |
|---|------------|
| 2. PART TWO: PRACTICAL RATIONALITY IN HABERMAS'S COMMUNICATIVE PARADIGM..... | 122 |
| 2.1 INTRODUCTION..... | 122 |
| 2.1.1. Taking Stock..... | 122 |
| 2.1.2 New Departure | 124 |
| 2.2 FUNDAMENTALS OF HABERMAS'S UNIVERSAL PRAGMATICS OF LINGUISTIC COMMUNICATION | 128 |
| 2.2.1 A synoptic presentation | 128 |
| 2.2.2 Some objections and clarifications | 134 |
| 2.3 TYPOLOGY OF ACTION IN THE NEW PARADIGM | 146 |
| 2.3.1 Presentation of main ideas..... | 146 |
| 2.3.2 Some objections and clarifications | 151 |
| 2.4 HABERMAS'S MATURE MODEL OF PRACTICAL RATIONALITY | 154 |
| 2.4.1 The Pragmatic Employment of Practical Rationality..... | 155 |
| 2.4.2 The Ethical Employment of Practical Rationality..... | 158 |
| 2.4.3 The Moral Employment of Practical Rationality..... | 169 |
| 2.4.4 A conceptual lacuna | 180 |
| 2.5 HABERMAS'S SUBSTANTIVE THEORY OF SOCIETY : LIFEWORLD AND SYSTEM | 183 |
| 2.5.1 Main Theses..... | 183 |
| 2.5.2. Some Important Critical Perspectives | 197 |

| | |
|--|------------|
| 3. PART THREE: TOWARDS A THOROUGHLY COMMUNICATIVE CONCEPTION OF SOCIAL ACTION AND ITS RATIONALITY..... | 210 |
| 3.1 A CONCEPTUAL GAP IN HABERMAS: BRINGING TOGETHER CRITICAL THEMES DEVELOPED IN PARTS ONE AND TWO..... | 210 |
| 3.2 A MORE PRECISE SPECIFICATION OF THE CONCEPTUAL GAP | 214 |
| 3.3 FAIR BARGAINING AND WEAK COMMUNICATIVE RATIONALITY: SOME RECENT THINKING BY HABERMAS CONCERNING ACTION AND ITS RATIONALITY..... | 224 |
| 3.3.1 Fair Bargaining | 224 |
| 3.3.2 “Weak” communicative action and rationality | 229 |
| 3.3.2.1 Declarations of intention..... | 231 |
| 3.3.2.2 Naked imperatives | 235 |
| 3.4 COMMUNICATIVE NEGOTIATION: FULLY OR RESTRICTEDLY SUCCESS-ORIENTED COMMUNICATIVE INTERACTION | 239 |
| 3.5 A RECONSIDERATION, IN THE LIGHT OF THE CONCEPT OF COMMUNICATIVE NEGOTIATION, OF MORAL AND NON-MORAL NORMS AND RIGHTNESS AND THEIR UNIVERSAL-PRAGMATIC GROUNDING | 248 |
| 3.5.1 Applications of the universalization principle as communicative negotiations | 248 |
| 3.5.2 The contingency of norm-based morality | 250 |
| 3.5.3 The social functionality of norm-based morality | 253 |
| 3.5.4 Moral and non-moral normative frameworks | 256 |
| 3.5.5 Inadequacy of the instrumental justification of moral and non-moral norms | 259 |
| 3.5.6 From the validity claim to normative rightness to the validity claim to intersubjective acceptability..... | 260 |
| 3.6 TAKING STOCK: THE HABERMASIAN AND THE REVISED CONCEPTUAL SCHEMES REGARDING PRACTICAL RATIONALITY | 265 |

| | | |
|---------|---|-----|
| 3.7 | FURTHER ELABORATION OF THE REVISED MODEL OF PRACTICAL RATIONALITY: RECIPROCAL DEPENDENCES BETWEEN ETHICAL, PURPOSIVE AND NORMATIVE DISCOURSES, AND THE CONCEPT OF COMMUNICATIVE NEGOTIATION | 271 |
| 3.7.1 | Ethical, Purposive and <i>prima facie</i> Normative Discourses and relations of interdependence between them..... | 271 |
| 3.7.1.1 | Ethical Discourses..... | 271 |
| 3.7.1.2 | Purposive and Ethical Discourses..... | 275 |
| 3.7.1.3 | Ethical, Purposive and <i>prima facie</i> Normative Discourses | 284 |
| 3.7.2 | Communicative Negotiation..... | 285 |
| 3.7.2.1 | Introductory ideas..... | 285 |
| 3.7.2.2 | Elements of Rational Choice Negotiation Analysis..... | 288 |
| 3.7.2.3 | From Rational Choice Negotiation Analysis to Communicative Negotiation .. | 301 |
| 3.8 | TOWARDS A PRACTICABLE IDEA OF A COMMUNICATIVELY RATIONAL SOCIETY: PRELIMINARY CONSIDERATIONS ON THE SOCIAL APPLICATION OF THE REVISED CONCEPTION OF PRACTICAL RATIONALITY | 310 |
| 3.8.1 | The problem of the revised concept's practical application in modern society | 310 |
| 3.8.2 | Towards overcoming obstacles to the revised concept's practical application..... | 316 |
| 3.8.2.1 | Social Structuring Devices | 317 |
| 3.8.2.2 | Enabling Conditions | 323 |
| 3.8.2.3 | Lines of Further Research..... | 328 |
| | CONCLUSIONS | 331 |
| | APPENDIX I: HUME'S SUBJECTIVE-INSTRUMENTAL CONCEPTION OF PRACTICAL RATIONALITY..... | 344 |
| | APPENDIX II: KANT'S TRANSCENDENTAL CONCEPTION OF PRACTICAL RATIONALITY | 350 |
| | APPENDIX III: MAX WEBER AND THE DISCONTENTS OF MODERN RATIONALIZATION | 362 |
| | NOTES | 392 |
| | BIBLIOGRAPHY | 400 |

GENERAL INTRODUCTION

1. THE THEME AND ITS RATIONALE: PRACTICAL RATIONALITY AND HABERMAS

The present study is intended as a small contribution to the theoretically and practically important, as well as long-standing, debate on the capabilities and limitations of practical rationality, that is, of rationality in its capacity as a guide to action. The study's overall theme can be broadly defined in terms of the question "To what extent and in what ways can human beings today justifiably rely on the exercise of their universal rational faculties for deriving guidance for their actions, individual and collective, private and public?". This general theme is being specifically approached through the critical exploration of Habermas's uniquely insightful, rich, practically pertinent and influential lifelong contributions on the subject. The many strengths as well as the shortcomings of Habermas's work on practical rationality are considered, with a view to offering constructive proposals for remedying some of the latter.

The post-medieval western world, and through its global influence also large sections of the rest of humanity, have increasingly come to rely on rationality both for acquiring theoretical knowledge about reality and for answering questions of the form "What should I/we do?". Time-honoured alternative sources of knowledge and action-guidance such as sacred texts, divine revelation and tradition have gradually been losing their unquestioned authority and social dominance. In an ongoing historical process that culminated in the 18th century "Age of Enlightenment", rationality has become the core of modernity and the foundation of modernity's promise of humanity's continuous progress towards freedom, justice and happiness. This process had its roots in ancient Greece, and in some

respects can even be traced in the midst of the theocratic Middle Ages. Thus, in Greek philosophy we have perhaps the first systematic attempts to establish rational foundations for knowledge and action. Plato, notably, considered the forms of Good and Justice as parts of the rationally knowable objective reality, on a par with the forms of mathematical and natural objects; while Aristotle grounds his social-contextualist virtue ethics on a teleological conception of human beings as rational and social animals. The European Middle Ages, on the other hand, predominantly considered God's Wisdom and Will, as mediated by the Church, to be the unquestioned authority on matters of truth and falsehood as well as on issues of right and wrong. Human reason was dismissed not just as an inferior way of approaching those questions but as totally inappropriate for them. Nonetheless, as early as the 13th century, an (Aristotelian) Christian theologian such as Thomas Aquinas felt the need to show that human reason can in fact arrive at conclusions (in the form of Natural Law) which are identical to the divinely revealed precepts.

With the advent of the Renaissance there was a considerable shift in attitudes towards rationality, evidenced, for example, in Machiavelli's political writings and Francis Bacon's advocacy of an empiricist-inductivist method for science and his simultaneous understanding of scientific knowledge as a form of power. That shift brought to an end the long centuries of religion's unchallenged supremacy over rationality, or of the harmonious coexistence between the two. With rationality's subsequent assertiveness – above all in the form of the 17th century Scientific Revolution and radical philosophical developments such as those represented by the work of Descartes and Hobbes – modernity had dawned. Now reason could challenge the authority of religion, though the balance of social power was still not in its favour, as cases such as those of Galileo and Giordano Bruno amply

demonstrate. In fact, it was only with the Enlightenment that rationality would take root and become a social force to be reckoned with.

However, in the two centuries since the Enlightenment, rationality and modernity have been severely challenged both on persisting pre-modern grounds and (beginning in the very Age of Enlightenment and increasingly since) on the basis of standards and ideals grown on the soil of modernity itself. Thus, there is considerable disappointment with rationality's performance so far, and growing skepticism concerning its ability to fulfill its promises. Extreme poverty, exploitation and inequality are still widespread on the planet, while international, inter-cultural and other conflict is still largely dealt with through the use of force. On the other hand, the impressive advancement of science and technology does not seem to be delivering what had been expected.

In addition, rationality is held responsible for new, specifically modern problems. These include the serious and perhaps irreversible damage to the natural environment, both globally and locally, new sophisticated forms of domination and exploitation, the alienating conditions of modern city life and meaninglessness.

At the same time, just as rationality's action-guiding abilities are being challenged in these ways, our immensely complex, pluralistic and fast changing world demands more action-guiding capacity than ever. It can therefore be argued that contemporary humanity is suffering from an action-steering deficit. It is of course also obvious that such pre-modern or other alternative action-guiding principles as religion, unreflective traditionalism, authoritarian political rule or the use of naked power cannot meet today's challenges.

Thus, we cannot but turn even more decisively to humanity's capacity for rationality. We must make all the necessary effort to gain a better understanding of the nature of rationality, bring to light its hitherto unrealized potential and devise

means of realizing that potential in full; and for joining this effort one can hardly think of a better way than through critically exploring Habermas's relevant contributions.

Habermas himself sets his work in the context of the Enlightenment/counter-enlightenment debates, starting off by settling accounts with the "Dialectic of Enlightenment" thesis of his Frankfurt School mentors. This is precisely the context in the light of which his work can best be seen.

2. THE ENLIGHTENMENT / COUNTER-ENLIGHTENMENT CONTEXT OF THE DEBATE ON PRACTICAL RATIONALITY

2.1 Reason and the Enlightenment

That important part of modern history known as the Age of Enlightenment is geographically located in Western Europe – above all France, England, Scotland and Germany – and North America. Temporally, it is normally demarcated by the dates 1688–90 and 1776–90. Its beginning is marked by the English Revolution and Bill of Rights, and by the publication of Newton's *Principia* and Locke's *Essay Concerning Human Understanding* and *Two Treatises of Government*; its end, by the American and French Revolutions (with the accompanying Declaration of Independence and Declaration of the Rights of Man, respectively) and the publication of Kant's Critiques.

At the core of the Enlightenment – in the sense of the intellectual-cultural movement that prevailed in the "Age of Enlightenment" – is a thoroughly positive and optimistic view of human reason, both theoretical and practical. Accompanying it is a belief in the essential goodness of human nature and a this-worldly concern for human well being. According to this view, reason has the unique capacity to put humanity on a path to continuous improvement towards

greater freedom and happiness. When applied resolutely, systematically and publicly to natural as well as to cultural (social, historical, political, legal, moral) phenomena, reason will furnish humanity with true, reliable and useful knowledge concerning these phenomena: it will thus emancipate us from prejudices and superstitions about them and enable us to attain mastery over them. Human beings will stop being at the mercy of forces they do not adequately understand and become autonomous agents in control of their own destiny.

The Enlightenment refuses to accept anything uncritically – “on faith”. It considers reason to be the final judge of all beliefs and practices, including beliefs about reason's own nature, powers and limits, and practices involving its own employment. The Enlightenment is thus a fierce and uncompromising opponent of any other authority, such as tradition, divine revelation, sacred or other revered texts, or despotic political power.

This reason-centered conception of the Enlightenment is famously captured by Kant in his influential essay “An answer to the question: ‘What is Enlightenment?’” published in 1784, near the close of “The Age of Enlightenment”. Approaching the question from the point of view of the subject and emphasizing the subjective conditions for enlightenment, Kant maintains:

Enlightenment is man's emergence from his self-imposed immaturity. Immaturity is the inability to use one's own understanding without guidance from another. This immaturity is *self-imposed* when its cause lies not in lack of understanding, but in lack of resolve and courage to use it without guidance from another. *Sapere Aude!* "Have courage to use your own understanding" – that is the motto of enlightenment (Kant, 1983: 41).

Thus, for Kant the essence of enlightenment lies in human beings mastering the necessary courage to think for themselves; to use their understanding, their reason, without guidance and follow wherever it leads (Kant in this essay uses the terms "understanding" and "reason" interchangeably).

Kant does not think that enlightenment comes easily. On the contrary, he is convinced that, whereas "it is easy to be immature", "it is difficult for any individual man to work himself out of the immaturity that has all but become his nature" (*ibid.*: 41). Individuals who have for long been relying on others for guidance do not have either the necessary mental skills or the will to pull themselves, by their own efforts, out of that state and start thinking for themselves.

This, however, does not drive him to despair. For what is almost impossible for an individual is not so for a "public": "That the public should enlighten itself is more likely; indeed, if it is only allowed freedom, enlightenment is almost inevitable" (*ibid.*: 41–42). The freedom that Kant considers both a necessary and a sufficient condition for a public to achieve enlightenment – that is, the capacity for unguided thinking – is nothing more nor less than "the freedom to use reason *publicly* in all matters" (*ibid.*: 42). The free public use of reason will spread enlightenment from the – perhaps few – mature individuals, who are likely to exist in every public, to greater and greater numbers, releasing in each the (hitherto suppressed) potential for using reason, which Kant firmly believes that every human being is endowed with. Progress towards enlightenment is for Kant humanity's "essential destiny". As a way of achieving such progress, the free public use of reason in all matters is – Kant admits – a slow, painstaking process, not a revolutionary one, for a revolution "can never truly reform a manner of thinking" (*ibid.*: 42). But it is a reliable process; and the only one available.¹

Kant's immense confidence in reason was equally exhibited by such quintessential Enlightenment figures as Diderot, Voltaire, Montesquieu and generally all those collectively known as the French *philosophes*, including the more ambivalent Rousseau. But it can legitimately be attributed to David Hume too, despite his radical philosophical skepticism with regard to reason. For Hume's following celebrated admonition is as expressive of the spirit of the Enlightenment as any:

If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, Does it contain any abstract reasoning concerning quantity or number? No. Does it contain any experimental reasoning concerning matter of fact and existence? No. Commit it then to the flames: for it can contain nothing but sophistry and illusion (1975: 165)

However, the Enlightenment as a movement should not be thought of as identical with "The Age of Enlightenment", and this for two reasons. On the one hand, "The Age of Enlightenment", as Kant knew, was not an entirely "enlightened age" (Kant, 1983: 44); not only had the considerable efforts of the Enlightenment not succeeded during that period in fully defeating the heavy legacies of Europe's non-enlightened past, but also the Age had witnessed the birth of new forms of counter-enlightenment (Berlin, 1979). On the other hand, central aspects of the Enlightenment – such as confidence in reason – had a strong presence in other periods too, both before and after "The Age of the Enlightenment".

So far we have identified the core of the Enlightenment in minimalist and rather imprecise terms; we have done so mainly in terms of certain (positive and optimistic) views, attitudes and practices concerning reason, while leaving the notion of reason itself largely unclarified. This "thin" characterization – which is more or less all that can be done about defining the Enlightenment *in general* –

provides us with an initial, approximate understanding of the relevance of the Enlightenment to present-day discussions of practical rationality. However, for a fuller appreciation of the philosophical issues the Enlightenment raises for these discussions it is necessary to move beyond the Enlightenment in general to a consideration of *specific forms* of it.

Of course, we are not here concerned with forms of the Enlightenment in the sense of different sets of substantive doctrines or practical projects that result from the *application* of reason. Rather, we are interested in different sets of views concerning the nature and power of reason itself. In particular we are concerned with different Enlightenment approaches to the action-guiding aspects of reason, that is, with different Enlightenment conceptions of *practical rationality*.

2.2 Three Enlightenment conceptions of practical rationality

Enlightenment theories of practical rationality can be classified in terms of their position on two questions: (1) Can rationality provide *ultimate* guidance for human action and, if yes, on what grounds? (2) What form does rationality's guidance take? On this basis, we can identify three distinct conceptions:

1) The objective-substantive conception of practical rationality

This conception is typically embodied in Natural Law theories. According to it rationality does have the capacity to provide ultimate guidance for human action. It can do so by virtue of its ability to know *objective* reality, including human nature, which is the fundamental source of such guidance. Rationality's guidance comes in the form of *substantive* ends and values.

The derivation of substantive values from properties of objective reality, which has been severely criticized by David Hume and was later charged with committing the

“naturalistic fallacy”, was not of course an invention of the Enlightenment. It is a type of inference encountered already in important works such as Hobbes’s *Leviathan*, and employed long before by Thomas Aquinas who had himself derived it from Aristotle. Natural Law theories, however, with their reliance on such a form of argument, gained particular prominence in the Enlightenment. They did so mainly through the influence of John Locke, who in his *Second Treatise of Government* (in Locke, 1988) derives the natural rights of life, health, liberty and property – and from them a particular normative theory of government – on the basis of a view of the state of nature according to which all human beings are created by God equal, independent and rational.

2) The subjective-instrumental conception of practical rationality

This conception was famously articulated in the 18th century by David Hume, but has since had widespread appeal, having been espoused by positivism and presently underlying the Rational Choice paradigm of Decision and Game Theory. Its central tenet is that ultimate values and norms derive directly from the *subjective* realm of human affective states, without any mediation by the rational faculties. Rationality can know those states but does not have the power to judge or influence them; it is not, therefore, able to provide ultimate guidance for action. What rationality can do, with respect to action, is to show us how to attain the affectively determined ultimate values and norms, being thus an *instrument* of human affections. (Hume’s account of the *subjective-instrumental* view of practical rationality is briefly presented in Appendix I of this study).

3) The transcendental-procedural conception of practical rationality

This third major Enlightenment conception of practical rationality is put forward in Kant’s theory of morality and Practical Reason. In Kant’s view, ultimate guidance

for action derives neither from objective nor from subjective reality but from a *transcendental* realm accessible to our rational faculties. In this way rationality can guide action in an ultimate sense. Such guidance comes not in the form of substantive ends and values but primarily in the form of *procedural* principles on the basis of which substantive maxims of action can then be determined. (For a more detailed account of Kant's theory, see Appendix II). This Kantian conception of practical rationality has, in its broad structure, important contemporary adherents, including John Rawls and – in his later, communicative paradigm – Habermas himself.

2.3 Critical responses to Enlightenment views on practical rationality

Enlightenment ideas on practical rationality have been strongly challenged both with regard to reason's ability to establish ultimate values or norms, and with regard to its role as a neutral instrument in the service of non-rationally chosen human ends. Three main strands of criticism can be distinguished, two of them concerning the former issue, and a third focusing on the latter.

1) Contextualist critiques of rationality

Beginning with Hegel's critique of Kant, it has been argued that rational deliberation on values and norms is inconceivable independently of the evaluative/normative context in which it is carried out. No "Archimedian point" can be found from which reason can autonomously determine values or norms. Against Kant's alleged empty moral formalism – also in opposition to Natural Law objectivism, but very much in line with Aristotle's virtue ethics – Hegel puts forward the idea of *Sittlichkeit*, or "ethical life", which comprises all of society's customs, traditions, habits, values, laws, institutions, etc. The ethical life constitutes for Hegel the (insurmountable) context within and on the basis of which rational

human beings must *unavoidably* derive guidance for their decisions and actions. In a similar, contextualist vein are also several more recent lines of thought such as Gadamerian hermeneutics, Wittgensteinian language-game theory, and communitarian/neo-Aristotelian ethics.

2) Radical critiques of rationality

According to a line of argument originating in Nietzsche and in certain respects resembling that of the contextualists, what normally pass for purely rational processes are thoroughly permeated by non-rational, and often sinister, elements such as unconscious motives, interests and power. These elements – usually without being detected – leave their stamp on the outcomes of “rational” deliberations. This position must be clearly distinguished from that of Marx. For while Marx also considers the political and other values of bourgeois society as nothing but ideological constructs ultimately determined by the interests of the ruling, capitalist class, he looks forward to an emancipated, classless society in which evaluations and decisions would be made through rational deliberations free from ideological distortions. Nietzscheans, by contrast, view the influences of interests, power, etc. on deliberative processes not as distortions of a potentially “pure” rationality but as intrinsic aspects of rationality as such. This Nietzschean approach to rationality characterizes much of poststructuralist/postmodernist thought, most prominently that of Foucault and Lyotard.

3) The “Dialectic of Enlightenment” thesis

The third major line of critical response to the Enlightenment centres on the idea that the spread of rationality in modern society, far from bringing about human

emancipation and happiness as the Enlightenment had postulated is, on the contrary, leading to unfreedom and meaninglessness.

This idea appeared in its first important version in the work of Max Weber, and – after being construed in Marxist terms – formed the basis of Georg Lukács's theory of reification. Then, under the strong influence of both Weber and Lukács, the leading thinkers of the early Frankfurt School, above all Horkheimer, Adorno and Marcuse, developed their own highly original and radical formulations of the idea, which, after them, came to be known as the "Dialectic of Enlightenment" thesis.

Max Weber

For Weber the predominant feature of modernity is the spread of instrumental, or "purposive" rationality in all areas of social life, from the economy and administration to science and culture. Modern societies, surely, make so much use of instrumental rationality because they see in it – in Humean fashion – a most powerful tool in the service of their own purposes. But its generalized employment, according to Weber, is having seriously adverse effects on its own human users. The (instrumental) rationalization of society leads to the "disenchantment of the world" by bringing about the loss of the world's old religious-metaphysical meaning without replacing it with any new meaning. Moreover, it results in the most effective domination of nature and society and forces human beings into a dehumanizing, bureaucratic "iron cage".

This melancholy condition of modern humanity is for Weber virtually irreversible, precisely because instrumental rationality is, indeed, "technically superior" to any other way of pursuing the realization of human ends. Moreover, his pessimism regarding the action-guiding role of rationality is compounded by his – also

Humean – conviction that ultimate values or norms cannot be determined through rational deliberation (see Appendix III for a more extended account of Weber's views on modern rationalization).

Georg Lukács

In his seminal Hegelian-Marxist work *History and Class Consciousness* (1971), Lukács introduces the crucial concept of *reification*, which is based on Marx's analysis of commodity fetishism, in *Capital*. It refers to the transformation of human beings and their experiences and social relations into objectified, manipulable "things". Lukács uses the concept of reification in order to shed new light on Weber's analysis of rationalization. To begin with, he interprets what Weber identified as the adverse consequences of the spread of instrumental rationality as cases of reification. Unlike Weber, however, he does not see those reifying consequences of rationalization as inevitable effects of rationality as such. Rather, he attributes reification to the partial and distorted nature of capitalist rationalization; and for that he puts the blame squarely on the "commodity form" that lies at the heart of capitalism and permeates every aspect of capitalist society. This opens for Lukács the prospect of an overcoming of reification. For that purpose the commodity form must be abolished and the distorted rationalization of parts of society must be replaced with the rationalization of the social totality.

Lukács, just like Marx, reserves the task of such a transformation of society for the class of the proletariat. However, the proletariat will not play this historical role by historical necessity. Given the existing conditions of pervasive reification, it will develop the class consciousness necessary for doing so only by being appropriately enlightened by an *avant-garde* who can grasp in thought what it itself has to realize in history: the idea of a non-reified rational society.

The early Frankfurt School

The first generation of Frankfurt School Critical Theorists, especially in the post-1930's period, largely endorse and build on Weber's and Lukács's theses concerning the inextricable relationship between modern rationalization on the one hand and meaninglessness, domination and reification on the other. In works such as *Dialectic of Enlightenment*,² *Eclipse of Reason*,³ *One-Dimensional Man*⁴ and *Negative Dialectics*,⁵ Adorno, Horkheimer and Marcuse paint a picture of a totally administered, integrated, reified society; a society without any oppositional social forces or forms of thought; one in which social ideals – with the possible exception of those embodied in modern art – do not transcend social reality, and thus do not provide any potential or hope for emancipation. Blame for this bleak situation is clearly placed on instrumental reason, which is thought to dominate all thinking and practice.

Thus, the Enlightenment project is taken to have failed totally, turning into its exact opposite. In addition to rejecting instrumental rationalization, these theorists were also darkly pessimistic about the prospects of an alternative. Indeed, although they often alluded and at times explicitly referred to the idea of a non-instrumental form of rationality, they never specified its character in adequate conceptual terms, nor did they explain how it could take root in social reality. In fact, apart from not sharing Lukács's faith in the emergence of a revolutionary working class movement – or any other radical movement for that matter – which would transform the structure of capitalist society, their analyses seem to rule out the historical and even the anthropological possibility of a non-reifying reason that would characterize an alternative society. Furthermore, these theorists – Adorno in particular – often seem to rule out also the very conceptual possibility of a non-reifying, non-instrumental form of reason; for they tend to either charge

encountered instances of rationality that are claimed to be non-instrumental with openly having instrumental purposes, or suspect them for being covertly instrumental. To that extent, Frankfurt School Critical Theory comes very close to Nietzschean/poststructuralist critiques of reason.

3. HABERMAS AND THE DIALECTIC OF ENLIGHTENMENT

For Habermas, the Enlightenment, and modernity more generally, are not so much failed projects, as projects that are one-sidedly and distortedly implemented and thus unfinished. Although himself a major second generation Critical Theorist, he disagreed with the total-reification thesis of his Frankfurt mentors, while at the same time firmly denying that there are in the post-liberal phase of modern society, oppositional forces – working class or other – that guarantee the eventual victory of reason in history. His own view was that, despite the considerable degree of integration, there was still potential for emancipatory change; but whether and to what extent that potential would be realized was a historically contingent matter.

Habermas has always differed from the early Frankfurt philosophers also with regard to the central question concerning the nature of rationality itself. On the one hand he does not consider all the uses of instrumental reason as objectionable or indeed avoidable. In the context of our relationship to the natural world the employment of instrumental reason is for Habermas both necessary and acceptable. What is objectionable is its employment in human/social contexts, as is the case, for example, with the instrumentalization of politics or other areas of the “lifeworld”. And even here, the culprit, in the last analysis, is not instrumental reason as such but its application in the wrong fields, as a result of “technocratic

consciousness”, which typically mistakes evaluative and normative questions for technical ones.

Moreover, Habermas has been from the beginning confident that a form of rationality capable of determining values and norms – one that would complement an instrumental rationality kept to its proper field of application – is both conceptually possible and socially feasible. Such a form of rationality, however, would have to avoid the weaknesses of older versions of it, such as Natural Law theory and Kantian practical reason. In defending the possibility of an evaluative/normative rationality, Habermas has been at odds not only with the early Frankfurt School thinkers but also with Humean and Weberian “decisionism”, contextualist relativism and Nietzschean/poststructuralist radical critiques of reason.

Throughout his career, Habermas has aspired to contribute to the project of modernity mainly through renewing and invigorating Critical Theory, which in his view had been driven into a cul-de-sac by the older Frankfurt School generation. To that end, he has endeavoured to develop an adequate concept of value- and norm-oriented rationality, suitable for issues concerning intersubjective human interaction, or “*praxis*”. I shall frequently be referring to this type of rationality as “praxial”. It is a type of rationality that, although not entirely absent from modern social practice as the Frankfurt theorists argue, has nonetheless been seriously neglected by modernity, both in theory and in practice. For Habermas, praxial rationality would be the cornerstone of a substantive social theory that could once more engage not just in the description but also in the rational assessment and criticism of existing social reality. Complementary to praxial rationality, in a binary classification of rationality types, would be instrumental rationality, suitable for dealing with technical issues.

4. THE PRESENT STUDY

This study critically examines Habermas's views on practical rationality, praxial as well as instrumental, in both the pre-communicative and communicative phases of his work. In both phases Habermas constructs a sharp and exhaustive dichotomy between instrumental and praxial types of rationality. The former type is largely conceived along Humean lines. It concerns the pursuit of ends by means of causally manipulating objective or objectified processes. The basic structure of this form of rationality remains essentially unchanged in the two phases of Habermas's work. Praxial rationality, on the other hand, concerns the determination of values and norms through intersubjective interaction. Here, there is a significant conceptual shift between the two phases. In Habermas's pre-communicative theory, praxial rationality combines (conventional) hermeneutics with a depth-hermeneutical form of self-reflection based on a "therapeutic" relationship between enlightener and enlightened-to-be. In his communicative paradigm, however, to which Habermas turns mainly because of his dissatisfaction with precisely this form of self-reflection, praxial rationality is conceived as a combination of ethical and moral discourse. The former component is concerned with the dialogical or "communicative" clarification of values and identities, the latter with the determination of universal moral norms through the communicative application of a universalization principle. The difference between this concept of praxial rationality and the pre-communicative one lies mainly with the second component. For whereas ethical discourse can largely be construed as a form of hermeneutics, just as the first component of the early concept, with the latter component there is a drastic shift from depth-hermeneutical self-reflection to a communicatively rendered, Kantian-type normative morality.

Despite this difference, however, both the early and the later versions of the dichotomy between instrumental and praxial rationality seem to suffer from the same serious defect – one that is encountered also at the deeper level of the dichotomy’s philosophical underpinnings: It looks as if the dichotomy – by grouping together on one side means-ends relations with causal manipulation and on the other side communicative relations with an orientation to values and norms – makes it categorially impossible for human beings to pursue their goals by means of each other (i.e. to use the services of each other) in dialogical, communicative ways. They must – by conceptual necessity – do so through causal manipulation. The repercussions of such a conceptual arrangement on the prospect of a non-reified, rational society, are obviously momentous.

This study mainly focuses on this important problem. Its central axis constitutes an effort to confirm the existence of the problem, to clarify its nature and to offer solutions to it. The study is divided into three main Parts. Parts One and Two explore, respectively, Habermas’s pre-communicative and communicative views on practical rationality. In the course of these explorations the above-mentioned problem is identified and discussed in a preliminary manner. Part Three attempts a more in-depth analysis of the problem and proposes remedies for it in terms of a revised and suitably underpinned model of practical rationality, albeit one that broadly remains within Habermas’s later, communicative paradigm. The central component of this revised model is the concept of communicative negotiation. It is a concept that makes an entirely non-reified, rational society conceivable, by specifying how human beings can pursue their goals by means of each other in purely communicative ways. Finally, the study tentatively considers ways in which the proposed conception of practical rationality can be made *practically* applicable in society, under conditions of modernity.

PART ONE

HABERMAS'S EARLY WORK ON ACTION AND PRACTICAL RATIONALITY

1.1 INTRODUCTION

The Frankfurt tradition of Critical Theory to which Habermas belongs has always understood both itself and the type of Marxism from which it originates as being concerned, primarily, with the *critique of ideology*; that is, with the identification and critical exposure of the central misconceptions which hold sway over human thought and action in capitalist societies, concealing on the one hand the exploitative and oppressive aspects of those societies and on the other the potentiality for emancipation and happiness inherent in them. However, the ideology singled out for critique, within this tradition, varies from one generation of theorists to the next. More specifically, there are significant shifts regarding the target of critique, between Marx, the first generation Frankfurt School and the second generation, including Habermas.

Thus Marx is mainly concerned with a critique of the ideology of "equal exchange of commodities"; an ideology that conceals the unequal and unjust capitalist relations of production, as well as the socialist-revolutionary potential of the proletariat.

Horkheimer, Adorno and Marcuse, on the other hand, looking at modernity from the vantage point of post-liberal, organized capitalism, rather than the liberal capitalism of Marx's time, and having studied Weber's and Lukács's analyses of modern, capitalist rationalization, focus their critique on Instrumental Reason

instead. According to their diagnosis, the fundamental ideological misconception of modern society lies in the belief in Instrumental Reason's value-neutrality, or more precisely, in the concealment of the inherent link between Instrumental Reason (above all, in the form of science and technology) and domination over nature and human beings alike. This misconception, combined with the recognition of Instrumental Reason's immense power as a tool in the service of diverse human purposes, results in modern humanity's worship of Instrumental Reason as the method whose systematic employment guarantees unending historical progress. One of the important consequences of this ideology is that it conceals the necessity of any search for alternative – non-instrumental, non-dominatory – ways of thinking and acting.

With Habermas, now, we have yet another important shift in this tradition of ideology critique; a shift – *vis à vis* the early Frankfurt thinkers – which is not attributed by Habermas to the emergence, since their time, of any radically new phase of social development, comparable to that which separates them from Marx but rather to a different theoretical understanding of largely the same state of affairs as they had experienced: post-liberal, organized capitalism, or “late modernity”.

Habermas, particularly in the early phase of his work, contends that the dominant ideological misconception of the age is no longer that of equal exchange, as was the case in Marx's liberal-capitalist times. But at the same time he rejects the view that the ideology in question is the worship of Instrumental Reason as such; he does so, however, not because he denies that such a worship is strongly present in late modernity, but because he does not think that it is ideological in the sweeping, undifferentiated manner claimed by his Frankfurt progenitors. The central ideology of late modernity, according to Habermas, is *technocratic*

ideology, or, as he often refers to it, *technocratic consciousness*. This consciousness is said to consist in failing to distinguish adequately between two qualitatively different action-related questions, the “practical” and the “technical”, and treating all as if they were of the latter type. (For now suffice it to say that by “practical” Habermas here means, roughly, moral, political, and evaluative, referring to what he calls *praxis*,¹ whereas by “technical” he means whatever involves the devising and using of techniques for the realization of goals). Thus, instrumental or purposive rationality, which according to Habermas is proper to technical issues – and to them only – in late modernity tends to be wrongly applied to all issues in an undifferentiated way.

Ironically, the early Frankfurt thinkers in their wholesale rejection of Instrumental Reason are, for Habermas, as guilty of confusing the “technical” and the “practical” – a misconception that constitutes the core of the technocratic ideology – as those who give Instrumental Reason wholesale endorsement.

With his critique of technocratic ideology, Habermas wants to restore and clarify the distinction between the “technical” and the “practical” so as to be able, firstly, to define the proper space of instrumental rationality (that is, the space of “technical” issues), and secondly – and more importantly for him – to show that “practical” issues are also amenable to rational treatment, albeit by means of a rationality fundamentally different from the instrumental or purposive kind.

The two main sections of Part One – section 1.2 and 1.3 – provide, respectively, an exposition and a critical analysis of Habermas’s early views on action and practical rationality.

1.2 AN EXPOSITION OF HABERMAS'S EARLY VIEWS ON ACTION AND PRACTICAL RATIONALITY

1.2.1 *Techne vs Praxis*

In *Theory and Practice*, Habermas confesses that he became aware of “the fundamental significance of the Aristotelian distinction between *techne* and *praxis*” through studying Hannah Arendt’s *Vita Activa* and Hans-Georg Gadamer’s *Truth and Method* (Habermas, 1974: 286, n.4).

Aristotle, crystallizing notions already present in ancient Greek language and thought, sharply differentiates two types of human action: *poiesis* or making, on the one hand, and *praxis* or doing, on the other. The fundamental difference between the two is specified by Aristotle in the following way:

[D]oing and making are different in kind, since the maker of a thing has a different end in view than just making it, whereas in doing something the end can only be the doing of it well (Aristotle, 1955: 177).

Corresponding to these two types of action, are two very different types of ability or skill; two different “rational faculties”.² The faculty exercised in making is *techne*, practical art, covering also what we today classify as fine art. The faculty associated with *praxis* is *phronesis*, prudence, practical wisdom. Thus the domain of *poiesis* and *techne* is that of the arts and crafts. Here, the object, the end of *techne*, the result of “technical action”, is a particular product which is external to the action itself: a statue, a table, a tool. Moreover, not only is it the purpose of *techne* to produce something which is external to the productive activity, but it is essential to the nature of *techne* to bring about its products *causally*, in the sense of Aristotle’s efficient causality:

The business of every art is to bring something into existence, and the practice of an art involves the study of how to bring into existence something which is capable of having such an existence and has its efficient cause in the maker and not in itself (*ibid.*: 175).

By his last remark, above (“something which.... has its efficient cause... not *in itself* [my emphasis]), Aristotle refers to the distinction between *things produced* through *human action* and *things in nature* that come about through an inner causal necessity, thereby stressing the specifically human character of *techné*.

In contrast to productive action, the domain of *praxis* is that of ethics, family life and above all, politics. And *phronesis*, prudence, or practical wisdom, is “a rational faculty exercised for the attainment of truth in things that are humanly good or bad” (*ibid.*: 177). Good *praxis* is conduct which is conducive to a good human life. But since the end of *praxis* is not a product external to action but an end that lies in the action itself, good *praxis* is conduct that is constitutive of the good life. Accordingly, practical wisdom is the ability – for Aristotle one that involves deliberation – to make judgements about what conduct is conducive to – in the sense of being constitutive of – the good life. Thus, in private life, a practically wise man

is supposed to be characterized by his ability to reach sound conclusions in his deliberations about what is good for himself.... not in one department of life... but what conduces to the good life as a whole (*ibid.*: 176).

In the same way, in public life, Pericles and others like him have the reputation of being “men of practical genius” because they “have the power of seeing what is good for themselves and for humanity” (*ibid.*: 177). Similarly, Aristotle adds, we

consider practically wise those who “display an aptitude for governing a household or a state” (*ibid.*: 177).

Finally, Aristotelian practical wisdom has this important characteristic: It is not simply about knowing which general norms of action are conducive to the good life. It is also – one might even say, ultimately – about identifying which *particular action* in a given *particular situation* is conducive to the good life:

Prudence is something more than a knowledge of general principles. It must acquire familiarity with particulars also, for conduct deals with particular circumstances, and prudence is a matter of conduct (*ibid.*: 180).

Practical deliberation about particulars, Aristotle insists, is not merely an intellectual enterprise, but involves a kind of “perception” which can only be developed through many years of practical experience (*ibid.*: 182).

Concern with the particular, and learning through practical experience rather than through a purely intellectual endeavour, are among the features which practical wisdom shares with *techné*. Indeed, whereas from one point of view Aristotle distinguishes between *techné* and *praxis*, from another he groups the two together and contrasts both with another type of activity, *epistémé*, science, or more precisely, *epistémé theoretiké*, theoretical or contemplative science. *Techné* and *praxis* are both concerned with bringing about new states of affairs by changing particular concrete things which are contingent and changeable through human action. Science, by contrast, is about knowing things that are universal, necessary and not changeable through human action:

Science is the coming to conclusions about universals and necessary things (*ibid.*: 178).

Clearly, Aristotelian science is not oriented to intervening in the world in order to change it in a technical, practical or any other way. Both Hannah Arendt and Hans-Georg Gadamer adopt the Aristotelian distinction between making and doing, and, having impressed their own interpretative stamp on Aristotle's terms, make extensive use of them in their own work. Some of their interpretations are reflected in – and might therefore illuminate – Habermas's own corresponding conceptual duality between *techne* and *praxis*.

Hannah Arendt divides *vita activa* – human “active” life, as opposed to *vita contemplativa*, contemplative life – into three types of activity: (1), labour, (2) work or fabrication, and (3) action. The latter two correspond, by and large, to Aristotle's *poiesis* and *praxis*. Of these, it is Arendt's concept of action which warrants our special attention. This is how she conceives of action, in contrast to work or fabrication:

Action, the only activity that goes on directly between men without the intermediary of things or matter, corresponds to the human condition of plurality, to the fact that men, not Man, live on the earth and inhabit the world (Arendt, 1958: 7).

and,

Action, as distinguished from fabrication, is never possible in isolation; to be isolated is to be deprived of the capacity to act. Action and speech need the surrounding presence of others no less than fabrication needs the surrounding presence of nature for its material and for a world in which to place the finished product. Fabrication is surrounded by and in constant contact with

the world: action and speech are surrounded by and in constant contact with the web of the acts and words of other men (*ibid.*: 188).

Thus, whereas work concerns above all a relationship between human beings and the material world, action in Arendt's understanding is essentially *interactive* and closely related to *language*. One cannot engage in action, just as one cannot use language, outside a network of human relations. Action, *praxis* – just as much as language – cannot be a solitary, asocial activity.

Gadamer's interpretation of Aristotle's dichotomy between *poiesis/techne*, on the one hand, and *praxis/phronesis* on the other emphasizes another point of difference: one concerning the significance, in each type of activity, of the concrete act; or, more precisely, the relationship, in each, between the concrete act and the general idea held prior to, and guiding, the concrete act. It has already been pointed out that making and doing, *techne* and *praxis*, – in contrast to science – are both primarily concerned with concrete acts, carried out in concrete situations. In this very respect however, according to Gadamer, the two are also fundamentally different.

In *poiesis*, the idea of the object to be made is present in the mind of the maker in a complete, determinate form, prior to the concrete act of making. Thus, *techne* is about the application, through some kind of technique, of an idea fully determined outside the "technical" action itself. In the case of *praxis*, on the other hand, there are only incomplete, indeterminate prior ideas, giving only general guidance to concrete action. The precise determination of the content of any concrete "practical" action, therefore, takes place in the concrete situation where the action is called for, and constitutes at the same time the assignment of a precise

meaning, applicable to that particular situation, to the prior, not fully determinate, idea guiding the action.

Consequently, *phronesis*, or practical wisdom (rendered by Gadamer as "practical" or "moral" reasonableness), being concerned with the deliberate determination, by the acting agent, of what to do in specific situations, always involves over and above possessing a "sketchy" [Gadamer's term] notion of the good life, also a further irreducible evaluative, or normative element: an element of concrete choice, or decision (*proairesis*) about what is good – an aspect of the good life – in a particular situation.

The following passages express eloquently Gadamer's understanding of this difference between *techne* and *praxis*.

(1) The image that the person has of what he ought to do, thus, for example, his concepts of right and wrong, of decency, courage, honor, solidarity and so on... are guiding images in a certain sense, to be sure. But still a fundamental difference is recognizable from the guiding image that the blueprint of an object to be produced represents for the artisan; what is right, for example, is not fully determinable independently of the situation that demands the right from me, whereas the *eidos* of that which an artisan wants to produce is certainly fully determined, namely by the use for which it is intended. (Gadamer, 1975: 283).³

(2) To speak with Aristotle, the conclusion of the practical syllogism and of practical deliberation is the resolve. This

resolve, however, together with the whole path of reflection, from the willing of the objective to the thing to be done, is simultaneously a concretization of the willed objective itself... With [practical reason] the point, which is delimited against any technical rationality, is that the aim itself, the “universal” derives its determinacy by means of the singular (Gadamer, 1981: 81).

(3) The knowledge that gives direction to action is essentially called for by concrete situations in which we are to choose the thing to be done; and no learned and mastered technique can spare us the task of deliberation and decision (*ibid.*: 92).

In the language of means and ends, this difference between making and doing can be put as follows: In making, the concrete action producing an object is merely the means to the production of something that was conceived prior to the action – in a fully determinate form – as the action’s end or purpose; in doing, by contrast, the concrete action is itself a fully determinate end chosen through practical deliberation as a situationally appropriate form of a value or norm adopted prior to the action as a general, not fully determinate, end.

The importance of the distinction between *techne* and *praxis* for Habermas cannot be exaggerated. The analysis of the demise of that ancient distinction in modern thinking and social practice – in the form of the “technocratic pathology” – together with his sustained efforts to redefine and reestablish the distinction for modernity, rehabilitating *praxis*, constitute a central axis of Habermas’s philosophical and social-theoretic endeavours.

Looking, first, into Habermas's work prior to the communicative turn, we come across various formulations of the distinction, each emphasizing different aspects of it, albeit all in keeping with the spirit of Aristotle, Arendt and Gadamer. To begin with, he differentiates between "technical" questions, i.e. questions pertaining to *techne*, and "practical" questions, i.e. questions pertaining to *praxis*, in the following way;

Technical questions are posed with a view to the rationally goal-directed organization of means and the rational selection of instrumental alternatives, once the goals (values and maxims) are given. Practical questions, on the other hand, are posed with a view to the acceptance or rejection of norms, especially norms for action, the claims to validity of which we can support or oppose with reasons (1974: 3).

Here, the distinction is made in terms of the different ways in which the two kinds of activity relate to means and ends. *Techne* is about finding, and employing, appropriate means to given ends. It is this – the means-ends – sense of technical action that Habermas conveys by the frequently used terms "goal-directed", "purposive", "purposive-rational" and "instrumental" action. *Praxis* on the other hand concerns the taking of reasoned decisions about the ends themselves. Accordingly, *praxis*, frequently referred to as the "*praxis* of life", is associated by Habermas with ethics and politics as well as with moral and political philosophy. It is worth noting that in this formulation of the distinction, ends, or goals – whether given, as in the case of *techne*, or chosen, as in the case of *praxis* – are rather general and abstract ends (values, maxims, norms) than concretely situated, specific ones.

Another important aspect of *techne* and technical action is brought out by expressions such as “technical” control over (objective and) objectified processes”, which occur frequently in Habermas’s texts (e.g. 1971: 57; 1974: 275, 269). This emphasis on control and mastery is in line with Habermas’s understanding of Greek *techne* as “the skillful production of artifacts and the expert mastery of objectified tasks” (1974: 42) and accords with Aristotle’s conception of *poiesis* as bringing about changes in reality through the harnessing of efficient causes. It is in connection with this property that Habermas often associates technical action with “work”, or “labour”, understood as the type of activity through which human beings exercise control over natural processes for the production of goods. However, it would be a gross misinterpretation to think that Habermas restricts technical action to humanity’s relationship with nature. Far from it, it is of paramount importance for his Critical Theory that technical action is applicable – and indeed applied, often objectionably so – to society as well. It concerns “the technical control over the objects or... objectified processes of nature and society” (1974: 264).

Coming now to the side of *praxis*, we also encounter a second aspect, over and above its being concerned with examining and settling evaluative, normative, goal-setting questions; namely the interactive nature of practical action. Echoing Hannah Arendt, (without, though, distinguishing between labour and work as she does) but also drawing on the early Hegel,⁴ Habermas contrasts labour, work, technical action – which involves a relationship of a subject to an object – with interaction, which he conceives as the reciprocal relationship between mutually recognizing subjects (*ibid.*, p.160). Intersubjective interaction is mediated through linguistic communication. Indeed Habermas used the terms “interactive” and “communicative” interchangeably long before he worked out the thoroughly communicative philosophy that found expression in *The Theory of Communicative Action* (Habermas, 1984, 1987).

These and certain other elements that for Habermas comprise the technical and the practical types of activity are brought together in the respective definitions of “work” and “interaction” given in the important essay ‘Technology and Science as “Ideology” ’ (1971: Chapter 6).

By “work” or *purposive-rational action* I understand either instrumental action or rational choice or their conjunction. Instrumental action is governed by *technical rules* based on empirical knowledge. In every event they imply conditional predictions about observable events, physical or social... The conduct of rational choice is governed by *strategies* based on analytic knowledge. They imply deductions from preference rules (value systems) and decision procedures;... Purposive-rational action realizes defined goals under given conditions. But while instrumental action organizes means that are appropriate or inappropriate according to criteria of an effective control of reality, strategic action depends only on the correct evaluation of possible alternative choices, which results from calculation supplemented by values and maxims (1971: 91-2).

Some noteworthy points emerge in relation to this definition. Purposive-rational action is, according to Habermas, of two kinds: (a) “instrumental action” and (b) “rational choice”, often also referred to as “strategic action”, in a sense of this term very different from the one given to it in Habermas’s later work. Both kinds have to do with means-ends relations. In instrumental action ends are realized causally, through the effective control of physical or social reality, a control that is achieved

on the basis of predictive empirical knowledge. In strategic action, on the other hand, the relationship between means and ends is an analytic one. The “choices” between alternative means are not “evaluative” ones; they are not choices involving decisions about the good life, as in Aristotelian *praxis*. They are, rather, choices fully determined by deductive calculations made on the basis of non-evaluative decisions. This strategic type of purposive-rational action seems to reflect an acknowledgement, on Habermas’s part, of the internal, conceptual and logical complexity of the realm of “ends” (values, interests, norms, etc).

In contradistinction to “work” or “purposive-rational action”, Habermas defines “interaction” in this way:

By “interaction”... I understand *communicative action*, symbolic interaction. It is governed by binding *consensual norms* which define reciprocal expectations about behaviour and which must be understood and recognized by at least two acting subjects... while the validity of technical rules and strategies depends on that of empirically true or analytically correct propositions, the validity of social norms is grounded only on the intersubjectivity of the mutual understanding of intentions and secured by the general recognition of obligations (*ibid.*: 92).

Thus, at the opposite pole to purposive-rational action Habermas installs intersubjectivity: Interaction is communicative action, symbolic exchange, between subjects. This form of intersubjective activity relates to norms in two different ways, (though, in the above passage, the two are not differentiated with sufficient clarity). On the one hand, communicative action is bound by norms that are mutually

acceptable to the interacting subjects; on the other hand, communicative action is the way in which norms come to be accepted as valid, and thus as binding. Though these two aspects of communicative action – that is (a) being governed by norms and (b) determining the validity of norms – go hand in hand in Habermas's work (with relative emphasis varying according to context) it is only the latter that has a genuinely evaluative character, which is the hallmark of *praxis*, as distinct from *techne*.

1.2.2 The modern triumph of *techne* and demise of *praxis*.

A great deal of Habermas's early work was devoted to tracing, and working out the philosophical basis for reversing, what he saw as the unfortunate demise of *praxis* and practical reason in modern times. For the ancient distinction between *techne* and *praxis* has been gradually reduced, both in social thought and social practice, to a near one-dimensional state of affairs. *Techne* and instrumental reason, empowered through an alliance to a new, technically-oriented form of theory, not only has achieved triumphant success in its own traditional field of activity, but has also deeply encroached on realms properly belonging to *praxis* and practical reason. This encroachment has indeed been so extensive that *praxis* and the practical are virtually squeezed out of existence.

In tracing this line of development, Habermas concentrates in particular on the changing understanding and practice of politics. For Aristotle, the *polis* is not an association formed by members of society in order to secure means for promoting their private interests of survival and security; that instrumental pursuit is the business of private social life, in the case of the Greek city the business of the *oikos*, the household. (Habermas, 1974: 47) The *polis* is, rather, a community concerned with promoting the virtue and good life of its members. The constitutional and legal arrangements of the state are studied and judged as good

or bad from the point of view of this ethical criterion. Politics – the activity of the *polis* – is continuous with ethics; a form of *praxis* as opposed to *techne*.

This Aristotelian conception of politics was dominant in the Greek and Roman world and survived, albeit in a significantly modified form, in the European Middle Ages up until the Renaissance, and was, according to Habermas, finally defeated by the modern technicised conception only in the 19th century. It did, however, undergo a significant transformation already in theocratic medieval Europe, as is shown in Aquinas's work. Thomas Aquinas, writing in the 13th century, like Aristotle conceived of the state with respect to its purpose not as an institution concerned, in an ethically neutral manner, merely with the private interests of individuals and groups; but as a moral order, in continuity with ethics, concerned with the virtue of its members. With respect to the manner in which it operates, however, the political order as conceived by Aquinas is not, as in Aristotle, "anchored in the *praxis* and *lexis* of free citizens, in public political life" (*ibid.*: 49). Instead, the Thomist state is an autocratic social-moral order largely true to the reality of Feudal Europe, and assumed to be in accordance with both Natural Law and biblical revelation. But the technicized conception of politics first appeared during the Renaissance, in the work of Machiavelli and Thomas More, and was presented in a more complete form in early modernity, in the works of Hobbes.

Machiavelli and More

In the Renaissance we have the first signs of a transformed, instrumentalist understanding of politics, which Habermas – in an interpretation which, as he admits, is "not without a degree of stylization" – finds in Machiavelli's works, especially *The Prince* (1513) and in More's *Utopia* (1516). In contrast to both Aristotle and Aquinas, Machiavelli and More divorce politics from ethics; in Habermas's words, "they extricate the structure of domination from its ethical

context” (Habermas, 1974: 52). They both view politics as being concerned not with establishing conditions that would enable individuals to lead a virtuous life, but first and foremost with securing the fundamental need of survival. They treat it as an activity that is morally neutral, at best related to ethics only as its precondition.

The two political thinkers, however, differ greatly in their views about what constitutes the main threat to life and, consequently, how best to confront that threat. For Machiavelli the prime threat is “violent death at the hand of one’s neighbour” (*ibid.*: 51) – a neighbour turned enemy, either from within one’s own society or from outside it. This danger is, for Machiavelli, a permanent feature of the human condition, stemming from the inevitability of human aggressive behaviour. To counter this ever present threat, Machiavelli recommends to the Prince political techniques whose purpose is “the assertion of princely power externally as well as the unity and obedience of its citizens internally” (*ibid.*: 55). The employment of such techniques is the main object of politics, and success in this employment the epitome of political – as distinct from ethical – virtue.

Thomas More, by contrast, singles out death by starvation as the greatest evil, avoidance of which must be the primary purpose of politics. The danger of dying from hunger, however, is not for More an immutable condition of humankind; it is, rather, a condition brought about by “compulsion toward exploitation which is established together with private property” (*ibid.*: 53). Consequently, More’s *Utopia* proposes a form of social organization which, if established, More believes would not produce the human compulsion toward exploitation which is responsible for the threat of death by starvation; it would be an effective means toward the preservation of life.

Thus both Machiavelli and More “deal not with practical questions, but with technical ones” (*ibid.*: 54); they transform politics from *praxis* to *techne*. This is so

not only in the sense that they are concerned with instrumental, means–ends relationships but also in that their political *techne* aims at affecting human conduct itself; as Habermas puts it, “it has – and this was inconceivable for the Ancients – human behaviour rather than nature as the material on which it operates” (*ibid.*: 59).

Finally, Machiavelli’s and More’s technicized politics share one more feature, for Habermas a shortcoming, which a century and a half later, Hobbes’s social philosophy would try to overcome. Both Machiavelli’s and More’s political thinking is untheoretical just as both *praxis* and *techne* were for the ancients who conceived theory as a purely contemplative activity unrelated with change-oriented action. The technical-political recommendations of Machiavelli and More “completely lack the scientific precision of *calculated technique*” (*ibid.*: 60). They are empirical, “pragmatically attained” rules, rather than conclusions drawn from a systematic body of scientific theory. Habermas refers in this connection to “Machiavelli’s book of recipes for the technically correct calculus of power”.

Thomas Hobbes

It was Hobbes, above all in his *Leviathan* (1651), who would first attempt to ground political recommendations (of a technical/instrumental character as in Machiavelli and More) on systematic theory, thus becoming “the founder of social philosophy as a science” (Habermas, 1974: 56). Hobbes is separated from his two Renaissance predecessors by a momentous intellectual development: the foundation by Galileo of modern natural science, a body of theory that would reach maturity – after the publication of *Leviathan* – in Newtonian mechanics. Galileo’s achievement was made possible through the application of a new scientific method that constituted an ingenious synthesis of systematic experimentation and observation on the one hand and mathematical methods on the other, two

elements each of which was, at around the same time, one-sidedly proposed as the sole pillar of scientific method, in Francis Bacon's empiricist-inductivist and Descartes' rationalist-deductivist philosophies of science, respectively.

What is radically new in this modern, Galilean method of studying nature compared to that of the ancients is, according to Habermas, that it seeks the kind of knowledge that can be used technically to bring about desired changes in nature: "The modern, scientific investigation of nature set about to pursue theory with the attitude of the technician" (*ibid.*: 61). But this interest in technically exploitable knowledge is not merely a "psychological" interest – one that happened to exist in the minds of modern scientists and philosophers, although such pioneers of modern science as Bacon and Descartes made no secret of their ambition to establish a science that would give human beings power over nature. Rather, it is a "transcendental" interest (see *ibid.*: 287, note 35), an interest logically presupposed by and imprinted on the structure of modern scientific theory itself: "From the days of Galileo on, theory is measured by its capacity for artificially reproducing natural processes. In contrast to episteme, it is designed for 'application' in its very structure" (*ibid.*: 61). Ironically, however, the first real conjunction of scientific theory and technique was to take place, in Habermas's view, in the field of social philosophy, to be precise in the work of Hobbes – whereas any significant technical application of natural science itself had to wait for another three centuries until the second Industrial Revolution of the latter part of the 19th century.

Taking seriously Galilean science's aim of providing knowledge that is technically usable with respect to nature, and insisting on the use of "a clear and exact method", Hobbes sought to establish a social theory which could empower a

political technique, one whose object would be not the realm of natural phenomena but human conduct itself.

Hobbes begins establishing a “modern physics of human nature” in the form of a “science of the state of nature” (*ibid.*: 64). He depicts the state of nature – that is the human condition in the absence of any political institutions – as one of the war of all against all, a condition where beings lead lives that are “solitary, poor, nasty, brutish and short” (Hobbes, 1962: 43) and are dominated by constant fear of violent death. This fear brings about – “by a certain impulsion of nature no less than that whereby a stone moves downward” (Habermas, 1974: 64) – a universal desire to escape from the fearful situation of permanent threat of death. This state of affairs, Hobbes continues, necessitates the creation, through a social contract among human beings, of an order of Law with punitive power – a Leviathan – that would bring an end to the fearful state of nature by coercing people through the use of sanctions to restrain their aggressive natural impulses and conduct themselves according to the prescriptions of the Law instead.

Comparing the relationship of theory to practice in Hobbes – as interpreted by Habermas – and in modern science, we can identify one major similarity and one major difference. The similarity concerns the way in which theory is used in order to bring about desired changes in the object which the theory is about. In both cases, the realization of certain ends is pursued through the technical control of the object on the basis of theoretical knowledge about it. This knowledge concerns regularities in the behaviour of the object and enables its possessor on the one hand to predict how the object will behave given a certain intervention in a given initial state, and on the other hand to bring about desired states of the object through appropriate intervention into the object’s given state. Thus, by analogy to the modern technologist’s science-based interventions in nature, a Hobbesian

political agency can exercise control on human conduct through interventions in the form of the use or the threatened use of legal sanctions that are determined on the basis of knowledge about human nature; more specifically, on the basis of knowledge about regularities concerning human desires, fears and beliefs and how these relate to human actions.

The dissimilarity between Hobbes and natural science-based technology lies in the origin of the ends that are pursued in the two cases (albeit they are both pursued by means of technical control). The ends of technology – the construction of a bridge or a new weapon, the design of a new computer programme, the development of a new, genetically engineered, type of grain – are in no way necessitated by or derivable from the scientific knowledge employed in the technical efforts to realize them, though, to be sure, their choice is made on the assumption that the scientific knowledge necessary for their realization either exists or is feasible. In Hobbes's theory, by contrast, the ends served by the legal order of the Leviathan which is established through the social contract, are thought to be necessitated by the laws of human nature itself, the very same laws that underlie the sanctions and other means that the Leviathan employs to control human conduct and thereby realize its ends.

Habermas objects to Hobbes's understanding of both (1) the relationship between the state of nature and the social contract and (2) the manner in which human conduct is influenced by the political and legal order. On both counts he criticizes him for being blind to the specific nature of *praxis* and its essential difference from causality and *techne*.

With regard to the first issue, Habermas argues, Hobbes begins with an "analysis of the natural state of the human species prior to all sociation [which] is not ethical at all [but] is purely physicalistic" (Habermas, 1974: 65), an analysis in terms of the

“apparatus” of sensation and “instinctive...casually determined modes of reaction” (*ibid.*: 65). He then moves to “commandments of natural reason” requiring the establishment of “rules of social life” that would facilitate the “satisfaction of needs” (*ibid.*: 66). The problem, Habermas maintains, is that Hobbes conceives this transition from the state of nature to commandments of reason about social rules as *causally* necessitated, whereas it can be properly understood only as a transition dictated by *practical* necessity. The mistake is covered by Hobbes’s (*ibid.*: 67) equivocal use of the term “law of Nature” in both a causal and normative sense: He calls “law of Nature” both “the *causal* connections of man’s asocial instinctual nature *prior* to the contractual constitution of society and the state; and the *normative* regulation of their social cohabitation *after* this constitution” (*ibid.*: 66).

Once the normative character of social rules is grasped, it becomes clear that they cannot be caused by human desires but must be adopted through deliberate decisions made by human beings as subjects (the very same human beings who in the science of human nature which informs their decisions are treated as natural objects). It is this moment of decision, of *praxis*, that is, according to Habermas, missing from Hobbes’s theory. Ironically, as Habermas points out, in the absence of *praxis*, Hobbes’s own theory would be “impotent – just like any other theory that failed to understand the nature of *praxis*; its own recommendations could not be implemented in a causal manner unless “generally published and accepted by the mass of the citizens” (*ibid.*: 73).

Habermas’s objection to Hobbes’s treatment of the relationship between the political-legal order and the behaviour of citizens is that he construes that relationship in technological terms, failing to see that “control over the processes of nature is essentially different from control over social processes” (*ibid.*: 75).

Technological control of nature is a manipulative process in which nature itself plays an entirely passive role. This model of control, Habermas argues, is inappropriate for redirecting human conduct, for in this case, the objects of the desired change are human beings themselves who – unlike nature – cannot be passive objects of manipulation but are actively involved in the process. Influences on human conduct are not possible without “a prior mediation through the consciousness of the citizens who discuss and act...” (*ibid.*: 75). Once more, *praxis* disappears from view. Thus, according to Habermas’s interpretation, at Hobbes’s hands far more systematically than at the hands of either Machiavelli or More, politics as practical activity is dealt a massive blow: its role as a process of discussing and deciding on norms and values is replaced by natural causality, whereas its capacity as a process of change is transformed from a form of human interaction into a theoretically informed technique.

However, at the same time as he laments the demise of *praxis*, Habermas acknowledges an important gain in Hobbes, an altogether modern achievement: the introduction of scientific rigour in social and political theory. Thus for Habermas the way forward after Hobbes is to restore to politics “the dimension of *praxis* to which the classical doctrine offered direct access” (*ibid.*: 74), but do so without sacrificing the genuine gains in theoretical rigour, without reverting to the ancient form of *praxis* that relied merely on non-theoretically- rigorous *phronesis*

Enlightenment: Natural Law and the Public Sphere

The age of Enlightenment is hailed by Habermas as marking a significant advance in this respect, with the emergence of the “bourgeois public sphere”, a new historical phenomenon which Habermas analyzed, and in certain respects idealized and celebrated, in his first major work, *The Structural Transformation of the Public Sphere* (1989).

Conceived as “the sphere of private people come together as a public” (Habermas, 1989: 27), the bourgeois public sphere presupposes a civil society – the “realm of commodity exchange and social labour” (ibid.: 30) – separate from and independent of the state, and, similarly, private and autonomous individuals who do not depend for their own and their family’s livelihood on political authority. Clearly, we are referring to a phenomenon that grew on the soil of a well developed capitalist market economy.

The bourgeois public sphere developed in two main areas, the world of letters (mainly literature and literary criticism) and – most importantly – the world of politics, and was characterized by the “rational-critical public debate” of important issues of general interest. The ideal of this form of public rationality that constituted the mode of operation of the public sphere concerned the formation of “public opinion” through human interactions based on the public use of reason, with a view to influencing developments in society, not least political decisions. It was undoubtedly a realm of *praxis*.

Furthermore, the bourgeois public sphere goes beyond the ancient form of *praxis* in that it is informed by systematic theory, specifically by the Natural Law theories dominant at the time. Here, however, according to Habermas, also lie some of the main shortcomings of the bourgeois public sphere. To be sure, Habermas has a high regard for the formal, procedural aspects of the bourgeois public sphere, the process of critical, rational public debate which Kant philosophically conceptualized for the Enlightenment and he himself has struggled to articulate and defend in much of his work, not least in its later, communicative phase. In its substantive aspects, though, the bourgeois public sphere is thought to be vitiated by its relationship with Natural Law theory.

Habermas discusses two examples of theory in its relation to public sphere practice: those of the French Physiocrats and the British (Scottish) moral and economic philosophers. Both the Physiocrats and the Scottish philosophers viewed civil society as a coherent whole, without major internal fault lines, in harmony with Natural Law. For the Physiocrats the public sphere remained outside their theory and was treated as its “practical complement”. Its role was to draw, with the help of philosophers and through the public use of reason, the political implications of their Natural Law theory of civil society and try to transmit those implications to political authorities who could implement them: “[scholars] determined the public opinion, [governors] converted into practice whatever conclusions were drawn from the critical reflection of the public guided by experts” (Habermas, 1989: 95). In the case of pre-revolutionary France, a public sphere informed by a theory such as that of the Physiocrats would push the state to transform itself so as to be in greater harmony with civil society. In the case of the Scottish thinkers, the public sphere was seen as part and parcel of bourgeois society – the object of their theory – and was thus incorporated into the theory itself. The bourgeois political public sphere followed the evolution of bourgeois society, “[unfolding] to the same degree as the natural laws of the market assert themselves” (Habermas, 1974: 78).

A public sphere informed by such a theory would also strive to bring the political system into line with bourgeois civil society; or, as in the post-revolutionary British situation where civil society and state were largely in harmony, it would strive to preserve and fine-tune that harmony through influencing day to day political decisions. Thus, in these and other analogous cases, social theory, failing to detect fundamental internal contradictions of its object – capitalist society – takes that society for granted, idealizing it, moreover, as a natural order; and, to the extent that it informs and affects the *praxis* of the public sphere, it does so in an

one-directional way: in the direction of the public sphere accepting and conforming to the theory's object. Responsible for this, Habermas maintains, is the fact that Natural Law social theories were constructed on the technically oriented model of natural science and could not properly conceive of political action as *praxis* and of their own relationship to it in accordance with such a conception. For social and political action to be informed by rigorous theory and at the same time possess the character of genuine *praxis*, we need, according to Habermas, a social and political theory different in structure from that of modern natural science.

Hegel and Marx

Habermas recognizes precisely such a theory in the dialectical aspects of the work of Hegel and Marx. Hegelian and Marxian dialectical social theory does not objectivistically describe its object from the outside, from the point of view of a subject external to it. It self-reflectively understands itself as being actively involved in its own object; as part of social *praxis* and at the same time as depending on social *praxis* for its validation. It is a theory that recognizes deep internal contradictions in its object, reveals these contradictions to the object itself and becomes engaged in the social *praxis* through which the object tries to overcome them by transforming itself. This kind of theory is thus "critical" of its own object. Above all it is critical of the ideology that conceals from the object – society – its own true being. The theory's intention – and criterion of validity – is not faithfully to describe its object but to relate to *praxis* so as to facilitate its object's self-development. The role of social *praxis*, conversely, is not to apply or conform to theory, as in the case of the relationship of practice to scientific and Natural Law theory, but to render the theory valid by realizing its critical intentions.⁵ It is precisely along these lines that the early Habermas, following the footsteps of the

first-generation Frankfurt School, develops his own conception of *praxis* and *praxis*-related critical reason.

Hegel's and Marx's theories themselves, however, in Habermas's estimation, went astray. The old Hegel "relinquished the dialectical relation between [theory and *praxis*]" (Habermas, 1974: 130). Just like the owl of Minerva that flies only when the day is over, philosophical theory now comes always after the event, capable of making sense of it, but not being entangled with it and influencing its nature in any way: "Philosophy...arrives too late to instruct, because it comprehends its time in thought only after reality has completed itself – a theory such as that is exempted entirely from *praxis* by the old Hegel" (*ibid.*: 130). Thus – philosophy having returned to the role of *theoria* in the Greek sense – social and political practice is left without theoretical guidance.

The young Marx of the *Paris Manuscripts*, inspired by Hegel's *Phenomenology of Mind*, conceives of human history as a process of self-generation of the human species through social labour, the process by means of which socialized humanity relates to nature (*ibid.*: 168). From this starting point Marx's theory develops in two directions which coexist in an uneasy relationship throughout his work.

On the one hand, we have a critical theory of a contradiction-ridden capitalist economy, a theory with the practical intention of relating dialectically to a social and political *praxis* that would overcome capitalism's contradictions. The most fundamental contradiction that the theory claimed to have revealed is one between the origin of the products of labour in human social activity and the fetishistic appearance of these products to their creators as external and alien objects. This approach can be found just as much in Marx's mature analysis of "commodity fetishism" in *Capital* as in his early analysis of alienated labour in the *Paris Manuscripts*.

On the other hand, Marx develops a materialist theory of history, where historical change is seen as driven by developments in the process of social production. The young Marx conceived of the development of the process of social production as resulting from the interrelations between the instrumental power over nature that human beings employed in the process of production (the “forces of production”, or “labour”) and the social relations holding between human beings engaged in the process of production (the “relations of production” or “interaction”). However, Habermas explains, this “brilliant insight into the dialectical relationship between the forces of production and the relations of production” (*ibid.*: 169)⁶ was interpreted one-sidedly, in favour of the former: “a precise analysis of the first part of the *German Ideology* reveals that Marx does not actually explicate the interrelationship of interaction and labour, but instead, under the unspecific title of social *praxis*, reduces the one to the other, namely: communicative action to instrumental action” (*ibid.*: 168). As a result, Marx’s theory of history could easily be interpreted as a technological determinist one. Indeed, such an interpretation came to overshadow the critical-dialectical aspect of Marx’s work and, in the hands of Marx’s epigones, led to the technocratic political practice that dominated the Marxist socialist movement down to our own times.

From the Bourgeois Public Sphere to Modern Technicized Politics

At the same time, a parallel development is traced by the early Habermas in mainstream bourgeois practice and theory, namely, the degeneration of the highly promising Enlightenment public sphere into a 20th century political and social life dominated by the technocratic ideology.

The demise of the 18th century bourgeois public sphere and, with it, of rational-critical public debate was, according to Habermas, the result of a growing intertwining of state and society, a process that culminated in the Welfare State

of the second half of the 20th century. As we have seen, Habermas considers the strict separation of state and civil society, more precisely of state and the market economy, as the foundation of the bourgeois public sphere. A separate market economy at the level of which the innumerable conflicts of interest between small and relatively autonomous private economic agents were resolved, could still be seen – ideologically – as a realm of equal, or just, exchange. Such a situation allowed the existence of a public sphere that left behind sectional interests and focused on the rational debate of issues of genuinely general interest.

Those conditions were gradually eroded. Capitalist crisis tendencies and the post-industrial-revolution concentration of economic power into the hands of fewer and fewer individuals and organizations, including workers' trades unions (to a large measure initiated by the industrial revolution), made it increasingly difficult to resolve conflicts of economic interests at the level of the market economy, and called for bigger, stronger, interventionist state institutions. Thus, the separation of state and civil society gave way to a situation where the state intervened more and more deeply into the economy and civil society while powerful economic and social agents got increasingly involved in the affairs of the state. Habermas speaks characteristically of a

dialectic of a progressive “societalization” of the state simultaneously with an increasing “state-ification” of society [that] gradually destroyed the basis of the bourgeois public sphere – the separation of state and society (1989: 142).

The key to the degeneration of the bourgeois public sphere and of rational-critical debate – its form of praxial rationality – was, for Habermas, the significant loss of credibility by the ideology of just exchange, which inevitably resulted from the intertwining of state and society.

The more pronounced the economic and social influence of powerful state and private agents became, the harder it was to view the functioning of the economy and society through the lenses of that ideology. The contradictions, injustices and inequalities inherent in society became increasingly visible. Developments such as Hegel's and Marx's theoretical accomplishments, the rise of socialist and other radical movements, and the revolutions of 1848, can indeed be interpreted as early evidence of such an enhanced visibility. Thus, conflicts of interests between holders of concentrated economic and social power – which were now not resolvable at the level of the market – could not be allowed to become objects of rational-critical public debate, despite their unquestionable public relevance. They were, instead, dealt with away from the public arena.

The process of the politically relevant exercise and equilibration of power now takes place directly between the private bureaucracies, special-interest associations, parties, and public administration (1989: 176).

But this form of decision making, together with the whole institutional framework of society, required a new kind of legitimation, in replacement of that formerly provided by the ideology of just exchange. And that need became all the more important as the masses came to acquire a formal political role through the extension, and eventually the universalization, of the right to vote.

The required legitimation can be provided, according to Habermas, by what he refers to as the technocratic ideology, or technocratic consciousness. This is a way of thinking which essentially consists in treating questions of *praxis*, involving normative and evaluative choices, as technical questions to be decided upon by experts, including decision theorists and social system analysts:

The ideological nucleus of this consciousness is the *elimination of the distinction between the practical and the technical* (Habermas, 1971: 113).

Accordingly, politically important decisions are claimed to be adaptive measures technically dictated by “objective exigencies”, the whole organization of society itself being ultimately justified in terms of its alleged superior ability to secure material well being, in particular through continuous science-based technological progress.

On the basis of that ideology even democracy becomes a technical issue:

[T]he process of democratic decision making about practical problems loses its function and “must” be replaced by plebiscitary decisions about alternative sets of leaders of administrative personnel (Habermas, 1971: 105).

Major strands of social theory itself – systems theory, above all – have already developed along such technocratic lines, viewing societies as systems of adaptation to externally imposed imperatives. Thus, referring to Niklas Luhmann’s systems theory, Habermas maintains:

[T]his theory represents the advanced form of a technocratic consciousness, which today permits practical questions to be defined from the outset as technical ones, and thereby withholds them from public and unconstrained discussion (Adorno, 1976: xxxii, quoted from *Theorie der Gesellschaft oder Sozialtechnologie?*, by Habermas and Luhmann).

By concealing the politically important normative and evaluative choices made away from public sight – a concealment that applies, by and large, to the decision makers themselves – technocratic consciousness depoliticizes and at the same time technicizes politics. By the same token it depoliticizes the mass of citizens.

To be sure, Habermas is not claiming that technocracy, as an actual mode of decision making or as a form of ideological consciousness, is today anything other than a partly realized potentiality. Citizens in the formally democratic and relatively affluent contemporary societies still have an ample range of opportunities to make politically significant decisions, both as voters and as consumers of material and cultural goods; and they are often still able to see these decisions not as technically dictated but as involving genuine normative and evaluative choices. To that extent, however, Habermas observes in modern societies further tendencies – mainly complementary to the technocratic ones – in the direction of the technicization of politics: tendencies toward the technical manipulation of people's voting and consumer behaviour and of public opinion more generally.

In the new situation, members of the public are less and less participants in rational-critical debate aiming at shared political views and decisions, and increasingly isolated, private individuals who form their opinions and make their choices in their capacity as separate voters and consumers. These opinions and choices are systematically influenced by sophisticated techniques of manipulation such as advertizing, marketing and public relations; techniques that are employed in the realm of politics as much as in the realms of culture. Aided by technically usable scientific disciplines such as psychology, such manipulation techniques normally work through individuals' unconscious. And despite the influence they exert, they are used by agencies which are in no way publically accountable for their activities. Speaking specifically of the manipulation of public opinion in order

to secure a certain behaviour by voters, Habermas points out what he considers to be true of manipulation techniques in general:

Especially manipulative are social-psychologically calculated offers that appeal to unconscious inclinations and call forth predictable reactions without on the other hand placing any obligation whatever on the very persons who in this fashion secure plebiscitary agreement (1989: 217).

As the main agents of public opinion manipulation (beyond the immediate culprits, i.e. the experts in advertizing, marketing and public relations) Habermas identified special interest groups, state administration, the political parties, and the press; the latter being accused of having changed "from being a merchant of news to being a dealer in public opinion" (ibid.: 182).

This transformation of politically significant decision making from an object of rational-critical public debate to an object of technical manipulation is associated by Habermas with the theoretical position he refers to as "decisionism": the view – prominently advocated by Hume and Weber – that normative and evaluative choices are not amenable to rationality, but must be made on the basis of non-rational decisions.

But even with this, decisionistic-manipulative, aspect of the technicization of politics added onto the technocratic one, the early Habermas still considers the future of politics and the public sphere in modern society as open-ended. In contrast to the first generation Frankfurt School thinkers, he sees modern societies as far from being completely engulfed in technical rationalization, discerning possibilities for the repoliticization of the masses and the regeneration of a *praxis*-oriented political public sphere; one that would now be guided by a

new, more adequate form of praxial rationality, which he confidently considered philosophically possible and persistently tried to articulate.

1.2.3 Habermas's early conception of praxial rationality

As we have seen in the previous section, Habermas, from the very beginning, expressed his strong belief in the philosophical and social possibility of praxial rationality. His first notion of such a rationality is that of "rational-critical debate", the kind of rationality that in his view characterized what he identifies as the first major modern form of social *praxis*: the eighteenth-century "bourgeois public sphere". Habermas thought highly of the formal-procedural aspects of rational-critical debate, broadly understood in terms of Kant's analysis of the "public use of reason". With regard to its substantive aspects, however, he was critical of its close association with Natural Law theories. His main objection was that the objectivist structure of those theories – leaving aside their bourgeois ideological content – tended to deprive participants in rational debate of any genuine role in determining their own identities and values, which for Habermas is an essential characteristic of true *praxis*.

Thus, in order to develop his own conception of praxial rationality, the early Habermas turns to a different kind of social theory, initiated – but in his view not employed consistently – by Hegel and Marx: a non-objectivist theory, whose aim is not simply to describe its objects (human beings or society) accurately, from an observer's point of view, but to remove obstacles to their self-understanding and thereby enable them to transform themselves in self-determined ways. What he had in mind was in fact close to what the Frankfurt School of the 1930's, most notably Max Horkheimer, called Critical Theory (Horkheimer, 1972).

Habermas broadly articulates his understanding of such a form of praxial rationality – mainly under the terms “dialectics” and “critique” – in essays such as “The Analytical Theory of Science and Dialectics” (in Adorno, 1976) and “Between Philosophy and Science: Marxism as Critique” (in Habermas, 1974). But that concept is fully developed only in *Knowledge and Human Interests*, Habermas’s major, systematic exposition of his early paradigm of thought.

The starting point of *Knowledge and Human Interests* is the rejection of “objectivism”; that is, of the “interpretation of knowledge as a copy of reality” (1972: 89), which Habermas identifies as the view of scientific knowledge that not only prevails in society at large but dominates the self-understanding of the sciences themselves as well as much philosophical thinking on the subject. Habermas, by contrast, in a Kantian manner, considers scientific systems of knowledge as “result[s] of interaction between the knowing subject and reality” (*ibid.*: 90); albeit, he views that subject not (like Kant) as an individual consciousness but (like Marx) as the whole of humanity conceived as a species that reproduces itself in history through its own action.

Thus, scientific knowledge is not knowledge of reality as it is itself, but of reality as it reveals itself when looked at from particular points of view. These points of view reflect human species interests which are *a priori* encoded in the very structure of the sciences. To identify those *a priori* elements of the sciences, Habermas undertakes a Kantian-type, “self-reflective” (*ibid.*: 212) “critique of knowledge” which, due to the connection between knowledge and human action, is at the same time a “critique of rational action” (*ibid.*: 17). Thus, he reflects on “the transcendental framework[s] of processes of inquiry” pertaining to different kinds of science, which establish “[t]he conditions of the objectivity of possible experience” within those sciences (*ibid.*: 195). Through this reflection, Habermas

differentiates between three categories of science, each governed by a different “knowledge-constitutive” or “cognitive” interest: 1) The “empirical-analytic” sciences are governed by the *technical* cognitive interest and disclose their object-domains from the perspective of possible technical control. 2) The “historical-hermeneutic”, or “cultural” sciences are governed by the *practical* cognitive interest and are oriented to communicative understanding. 3) The “critical” sciences are governed by the *emancipatory* cognitive interest and are oriented to critical self-reflection (1972: 308-311).

Knowledge-constitutive interests are “transcendental”, in the sense that they are a *priori* conditions of possibility of scientific knowledge. Habermas wants them to be clearly distinguished from contingent, empirical interests of scientists or social groups, which also exist and undoubtedly influence the specific directions taken by scientific research. But knowledge-constitutive interests, together with the *a priori* structures of scientific inquiry that reflect them, are transcendental in a sense different from Kant’s: They “have a transcendental function, but they determine the architectonic of processes of inquiry and not that of transcendental consciousness as such” (*ibid.*: 194). Furthermore, they are not just transcendental. They are at the same time embedded in the form of activity of the human species, which Habermas – following Marx – conceives materialistically as a product of nature. The cognitive interests and the corresponding structures of inquiry are “rooted in specific fundamental conditions of the possible reproduction and self-constitution of the human species” (*ibid.*: 196). More specifically, the technical interest is rooted in the structures of work, or instrumental action, the practical interest in the structures of language-based interaction, and the emancipatory interest in social relations of power.

Thus, the knowledge-constitutive interests are both transcendental and products of nature. To express this dual conception, which, as we shall see, he later came to consider unsatisfactory, Habermas frequently uses the term “quasi-transcendental”.

By being inherently linked to forms of human action, the three types of inquiry differentiated by Habermas – the empirical-analytic, the hermeneutic and the critical – *ipso facto* constitute for him forms of practical rationality, technical and praxial.

1) Empirical-analytic inquiry and technical rationality

According to the mainstream hypothetico-deductive model of science, which is adopted also by Habermas (1972: 113–139), empirical-analytic inquiry develops general hypotheses regarding lawlike, cause-effect connections between observable phenomena. These hypotheses are normally expressible in the form “Under initial conditions C, event A always causes event B”. Hypotheses are accepted (as laws) into the body of theory of a (nomological) science once they withstand empirical testing on the basis of *predictions* of the form “If A occurs under conditions C, then B will occur”, which are deductively derived from them. By employing a law, or theory, established in this way, we can *explain* the occurrence of event B in terms of the occurrence of event A under conditions C.

The rationality of technical action is strictly analogous to the above logic of scientific prediction and explanation: Since A under conditions C always causes B, B can be brought about if A is brought about under conditions C. Thus, if B is our goal, we can attain it by means of an action that produces A under condition C. Evidently, the (inherent) technical exploitability of the empirical-analytic, nomological sciences lies in that, by (inescapably) providing us with knowledge of

lawlike, cause-effect relationships pertaining to particular objectified processes, they enable us to causally control and manipulate those processes so as to attain our ends.

2) Hermeneutics and critical self-reflection: Forms of praxial rationality

As we have seen, for Habermas the rationality of *praxis* concerns not the causal manipulation of objectified processes for attaining pre-decided ends but the determination by human beings of the very ends, values and norms to which their actions should be oriented. The hermeneutic form of inquiry that in his view constitutes the methodological structure of the cultural sciences is said to be governed by the practical cognitive interest, that is by the interest in the mutual understanding between communicating subjects. As such, hermeneutics is a methodical dialogic process, based on the use of ordinary language, whereby evaluative and normative meanings, among others, can be clarified and agreed upon (1972: 161–186). It is thus for Habermas unquestionably a form of praxial rationality. When it comes to his more detailed understanding of this hermeneutic form of praxial rationality, Habermas is in essential agreement with Gadamer, particularly his analyses in Part II of *Truth and Method* (Habermas, *ibid.*: 348, n9).

Hermeneutics, however, whether of a Gadamerian or other equivalent variety, though undoubtedly an important form of praxial rationality, is for Habermas inadequate. As we shall see at greater length in the discussion of the Habermas–Gadamer debate in section 1.3.2.2, the fundamental weakness of hermeneutics according to Habermas is that it ultimately addresses only “meaning structures in the dimension of what is consciously intended” (*ibid.*: 217). It is directed at interpreting meanings which are either immediately present to the subject's consciousness or can be brought to consciousness by ordinary linguistic

means. Hermeneutics, in other words, assumes that misunderstandings, including, importantly, self-misunderstandings, are always due to “accidental” or “external” obstacles to communication, which can be identified and removed in the medium of ordinary language.

Habermas does not of course deny that difficulties in understanding and communication are often of this kind and can, therefore, be adequately dealt with by means of hermeneutic methods. He insists, however, that failures in understanding, particularly self-understanding, are frequently “systematic”, originating in deep-seated causes which are internal to the subject and yet inaccessible to it. They are cases of “systematically distorted communication” (see also Habermas, 1970), which are not amenable to hermeneutic analysis. Such cases, Habermas argues, require an entirely different method, one that “joins hermeneutics with operations that genuinely seemed to be reserved to the natural sciences” (Habermas, 1972: 214), or “unites linguistic analysis with the... investigation of causal connections” (*ibid.*: 217). This composite method, which constitutes a stronger form of praxial rationality – complementary to the purely hermeneutic one – is elaborated by Habermas primarily through a study of what he considers to be its best exemplification: Freud’s psychoanalysis (see Habermas, 1972: chapters 10 and 11).⁷ It is said to be a “depth-hermeneutic” method, termed “self-reflection” or “critical self-reflection”. I shall also be referring to it as “critical-emancipatory self-reflection”.

The core logical structure of the method of critical self-reflection (with reference to an individual subject but *mutatis mutandis* applicable also to a social class or other collective subject, including society as a whole) is as follows:

(i) There are systematic distortions in the patterns of understanding/communication/behaviour of a person (probably in the form a neurosis) whose

origins cannot be fathomed – not least by the person in question (henceforth “subject”) – through hermeneutic efforts that employ ordinary linguistic means and are directed at the subject’s conscious mind.

(ii) Someone else (assuming the role of “therapist”, possibly at the request of the subject, now in the role of “patient”) undertakes to generate causal hypotheses, in the empirical-analytic fashion, that would explain the distortions in question. The causes looked for and eventually included in the hypotheses are internal to the subject but inaccessible to its consciousness. In psychoanalytic cases they are typically unconscious motives.

(iii) The aims of an exclusively empirical-analytic inquiry would be to yield knowledge of cause-effect connections *on the basis of which* the behaviour of the object of knowledge (of the “subject”, in our case) could be causally explained, predicted and manipulated. Thus, such an inquiry has an inherent interest in the continued existence of the relevant causal connections. As an aspect of the method of critical-emancipatory self-reflection, however, empirical-analytic inquiry is interested in identifying causal connections not in order to *take advantage* of their existence, but on the contrary in order to *dissolve* them and thereby liberate the subject from the unwanted distortions in communication and understanding which result from them.⁸

(iv) To that end, the inquiry into the unconscious causes of systematic distortions must be joined with ordinary hermeneutic communication. Thus, the inquirer/therapist proceeds to communicate to the subject/patient the hypothesized unconscious cause of the problematical symptoms. If the subject, through self-reflection and possibly not without strong inner resistance, comes to recognize the proposed cause as truly a product of its own self-deceiving mind, the hitherto unconscious cause is thereby brought to consciousness. With such self-reflection

“[d]ePTH-hermeneutic understanding takes over the function of explanation” (1972: 272).

(v) The subject’s self-acknowledged “enlightenment” concerning the self imposed genesis of its own problems is – in the context of the method of critical self-reflection – the only acceptable verification of the therapist-inquirer’s causal hypothesis: “The experience of reflection is the only criterion for the corroboration or failure of hypotheses” (1972: 266).

(vi) Enlightenment in the above sense is closely linked to emancipation from the self-imposed problematical situation. In psychoanalytic contexts unconscious motives are supposed to lose their causal efficacy merely by virtue of being brought to consciousness. In contexts of social critique, however, the transition from enlightenment to emancipation requires organized political action. As we shall see in the critical discussion of Habermas’s concept of praxial rationality, in section 1.3.2.2, the relationship between enlightenment and emancipation, together with that between “therapist” and “patient”, is among the most questionable aspects of the method of critical-emancipatory self-reflection.

1.3 CRITIQUE OF HABERMAS'S EARLY VIEWS ON ACTION AND PRACTICAL RATIONALITY

This section will be devoted to critical – including self-critical – perspectives on Habermas's early (pre-linguistic-turn) ideas concerning action and the rationality of action. The most central of these ideas, as presented in the preceding section, are, in a nutshell, the following

1) There are two, mutually irreducible, types of action. On the one hand, there is the type of action concerned with the pursuit of pre-given ends through the control of objectified processes, which is referred to by the terms *techne*, work, instrumental action or purposive-rational action. On the other hand there is *praxis*, interaction or communicative action, concerned with human relations based on, or aimed at determining, mutually accepted norms. This dualistic view is, according to Habermas's own interpretation, in line with the Aristotelian tradition of thought on the subject of action, taken up in the 20th century by Hannah Arendt and Hans-Georg Gadamer and before them expressed (and then abandoned) by the young Hegel of the Jena period and manifested in Marx's concrete analyses of capitalism, though not in his theoretical self-understanding.

2) Each of these two types of action is governed by its own specific type of rationality. *Techne* is governed by "technical" rationality, a rationality of hypothetical imperatives, to use Kant's term, informed by scientific and technological knowledge. The latter is largely conceived by Habermas on the lines of the hypothetico-deductive model of mainstream philosophy of science, emphasizing the logical relation between explanation and prediction, which makes possible the use of science for the technological control and manipulation of

objects on the basis of their own patterns/laws of behaviour.

Unlike modern positivism, however, Habermas does not hold the "scientistic" belief that scientific knowledge of this kind is identical with knowledge as such. In particular, he rejects the idea that there can be no rationality with regard to the determination of norms and values in the context of *praxis*, and that such determination is simply a matter of non-rational decision. In this respect Habermas is also at odds with Hume and Weber, as well as with various non-positivist yet relativist schools of thought, such as communitarianism, hermeneutics and post-structuralism.

Habermas's belief in the possibility of a rationality pertaining to *praxis*, here referred to as "praxial rationality" – a belief which he considers the cornerstone of any critical theory of society – is first expressed in his earliest major work *The Structural Transformation of the Public Sphere* (1989) in terms of the notion of "rational-critical debate". His early ideas on praxial rationality were elaborated in his defence of Adorno against Popper in the context of the "Positivist Dispute" of the 1960's (see Adorno et al, 1976), and were given their fullest expression in *Knowledge and Human Interests* (Habermas, 1972).

In the latter work Habermas maintains that in contrast to technical rationality, which is based on empirical-analytic knowledge, governed by the "quasi-transcendental" knowledge-constitutive interest in technical control, praxial rationality is based on a combination of hermeneutics, governed by the knowledge-constitutive interest in mutual understanding, and critical self-reflection. Critical self-reflection itself is a dynamic synthesis of empirical-analytic and hermeneutic elements that provides critical-emancipatory self-knowledge.

This form of praxial rationality differs from both its objectivist and Kantian

Enlightenment variants. Self-reflection – its most important aspect – is an idea of Hegelian origin that Habermas models on Marx's critique of ideology and, above all, on Freudian psychoanalysis. It aims to reveal and eliminate self-deception and all self-imposed, hypostatized powers, thus bringing about genuine self-understanding and freedom and creating the conditions for authentic mutual understanding.

3) In modernity, according to Habermas, there has been an increasing displacement of forms of *praxis* and praxial rationality by forms of *techne* and technical rationality, culminating, in the late 20th century, in technocracy – the reduction of ethical and other evaluative questions to technical ones.

Habermas rejects the idea which he claims Marcuse was toying with, namely that technical action and rationality could be historically overcome as a mode of relating to nature. However, he also disagrees on the one hand with the Weberian and post-1930's Frankfurt School view that the development of technical rationality necessarily leads to social domination and the destruction of praxial rationality, and, on the other hand, with the opposite view, namely that the development of technical rationality and the concomitant "liberation from hunger" inevitably leads to praxial rationality and the "liberation from servitude". He is convinced that whilst each form of practical rationality is possible and desirable in its own proper field of application, there is no inevitable link between them of either a positive or a negative kind.

The main criticisms advanced against these early Habermasian views can be grouped in two clusters, the first addressing issues relating to the dichotomy of *techne* and technical rationality vs *praxis* and praxial rationality, the latter concerned with the philosophical foundations of praxial – and to a lesser extent of technical – rationality, as these foundations have been developed in *Knowledge*

and Human Interests. I shall discuss these two clusters of criticism in turn.

1.3.1 Critical perspectives concerning the dichotomy between *techne* and *praxis*.

1.3.1.1 McCarthy's critique

In his major book on Habermas, covering the latter's writings until the mid-seventies, Thomas McCarthy discusses a number of ambiguities and confusions regarding Habermas's use of the distinction between labour and interaction and various other terminological versions of it (McCarthy, 1978: 16-40).⁹ Here I shall take up only some of the issues raised by McCarthy.

One of Habermas's main intentions behind the labour/interaction distinction has been to differentiate conceptually between two fundamental kinds of anthropological relationships and the very different and irreducible forms of rationalization to which each is amenable. These are the relationship between human beings and nature on the one hand and the relationship of human beings to each other in society on the other. The former relations concern the satisfaction of material needs and are in modernity facilitated above all by science and technology, as the most important elements in productive activity; the latter concern the establishment of forms of understanding, coexistence and cooperation between human beings that make social life possible. Labour, or purposive-rational action, in McCarthy's words "is governed by technical rules that imply conditional predictions, as well as preference rules and decision maxims that imply conditional imperatives; it is directed to the attainment of goals through the evaluation of alternative choices and the organization of appropriate means; it is sanctioned by success or failure in reality " (McCarthy, 1978: 26). Social interaction, on the other hand, "is governed by consensual norms that define

reciprocal expectations; it is grounded in the intersubjectivity of mutual understanding and secured by the general recognition of obligations; it is sanctioned conventionally" (ibid.).

McCarthy points out, justifiably, that this contrast implies, contrary to what Habermas is at pains to reject elsewhere, that the use of science and technology as well as all other activities involved in material production – "the activities of scientists and technicians, producers and planners" (ibid.) – are by definition not social, not in any way governed by social norms or interpersonal relations. Similarly, the distinction has the equally unacceptable implication that social interactions cannot be about activities of material production, for if a social activity were directly concerned with material production it could not be classified as purely interactive, or, for that matter, as purely purposive-rational, as demanded by the distinction.

Evidently, concrete spheres of action, or even single concrete actions, more often than not inextricably combine purposive-rational and social-interactive elements, and therefore the proposed dichotomy is incapable of sorting them out and clarifying their relationship in a coherent manner. McCarthy, following certain clarificatory cues from Habermas, suggests that there are two other ways in which the dichotomy can be interpreted and that these are in fact the ways Habermas understands and uses the dichotomy, though confusingly not specifying which of the two he is employing each time.

Firstly, the two concepts are applied to different aspects of actions rather than to whole concrete actions. Thus, the terms purposive-rational, technical, etc. "focus on task elements of actions or action systems... [thematizing] the means-ends orientation of the action, the technical knowledge and decision procedures on which it is based, its degree of success 'in reality', in short the economy and

efficiency with which means are employed to realize specified ends" (ibid.); whereas the terms social-interactive, communicative, praxial, etc, "focus on the intersubjective structure of action... [thematizing] the consensual norms, reciprocal expectations, mutual understanding..." (ibid.) In other words, the former set of terms refers to those elements of actions that concern the technical control and manipulation of an object by a subject who uses the object as a means to his or her own ends, whereas the latter set refers to those elements of actions that involve the norm-oriented, mutually respectful relations between subjects.

Secondly, the terms are applied by Habermas not to aspects of actions but to whole, concrete actions or systems of actions according to the aspect – the technical or the praxial – which is dominant in them.

This conceptual arrangement, however, raises certain questions of its own. To begin with, if the point of the dichotomy is to differentiate between actions that are predominantly *technical* (i.e., involving the control and manipulation of an object by a subject using the object as a means to its own ends) and actions that are predominantly *praxial* (i.e., involving mutually respectful intersubjective relations) the initial intention to use the dichotomy in order to differentiate between the sphere of humanity's transactions with nature in the form of activities of material production, on the one hand, and the sphere of interaction among human beings in social areas outside material production, on the other, cannot be sustained.

For it cannot be *a priori* assumed and conceptually guaranteed that social interactions in such spheres as politics, family, culture, ideology, etc, are and will always be predominantly of a praxial kind in the above sense. Indeed, as McCarthy points out, Habermas explicitly considers the concept of purposive-rational action to be applicable to social relations. In this regard, he repeatedly refers to purposive-rational action as action involving the technical control not only

of natural but also social processes. For example, when talking about the technical rules that govern instrumental action he makes it clear that "they imply conditional predictions about observable events, physical and social" (Habermas,1971: 92). Furthermore, it is a central thesis of Habermas as well as of all Critical Theory that social relations in capitalist modernity involve exploitation, oppression and reification, and are far from exhibiting intersubjective respectfulness.

Thus, the concept of technical or purposive-rational action cannot be employed, and is not employed by Habermas, to capture only activities concerned with the utilization of nature in processes of material production; it is also applied to objectifying relations between human beings. In this respect, Habermas's concept of *strategic* action is important. As McCarthy shows, Habermas, after some initial ambivalence as to his usage of the term, characterizes as strategic that sub-set of technical/instrumental actions for which the object of control is human beings and their behaviour. In this kind of action, in McCarthy's words,

an agent acts instrumentally towards others, [that is] he comprehends their behaviour in terms of observable regularities and therefore as effectively controllable. He is involved in a subject - "object" relationship in which the "object's" capacity for a moral relationship, "its" potential for communicative relations with the agent, recedes into the background (McCarthy,1978: 29).

But if Habermas recognizes, as he clearly does, that much of social interaction is strategic in this sense, why does he conceptually associate interaction – having dichotomously distinguished it from work – with communicative, intersubjective action? Perhaps the answer is that, with the "interaction" - side of the dichotomy Habermas is interested not so much in descriptively capturing what, as a matter of

fact, is communicative, as in normatively expressing what he believes *ought to be* communicative, and which in his view lies exclusively within the sphere of social interaction.

Moreover, the simple dichotomy between (a) technical actions, involving manipulative, subject-object relations in the process of dealing with nature for the purpose of material production, and (b) praxial actions, involving communicative relations among human beings engaged in social interaction, is problematical not only because there can be, and are, subject-object relations between human beings in any social sphere but also with regard to the association of processes of material production – processes of work – exclusively with technical, manipulative, subject-object relations.

To examine this association we must analytically differentiate between two distinct, though closely interwoven, aspects of the process of material production: on the one hand there is the strictly physical relationship that a working person/agent of production, in performing a productive task, has with the material objects involved in production; and on the other hand there are the relationships into which working persons enter among themselves in performing their productive tasks. The latter kind of relationships include those involved in the structure and operation of work organizations such as relations concerning the division and sharing of work, cooperation, the distribution of functional organizational authority, as well as the determination of what knowledge, skills and equipment need to be used for various tasks and made available to the relevant working people.

Regarding the former aspect – that of productive activity that concerns the strictly physical relationship to nature – Habermas, in sharp contrast to an important line of early Frankfurt School thinking, rejects outright as of mystical provenance any idea that it can be anything other than technical, i.e., one of control and

manipulation of an object by a subject. In this aspect of humanity's relation to nature Habermas includes the very structure of modern natural science and technology, which he considers as inherently oriented to possible technical control, ruling out the possibility of an alternative science and technology, which would not be oriented to the technical control of nature while at the same time being of service to humanity in its efforts to satisfy its material needs.

Of course even if one agreed with the idea that nature can never be a partner in a communicative relationship in the way another human being can (though this could be questioned with regard to higher animals), one could criticize Habermas for not making an effort to differentiate between qualitatively different kinds of subject-object relationship between human beings and nature. For it is one thing to relate to nature in a mechanistic way, as to something little more than abstract matter open to control and manipulation on the basis of laws of physics obeyed in the same way by all matter in the universe; and quite another thing to relate to it – still in a broadly subject-object mode – with ecological sensitivity, as to a unique, concrete biosphere, an integral ensemble of interrelated ecosystems, with its own means of self-regulation and self-reproduction, which we humans must leave so far as possible undisturbed, and when necessary protect, when making use of nature's resources for meeting our material needs. But this lack of internal differentiation within Habermas's concept of the subject-object relation to nature is an omission which – though in need of remedy – does not seriously damage the broader framework of his ideas.

Habermas, however, does not restrict technical, manipulative, controlling action to the strictly physical relationship between human beings and nature; he goes further, "surrendering" to such action also the social-interactive aspects of processes of production. To be sure, Habermas accepts Marx's distinction

between forces and relations of production and emphasizes the historical-evolutionary importance of the "dialectical" relationship between the two, regretting that Marx, in certain influential formulations of historical materialism, gave primacy to the former. And Habermas would place relations of production on the side of *praxis*, these relations including such "class" relationships as the ownership of premises, equipment, materials and final products of work, as well as payment agreements between employers and employees. Moreover, as we have seen, the early Habermas identifies as the core ideological distortion of post-liberal capitalism the technocratic tendency to treat genuinely "practical" (in the sense of praxial) issues such as these – which are amenable to rationalization through "public unrestricted discussion, free from domination" (Habermas, 1971: 118) – as technical ones, to be dealt with by means of some kind of manipulative "human engineering" (*ibid.*).

Nonetheless, Habermas tends to include, though often implicitly, under the term "forces of production", and thus to place on the side of the genuinely technical, i.e., instrumental and strategic, other social-interactive aspects of the process of production such as the (so called "technical" as opposed to "social/class") division of labour and the relationships of authority involved in the structure and operation of work organizations. This is evident, for instance, when he differentiates between areas where instrumental and strategic action serves (acceptable) rationalization, and when it serves (unacceptable) "manipulation":

Whether it is a matter of rationalizing the production of goods, management and administration, construction of machine tools, roads, or airplanes, or the manipulation of electoral, consumer, or leisure-time behaviour, the professional practice in question will always have to

assume the form of technical control of objectified processes" (ibid.: 55).

Sometimes he goes further – following Weber and anticipating the explicit and systematic position he takes later in *The Theory of Communicative Action* – by referring much more broadly to "the economic system" and "the state apparatus" as primarily "subsystems of purposive-rational action" (Habermas, 1971: 93-4).

Thus, just as with the concept of interaction (in the sense of communicative action) Habermas aims to capture not all existing social interaction but rather, normatively, all the areas of interaction he thinks ought to be communicative, so with the complementary concept of work (in the sense of technical-instrumental, manipulative, control-oriented action) he seems to want to capture not only the purely physical relations of human beings to nature but also those social interactions which, again normatively, he accepts as properly technical and locates exclusively in the sphere of work processes.

Here of course the question arises: on what grounds does Habermas distinguish between those social-interactive aspects of production processes that should rightly be governed by technical/purposive-rational/strategic action and those that ought to be governed by praxial/communicative action? Or, putting the question somewhat differently, according to what criterion does he draw the line between the social-interactive aspects of production that are to be considered as forces of production and those that are to be considered as relations of production?

One possible, and eminently Habermasian answer (see Habermas, 1972: 280–1) would be that certain kinds of social interactions can be allowed to be purposive-rational because the agents involved do, or would, as a result of adequate processes of critical self-reflection, come to consider such purposive-rational

interactions as acceptable because they are inextricable aspects of work processes that are productive enough to satisfy their material needs and therefore unavoidable if that level of satisfaction of material needs is to be secured; whereas other kinds of social interactions are, or would be, considered acceptable to agents as purposive-rational only in the absence of self-reflection, that is, as a result of reification.

In that case Habermas would then have to provide empirical evidence or counterfactual arguments to establish that his placing of social-interactive areas of production processes, such as management, administration, the technical division of labour and the structure of work organizations, on the side of purposive-rational action satisfies the above criterion of adequate critical self-reflection; which, however, he does not. In fact, the question whether, on the basis of the above criterion, this or that area of social activity should rightly be a domain purposive-rational, rather than praxial/communicative, action, could not be answered adequately before systematic processes of critical self-reflection, on the part of the agents concerned, took place; and even then, the verdict would be fallible, subject to revision by further processes of critical self-reflection. In the absence of adequate self-reflection of this kind, it is best for Critical Theorists to leave these questions open.

But there is another possible explanation, of a conceptual kind, as to why Habermas divides social interactions into *techne* and *praxis* the way he does, and why he does so in a seemingly *a priori* manner, without apparently feeling the need to provide evidence in support of his decisions in this regard.

In the Habermasian conceptual framework, "work", or *techne*, or purposive rational action (including both instrumental and strategic action) is goal-oriented action, that is, action involving the employment of means to achieve ends, and, by

definition, also action involving the control and manipulation by a subject of something treated as an object. In the case of strategic action, human beings, or human behaviour, are at once the means to the subject's ends and the object of control and manipulation. On the other hand, "interaction", or *praxis*, or communicative action is, again by definition, action involving a relationship between mutually respecting subjects, and, at the same time, action oriented to norms, not to goals.

According to this conceptual architecture, therefore, actions which are predominantly involved in a direct way with the pursuit of goals – as in the case of pursuing the satisfaction of material needs through activities of production – cannot, by definition, be other than technical, in the sense of controlling and manipulating. Conversely, actions, by definition, can be praxial only if they are not primarily oriented to goals, or perhaps so oriented only indirectly.

To be sure, even on the basis of this strategy, decisions to put a certain activity on one side of the divide or the other would need a lot of explanation, much of it turning on the meaning of terms such as "predominantly", "directly" and "indirectly". More importantly, however, the very dichotomous nature of the work/interaction conceptual framework is seriously flawed, for, apparently, it rules out the possibility of human relationships that are at once both goal-oriented and communicative. This conceptual flaw reappears in Habermas's later, communicative paradigm, as will be seen in Part Two, and has very important implications for his social theory. In fact, Part Three of this study is largely devoted to clarifying this flaw and its implications, and to working out adequate remedies.

In the light of the preceding discussion I would like to refer briefly to certain criticisms of the Habermasian conceptual distinction between labour and interaction advanced by Axel Honneth and Anthony Giddens.

1.3.1.2 Honneth's critique

Axel Honneth (1982) focuses his critique on the concepts of work and instrumental action. He begins with Marx who, according to Honneth, developed a complex concept of work in which alienated and non-alienated aspects remained in tension with each other. This concept was based on the experience of early industrialization when both "meaningful, handicraft-oriented" and "atomized industrial" (Honneth, 1982: 23) forms of work coexisted. Marx's concept covered on the one hand meaningless, "abstract" acts of labour whose object is out of the control of the worker and on the other a "meaningful and self-regulating form of activity intimately related to its object", or "acts of artisanry which are complete in themselves" (ibid.: 25). Emphasizing the latter aspects, Marx – in the footsteps of Hegel – conceives of work as the core developmental process of the human species, a process that combines within itself technical-productive as well as moral and political elements; so much so that "world history is defined [by Marx] as the self-recreation, self-preservation and self-emanicipation of society through work" (ibid.: 22).

This rich and critical 19th century concept of work has been impoverished in 20th century social theory, Honneth maintains, in the wake of the extreme impoverishment of real work in industrial capitalism. Hannah Arendt (1958), among others, redefined work as an activity that shared a great deal with partial, repetitive, mindless, mechanized labour processes that came to prevail in the large-scale industrial plants of late 19th and early 20th centuries, and were then radicalized and systematized in Taylorism.¹⁰ As a complement to this "lowly" concept of work Arendt reintroduced the Aristotelian concept of *praxis*, to cater for political, cultural and other activities that are "free from all contact with things" and constitute "true action" (Honneth, 1982: 27).

For Honneth, 20th century Marxist theory too adopted this theoretical approach of "categorically removing any sort of potential emancipatory significance from the act of work" (*ibid.*: 28). Thus, theorists such as Lukács, the early Marcuse and Sartre assign emancipatory potentiality not to the concrete work experience of the worker but to the "*praxis* of a transcendental or collective working subject" (*ibid.*), whereas Horkheimer and Adorno, in the *Dialectic of Enlightenment*, sever even this tenuous link between work and emancipation, arguing instead for the existence of an intrinsic association between work and domination, and resting their hopes for human emancipation on "the aesthetic idea of a mimetic mode dealing with nature" (*ibid.*: 30).

Precisely this trend of 20th century social theory, Honneth points out, is followed by Habermas himself. Thus, Habermas adopts "too thin" a concept of work, one that is limited to "instrumental action which can be applied without distinction to any manipulative relation to an object" (*ibid.*: 36). Honneth charges Habermas with constructing a flat, uncritical concept of work that prevents him from differentiating between work processes that handle objects in significantly different ways:

A critical concept of work must grasp categorically the difference between an instrumental act, in which the working subject structures and regulates his own activity on his own initiative, according to his own knowledge, in a self-contained process, and an instrumental act, in which neither the accompanying controls nor the object-related structuring of the activity is left to the initiative of the working subject (*ibid.*).

While both the above kinds of work are forms of instrumental activity, Honneth

argues, the former possesses potential for resistance, conflict and emancipation, whereas the latter does not – a fact that Habermas, unlike Marx, does not, and conceptually cannot, recognize.

Thus, unnecessarily condemning work to the status of morally neutral, manipulative, instrumental action, Habermas, (Honneth concludes), in a manner similar to Arendt, explores potentialities for moral development and emancipatory action in terms of the concept of communicative interaction among human subjects, which he applies to areas of social activity that lie beyond the sphere of work.

I have already pointed out that Habermas's concept of technical action, or work, is too thin to differentiate between mechanistic and ecological ways of relating to nature. Now, Honneth argues that the concept is too thin to make a different internal distinction, that between Tayloristic and craft-like work processes, while the complementary concept of *praxis* cannot help in this respect because it is too far removed from the world of work, and – having been modelled on Arendt's concept – may in fact have "nothing to do with things" whatsoever.

It is true that Habermas does not differentiate between Tayloristic and craft-like work processes (i.e. forms of technical division and organization of work). At the same time, under the concept of work he tends to assimilate the strictly physical relationship to the object of work with the other aspects of the work process. But this problem could not be remedied by means of an internal differentiation within a concept of work understood as referring to an instrumental, manipulative relation to nature, as if different kinds of technical division and organization of work constituted merely different ways of relating to the physical objects of production. As argued above, such aspects of the production process are predominantly social-interactive. They should therefore be subsumed under some concept

different from work in the above physical sense, a concept of interaction which, in addition to recognizing their social nature and differentiating them from the strictly physical aspects of the work process, would preferably have the critical edge to internally differentiate between those social interactions that can be accepted as strategic and those that ought to be communicative, in Habermas's sense of these terms. That way, Honneth's critical concerns about the distinction between Tayloristic and craft-like work could be addressed in a more coherent and rigorous manner.

Nor would it be satisfactory to tackle Honneth's objection by assimilating the physical and the social-interactive aspects of work processes under an expanded concept of "work" or "*praxis*" in a manner akin to that of the young Marx, and then differentiating among work processes according to how Tayloristic or how craft-like they are, thus conceding that whatever kind a work process is involved, it is an inextricable bundle of physical and social elements. Such a conceptual strategy would, I think, be a step backwards from Habermas's distinction between communicative and purposive-rational action, the latter further differentiated into instrumental and strategic variants.

To be sure, such a move would help us regain some of the critical acumen lost by Habermas – a loss evident in the relative ease with which he accepts the existing forms of work organization, management and administration, while knowing that they are pervaded by controlling, manipulative, strategic action. And, even though this gain would have been achieved at the expense of analytical rigour, the trade-off could be acceptable if we could not have both the analytical rigour and the critical perceptiveness. However, as I will argue in Part Three, we can have both if we take forward steps to develop the Habermasian concepts further; in particular, if we move in the direction of uncoupling the ends-means orientation from the

manipulative aspect of "purposive-rational action", and the norm orientation from the intersubjective communication aspect of "communicative action", and making it conceptually possible to have social action which is at once goal (ends/means)-oriented and communicative.

1.3.1.3 Giddens's critique

In his contribution to the volume *Habermas: Critical debates* (1982: 149-61), Anthony Giddens, while noting the enduring importance of the labour/interaction distinction in Habermas's writings, as well as some positive effects the distinction has had within the framework of Habermas's work, such as the "identification of positivistic strains in Marx and the connecting of these to an analysis of the limitation of technical reason" (*ibid.*: 155), nonetheless focuses on certain features of the distinction which, in his view, render it ultimately inadequate. I shall concentrate on Giddens's objections concerning the notion of interaction, the objections that I think would be the most damaging if well founded, but which, on examination, proved not to be.¹¹

Giddens summarizes Habermas's distinction between labour and interaction as one between purposive-rational action, which is "governed by technical rules, and sanctioned by the likelihood of failure to reach objectives" (*ibid.*: 156) and communicative action, which is "governed by social norms, and sanctioned by convention or law" (*ibid.*). A serious problem with this conceptual scheme, according to Giddens, is the identification of social interaction with communicative action and of both with action oriented to norms, making norms the central component of social interactions, just as in Parsons' "normative functionalism". But "there is more to interaction than the norms to which it is oriented", maintains Giddens (*ibid.*: 158), arguing that "power is as integral a component of all social

interaction as norms are" (ibid.: 159). Giddens charges Habermas with conceptual blindness to crucially important power-related social phenomena such as disparities with regard to access to scarce resources and struggles over sectional group interests, a blindness that for Giddens amounts to an "absent core" in Habermas's conceptual scheme, one that makes it unfit "for grasping the production and reproduction of society" (ibid.).

This communicative and normative understanding of social interaction, Giddens claims, has serious consequences for the scope of Habermas's critique of modern society, not allowing it to extend beyond the moral, cognitive and ideological orders, to the material basis of social domination. For Habermas, according to Giddens, "the critique of domination comes to turn upon freedom of communication or dialogue, rather than upon material transformations of power relations" (ibid.).

It would be fair to say that before *The Theory of Communicative Action* Habermas had not – and did not claim to have – worked out a comprehensive set of social-theoretical concepts capable of accounting for "the production and reproduction of society"; Giddens is right when he says that "Habermas has little to say about, and proffers little in the way of concepts for analyzing, the social relations that are constitutive of social systems" (ibid.: 158). Moreover, even *The Theory of Communicative Action* does not, of course, amount to a substantive sociology of modern societies.

But irrespective of the comprehensiveness or otherwise of his social theory and the extent of his substantive sociological analyses, the claim that the conceptual scheme of labour and interaction makes it impossible for Habermas to grasp phenomena of social power and struggle and to criticize social domination other than in the form of communicative and ideological distortions is, I believe, the

result of a gross misunderstanding.

For as we have seen above, the very phenomena of power stressed by Giddens can be adequately captured in terms of Habermas's concept of purposive-rational action, which is clearly intended to be applicable to social as well as to physical reality. More specifically, such phenomena come under the variant of purposive-rational action Habermas calls *strategic*, as distinct from the *instrumental* variant, which refers to non-social purposive-rational actions.

As might be expected, Habermas, in his "A reply to my Critics" (Thomson and Held, 1982: 219-83), responds to Giddens by directing attention precisely to the concept of strategic action.

The concept of force already has a central place in the action theory sketched above: 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 and strategic action says nothing else than this (*ibid.*: 269).

But although the blindspot Giddens claims to have identified in Habermas's work seems not to be there, there is, I will argue, another gap in Habermas's work which does have its origin in a conceptual problem; the same problem which I pointed out above, in the course of my discussion of McCarthy's and Honneth's critiques. With reference to the above quotation of Habermas, this problem can now be approached from the slightly different angle of action coordination.

Habermas's conceptual scheme seems to allow for two kinds of action coordination between social agents: The first kind is coordination by means of "exercising force", or more precisely, and generally, by means of strategic, i.e. controlling, manipulative, goal-directed action by some human beings towards others whom they want to use as means to their own ends. The second kind of social coordination is through "achieving understanding", which Habermas understands as communicative, norm-directed action between mutually respecting subjects. This dichotomy categorially rules out the possibility of action coordination between goal-directed social agents who want to use other social agents as means to achieving their goals, but are intent on doing so not in a strategic, controlling, manipulative manner but communicatively, through reaching understanding. At first sight, this concept of a third way of coordinating social action seems to have considerable descriptive applicability as well as enormous prescriptive attractiveness. If this is so, its absence from Habermas's conceptual apparatus constitutes a theoretical lacuna that creates a major blindspot for his Critical Theory. As already mentioned, this point will be central to my line of argument in Part Three.

1.3.2 Critical perspectives on Habermas's ideas on the rationality of *techne* and *praxis*

The question concerning the nature, and indeed the very possibility, of praxial rationality is central for Habermas, as it is for all Critical Social Theory, that is, for all social inquiry that, in addition to comprehending social reality, wants to pass reasoned evaluative judgement on it, while seeing itself reflexively as part of that very same societal process.

The question of the possibility and structure of technical rationality, however, as opposed to that of technical rationality's scope of applicability and ideological social role, receives in Habermas's work much less attention. This is largely because on the issue of the structure of technical reason Habermas is in considerable agreement with the thinking of mainstream philosophy of science, while the possibility of technical reason is not, and can hardly be, questioned in the age of modern science-based technology.

Nonetheless, Habermas's ideas on the fundamental conditions of possibility of technical rationality are sufficiently distinct to justify some critical discussion. I shall begin with such a discussion, and then turn to a critical consideration of Habermas's views on the rationality of *praxis*.

1.3.2.1 Critical perspectives on technical rationality

Technical rationality, for Habermas, is concerned with the development and employment, through the use of reason, of technical means to given ends. The term "technical means" here refers to means that involve controlling and manipulating objects or objectified processes.

The technical capacity of reason is measured according to the effectiveness – in terms of reaching desired goals – of the technical means devised and employed through its use. Of course, often technical means can serve goals other than those for the sake of which the means were devised, even goals that are chosen later and in the light of the fact that the means to their satisfaction have come into existence. Telephone networks, for example, were not initially developed in order to be used as the basis for the internet.

The degree of "rationalization" of *techne*, or technical action, at any given time, can be defined as the extent to which technical means are developed and employed through the use of rationality. Traditional *techne*, largely in the form of craft, relied more on intuitive knowledge embodied in skills which were passed from craftsman to craftsman across generations, than on systematic knowledge arrived at through deliberate rational processes. The degree of rationalization of traditional *techne* was thus low, and at the same time its progress, in terms of increasing effectiveness of technical means, was limited. In modernity, however, the effectiveness of technical means has increased to an unprecedented degree, and continues to increase with an accelerating pace. This development has been achieved through growing rationalization, effected by means of the fusion of *techne* with modern scientific knowledge, in the form of science-based technology.

Thus, at the heart of the impressive advancement – through rationalization – of technical action in modern times, especially since the "second industrial revolution" in the late 19th century, lies the technical exploitability of modern science, that is the capacity of modern scientific knowledge to be utilized for technical purposes. Nomological science makes possible precise predictions about the effects of specified interventions into specified objectified processes. Thus, it enables us to technically bring about the predicted effects through making the specified intervention into the specified objectified process.

This much, concerning the technical exploitability of modern science, is by and large shared by mainstream philosophy of science. The early Habermas, however, is not content with taking the technical exploitability of science for granted, as an unproblematic incidental property of it; an attitude that might be understandable from the point of view of a realist interpretation of science: since science grasps the true nature of the reality it studies, it is bound to give clues as to how that

reality is to be controlled and manipulated. Nor is Habermas content with considering technical exploitability as a property of all forms of knowledge, as, in his view, positivism does, treating natural science as the paradigm of all science, and scientific knowledge as the paradigm of all knowledge.

Unlike positivism, Habermas inquires, in a Kantian-critical manner, into the epistemological conditions of possibility of scientific knowledge, and provides an answer in terms of the theory of *cognitive interests*. In contrast to the hermeneutic and critical sciences that are governed by the "practical" and "emancipatory" cognitive interests, respectively, the nomological, empirical-analytic sciences are said to be governed by the "technical" cognitive interest, the interest in developing knowledge that can be utilized for technical purposes. Thus, in a way reminiscent of Heidegger's "The Question Concerning Technology" (in Krell, 1993), Habermas considers technical exploitability not as an incidental property, but as a kind of prior "intention" of scientific activity.

This attempt by Habermas to provide epistemological underpinnings for empirical-analytic science, and thus for modern technical rationality, can be criticized on two main counts.

1. The first criticism concerns the nature of the cognitive interests. Specifically with regard to the technical interest Habermas has been criticized for making two claims which are incompatible with each other.

On the one hand, Habermas holds that the technical cognitive interest is the "transcendental" condition of the possibility of scientific knowledge of natural or other objective or objectified processes, and it has also played this role, though in a much less systematic way, with regard to prescientific technical/craft knowledge. But the technical cognitive interest is not "transcendental" in the strictly Kantian

sense of structuring the individual consciousness of the knowing subject. It rather structures the methods of inquiry of the human species – or, in the age of modern science, it structures, more precisely, the methods of inquiry followed by the community of scientists. In this way the structure of scientific inquiry is linked to the structure of technical action itself: The hypothetico-deductive logic of empirically testing scientific hypotheses is analogous to the logic of technical action guided by scientific theories, that is, by hypotheses that have withstood empirical testing (see section 1.2.3, above).

Thus, according to this view, the technical exploitability of science is a property inherent in the very structure of scientific method and scientific theory and not their contingent after-effect, whilst nature, as well as other objectified processes, is constituted by the technically acting and knowing subject as an object of possible technical control.

But on the other hand, Habermas insists that the human species and its technical cognitive interest are themselves products of nature, albeit making it clear that he does not want this latter position to be interpreted in a pre-critical, metaphysical-materialist or naturalistic manner.

The problem with this combination of views is succinctly described by Thomas McCarthy.

Habermas appears to be caught in a dilemma: either nature has the transcendental status of a constituted objectivity and cannot, therefore, be the ground of the constituting subject; or nature is the ground of subjectivity and cannot, therefore, be simply a constituted objectivity. Habermas wants, paradoxically, to hold on to both horns (1978: 111).

In a similar vein, Henning Ottmann maintains that "Habermas's conception of the technical interest finally sways between naturalism and idealism" (Thomson and Held, 1982: 92).

McCarthy (1978: 111-125) surveys Habermas's efforts to resolve the dilemma via a "reconstruction of Marx's implicit epistemology" (*ibid.*: 114), but concludes that these efforts are, in the end, unsatisfactory.

No matter what form it takes, nature-in-itself remains, within a transcendental framework, "an abstraction required by our thought". That such an abstraction should simultaneously be the ground of thought is incomprehensible (*ibid.*: 125).

Habermas described the cognitive interests as *quasi-transcendental*, in an effort to avoid over-privileging, terminologically, one of the two aspects of his intended concept. He was, however, himself fully aware of the serious tension between these two aspects, and in the end knew that he had failed to resolve it. In a review of his early work, referring to objections concerning "the inadequately clarified status of the interests that direct knowledge", he admits:

the formula "quasi-transcendental" is a product of an embarrassment which points to more problems than it solves (Habermas, 1971: 14).

2. Habermas's theory of cognitive interests can also be questioned with regard to the possible implications of its transcendentalist aspect concerning such issues as the nature of scientific knowledge and the ways in which human beings can relate to the natural world.

In particular, an objection to the idea that scientific knowledge is governed by a technical interest is one that – as McCarthy points out – could have come from the

first generation Frankfurt theorists themselves, mainly from Horkheimer, Adorno and Marcuse. The objection is that

it appears that this conception limits our knowledge of nature to information that is technically utilizable and our intercourse with nature to instrumental mastery of objectified processes (McCarthy, 1978: 66).

It is certainly true, as we have seen in an earlier section, that Habermas rejects the possibility of a qualitatively different, non-manipulative form of productive, physical need-satisfying relation to nature; and of a "mode of *cognizing* nature" (*ibid.*: 67) other than that offered by the technically governed empirical-analytic sciences. However, he does – as McCarthy suggests – allow for other, non-technical *attitudes* towards nature, such as the "mimetic, poetic, playful, mystical, fraternal" (*ibid.*) attitudes. Such attitudes, though, must by implication be understood as modes of consciousness that are empty of any cognitive content; and such a thesis is highly problematic.

Thus the theory of knowledge-constitutive cognitive interests proves unsatisfactory, and Habermas recognizes this fact. In the 1970s he abandons the theory, together with the whole paradigm of his early work, in favour of an alternative, communicative paradigm, which will be discussed in Part Two of this study.

However, it should not be concluded that because the theory of cognitive interests – which was intended to provide an explanation for the technical utilizability of science – is inadequate, therefore Habermas's views concerning the possibility and structure of technical rationality are invalid. For these views are premised only on the fact of the technical utilizability (including the technical effectiveness) of science; and that fact cannot be denied just because of a failure to explain it.

Indeed, Habermas's early conception of technical reason and technical rationalization, in terms of the idea of the technical utilizability of science, seems to have survived intact his communicative paradigm shift.

A potentially more damaging criticism of Habermas's conception of technical rationality, both pre-communicative and communicative, is closely related to the one raised at the end of our discussion of McCarthy's, Honneth's and Giddens's critiques of the work/interaction dichotomy. It concerns the identification of technical rationality, in the sense of rationality involving the control and manipulation of objectified processes, with means-ends rationality; an identification that conceptually rules out the possibility of means-ends rationality that is not technical in the above sense. This issue, as already mentioned, will be taken up and extensively discussed in Part Three.

1.3.2.2 Critical perspectives on praxial rationality

Critical, including self-critical, discussion on Habermas's ideas concerning the rationality of *praxis* centres, by and large, on a certain concept of self-reflection.

Replacing – perhaps also to some extent explicating – the early notion of "rational-critical debate" which represented Habermas's understanding of praxial rationality in *The Structural Transformation of the Public Sphere* (1989), this concept of self-reflection was first introduced in the context of the "positivist dispute" (see Adorno et al, 1976) and in essays contained in *Theory and Practice* (Habermas, 1974). In these works the terms mainly used for the self-reflective approach were "dialectics" and "critique". The concept was later fully worked out in *Knowledge and Human Interests* (Habermas, 1972), with the predominance of terms such as "reflection", "self-reflection" and "critical self-reflection". I am also referring to it as critical-emancipatory self-reflection.

The concept of self-reflection, which, as we have seen, is of Hegelian provenance and is modelled on Freudian psychoanalysis and to a lesser extent on Marx's critique of ideology, is proposed by the early Habermas as the most developed, most adequate form of praxial rationality.

It is a rational process in which the subject is supposed to gain self-understanding and insight into self-deceptions, and through it emancipation from self-imposed, hypostatized powers. Self-reflection of this type is proposed by Habermas as a form of praxial rationality associated with the emancipatory cognitive interest, just as analytic-nomological knowledge is associated with the technical cognitive interest and hermeneutic knowledge, itself a weaker form of praxial rationality, is associated with the practical cognitive interest.

In what follows, I shall discuss the main criticisms advanced against this Habermasian concept of self-reflection and Habermas's response to them. More precisely, I shall discuss, firstly, some criticisms coming from the side of hermeneutics; secondly, certain difficulties regarding the psychoanalytic analogue of self-reflection (the "therapist-patient" relationship, and the relationship between enlightenment and emancipation); thirdly, the status of self-reflection as praxial rationality; and fourthly, confusions concerning two different senses of critical reflection.

As we shall see, some of these criticisms have given Habermas an opportunity to further clarify and defend his ideas. Some others, however, have led him to a more radical questioning of his earlier views, and eventually to the paradigm shift to his later, communicative theory.

1) *Critique coming from the side of Gadamerian hermeneutics*

In his analysis of Gadamer's *Truth and Method* (1975), Habermas praises Gadamerian hermeneutics for marking a significant advance on both the 19th century "romantic" hermeneutics of Schleiermacher and Dilthey and Wittgensteinian language-games theory, themselves an advance, according to Habermas, on the positivist conception of social science.

To begin with he is in agreement with Gadamer's rejection of the "objectivistic" conception of interpretation, that is, of the idea that interpretation is unaffected by the interpreter's perspective. For 19th century hermeneutics, this conception is represented above all by the idea that successful interpretation requires, on the part of interpreter, an "empathy" with the author, i.e. psychologically putting oneself in the author's position; this alone would enable the interpreter to achieve understanding of the author's true meaning. For Wittgensteinianism, the idea of empathy is rendered in linguistic rather than psychological terms: to achieve understanding the interpreter must learn, and step into the author's language game.

Gadamer, by contrast, stresses the subjective and historical dimension of understanding: the ineradicable role, in every instance of understanding, of the interpreter's own language, tradition, "prejudices", historical point of view or "horizon". Gadamer's hermeneutics "insists that we learn to understand a language game from within the horizon of the language that is already familiar to us" (Habermas, 1988: 152–3). With regard to historical interpretation, the hermeneutic idea, which Habermas approves, is that "what [the historian] can know historically cannot be grasped independently of the framework of his own life–*praxis*" (*ibid.*: 160).



But if understanding is not an objective representation of the author's authentic meaning, it is not an arbitrary, subjective construction either. For Gadamer it is rather a kind of unity of subject and object, one resulting from the "fusion" of the author's linguistic-historical horizon with that of the interpreter. This fusion is the outcome of a dialogical process and involves the mutual adjustment of the two horizons. From the point of view of the interpreter – ordinary communicating social agent, historian or social scientist – understanding meaning involves self-reflection, self-understanding and self-change. In particular it has an effect on the interpreter's future actions. In this respect Gadamerian hermeneutics generalize, according to Habermas, from "the immanent connection between understanding and application [that] can be seen in the cases of theology and jurisprudence" (*ibid.*: 162), where one seeks to gain understanding of sacred texts or laws, respectively, with a view to applying it in his or her own situation. Indeed, Gadamer compares hermeneutic understanding with Aristotelian practical knowledge (*ibid.*: 153-4).

Habermas wholeheartedly endorses the re-establishment by Gadamer of the intrinsic linkage between understanding and *praxis*, fact and value, science and the ethico-political realm, which had been severed both by positivism and by other forms of hermeneutics:

I see Gadamer's real accomplishment as his demonstration that hermeneutic understanding is necessarily related, on the transcendental level, to the articulation of an action-orienting self-understanding.
(Habermas, 1988: 162)

At the same time, however, Habermas strongly objects to what he sees as Gadamer's profound conservatism. He rejects the view, which Gadamer seemingly

maintains, that the prejudices, or prejudgements, inherent in tradition, thus in language, are largely uncriticizable and ineradicable aspects of all understanding. For Habermas, the action-orienting aspect of Gadamerian hermeneutics is nowhere near as decisive as he would wish. It is limited to the application of traditional meanings in new conditions, ruling out their radical questioning and criticism. Indeed, with every situation of hermeneutic understanding, with its inseparable applicative moment, traditional prejudices, and the authority of tradition as such, are largely reaffirmed.

A major problem with Gadamer, Habermas argues, is that he is content with remaining within the bounds of tradition and language, assuming that all possible meaning lies within these bounds and that all genuine understanding takes place within them. He fails to see that language and tradition can be – and are – carriers of systematic, not just occasional, distortions of meaning, akin to the unconscious and repressed motives dealt with by psychoanalysis: "Language is also a medium of domination and social power" (*ibid.*: 172). And if this is so, understanding cannot be simply assumed to be veridical. Gadamer, in Habermas's view, also fails to see that there are significant aspects of social and historical reality that lie beyond tradition and language and cause systematic distortions in tradition and language, which should also be addressed in any effort at social and historical understanding.

Furthermore, Gadamer is too preoccupied with the "ontological" issue of showing – in opposition to objectivism – that in all understanding the interpreter's perspective, on the one hand, and tradition and language (the alleged repository of all meaning, i.e. of all possible objects of interpretation) on the other, are bound up with each other; and he is too unconcerned with questions of method. As a result he does

not search for methodological means of creating a distance between the interpreter and the social reality that is to be interpreted.

All in all, Gadamerian hermeneutics, according to Habermas, leaves us with no reference point from which traditional prejudices and authority could be critically questioned.

Habermas's own position is that, in addition to tradition and language, social reality is also constituted by two kinds of "real" constraints: (a) those of external nature, reflected on the social plane in labour processes involving the technological exploitation of nature, and (b) those of inner human nature, socially reflected in repressive power relationships: "The objective context in terms of which alone social actions can be understood is constituted conjointly by language, labour, and domination" (*ibid.*: 174).

These two types of constraint act "behind the back of language" (*ibid.*: 174) to influence language itself and thus the very processes of hermeneutic understanding. Thus, true understanding of social reality, one that also makes possible the critical questioning of the prejudgements contained in tradition and language, can be obtained only if we also look beyond tradition and language, into the "real" world of our (exploitative) relations to nature and (oppressive) relations to each other.

Methodologically, Habermas proposes that genuine comprehension of this complex of language, labour and domination can be achieved through a type of self-reflection akin to that of Freudian psychoanalysis and Marx's critique of ideology, one that is achieved not by hermeneutic methods alone but by "joining empirical-analytic methods and hermeneutic ones" (*ibid.*: 167). Such a combination of causal and interpretative methods makes possible a critical and

emancipatory form of self-reflection, one that "understands the genesis of the tradition from which it proceeds and to which it returns [and thereby shakes] the dogmatism of life-*praxis*". (*ibid.*: 168)

In his response to Habermas, Gadamer rejects the latter's charge of conservatism. He denies that hermeneutics considers all authority and tradition as good and acceptable, and that it is incapable of subjecting it to any kind of criticism: "It is an inadmissible imputation to hold that I somehow meant there is no decline of authority or no emancipating criticism of authority" (Gadamer, 1967: 34). However, one cannot *a priori* reject all authority either: "Authority is not always wrong" (*ibid.*: 33). Hermeneutics brings aspects of authority and tradition to awareness and renders them open to conscious acceptance or rejection. Authority accepted on the basis of hermeneutic understanding, far from being dogmatic and arbitrary, is true, legitimate authority: "Authority can rule only because it is freely recognized and accepted. The obedience that belongs to true authority is neither blind nor slavish" (*ibid.*: 34).

To be sure, the interpreter, at any given time, can become aware of, and consciously assess, only a small fraction of his or her tradition and the prejudices inherent in it. The rest operates "behind his or her back":

Reflection on a given preunderstanding brings before me something that otherwise happens *behind my back*. Something – but not everything, for what I have called the [consciousness of effective history] is inescapably more *being* than consciousness, and being is never fully manifest (*ibid.*: 38).

But if hermeneutics does not conservatively lead to the blind acceptance of all authority, Habermasian reflection – Gadamer retorts – has the dogmatic tendency to reject all authority and traditional prejudgements, a tendency allegedly of anarchistic provenance:

The basically emancipatory consciousness must have in mind the dissolution of all authority, all obedience. This means that unconsciously the ultimate guiding image of emancipatory reflection in the social sciences must be an anarchistic utopia (*ibid.*: 42).

This excessively critical stand towards authority and tradition Gadamer associates with Enlightenment's "abstract antithesis" between reason and authority, a view that allegedly ignores the "true dependencies" between the two:

[T]he thing that hermeneutics teaches us is to see through the dogmatism of asserting an opposition and separation between the ongoing, natural "tradition" and the reflective appropriation of it (*ibid.*: 28).

Reason, for Gadamer, is an aspect of tradition, not something outside and opposed to it.

Linked to this argument is Gadamer's criticism of Habermas's thesis that labour and political power are "real" factors that lie beyond and operate "behind the back" of language and tradition, thus being outside the scope of hermeneutics but nonetheless accessible to reason. For Gadamer "it is absolutely absurd to regard the concrete factors of work and politics as outside the scope of hermeneutics" (*ibid.*: 31). Hermeneutics is concerned with "everything that can be understood" and "being that can be understood is language" (*ibid.*). But work, politics and all

other aspects of society must take a linguistic form – thus place themselves within the hermeneutically accessible realm of language and tradition – before they can present themselves to us as parts of social reality and become objects of reflection:

there is no societal reality, with all its concrete forces, that does not bring itself to representation in a consciousness that is linguistically articulated. Reality does not happen "behind the back" of language; ... reality happens precisely *within* language (*ibid.*: 35).

Of course, Gadamer explains, this is not the same as the idealistic thesis that "linguistically articulated" consciousness determines material reality.

Habermas should not have a problem with the view that all social reality, including social labour and power relations, cannot but present itself to us in a linguistic form. On the contrary, he virtually states as much himself:

There is good reason to conceive language as a kind of metainstitution on which all social institutions depend. For social action is constituted only in ordinary-language communication (Habermas, 1988:172).

For both philosophers social reality is a potential object of interpretative understanding. Also for both, traditional authority and prejudgements, once understood, can be consciously accepted or rejected, though in this respect there is a considerable difference in emphasis between the two thinkers. For surely, Habermas – contrary to Gadamer's unjustified claim quoted above – does not advocate or imply an "anarchistic" rejection of all tradition, despite emphasizing

the critique and overthrowing of traditional authority and prejudgements, unlike Gadamer who stresses their confirmation.

The fundamental difference between them lies in the methodologies which they propose for the successful understanding of social reality. For Gadamer, hermeneutic reflection, in the form of dialogical, communicative exchanges in the medium of ordinary language, is proper and sufficient for genuine understanding. Dialogical hermeneutic reflection is considered capable of seeing through the "dogmatic power [that] authority exercises ... in innumerable forms of domination" (Gadamer, 1976: 33) and enabling us to judge between acceptable and unacceptable authority, justified and unjustified prejudgements. Even unconscious motives, in the psychoanalytic sense, are claimed by Gadamer to "[fall] within the larger perimeter of hermeneutics" (*ibid.*: 41).

Habermas denies precisely the capacity of hermeneutic reflection to create the methodological "distancing" necessary for seeing through distorted communication, above all the kind of systematic distortions in communication – the "internal foreign territory" (Habermas, 1972: 218) – produced by social forces akin to psychoanalytic unconscious and repressed motives.

To overcome this weakness, Habermas proposes a combination of hermeneutic reflection with empirical-analytic, nomological methods that would make possible an emancipatory, critical reflection, one that would give access to meanings that would remain inaccessible to hermeneutic dialogue. Empirical-analytic methods would yield hypothetical causal explanations of alleged systematic distortions, and ultimately might lead to their dissolution. The capacity of self-reflection to dissolve a causal connection is of course due to the fact that we are dealing here with a "causality of fate, not of nature" (Habermas, 1972: 256), that is a causality determined not by external "natural" factors, but by factors internal to the mind.

To be sure, this emancipating effect occurs only if those allegedly under the spell of the presumed distortion ultimately come to accept – through self-reflection taking place in the medium of ordinary language discourse – the validity of the relevant causal hypotheses: "Only the patient's recollection decides the accuracy of the construction" (*ibid.*: 230). Thus, Habermas, in agreement with Gadamer and in opposition to positivism, in the final analysis returns to the medium of hermeneutic dialogue as the ultimate locus of judgement concerning genuine understanding. But only in the final analysis; not before the detour of causal, empirical-analytic methods. For this reason Habermas characterizes his approach as a hermeneutic one, though as a special form of "depth hermeneutics" (*ibid.*: 272), in contrast to what can be termed as Gadamer's "surface" hermeneutics, which is considered capable of revealing and removing only shallow impediments to true understanding.

Both hermeneutic and critical-emancipatory reflection rely in the end on agents' own self-understanding. The question that separates them is by what method, if any, can we probe deep enough for authentic self-understanding. And in this respect Habermas offers a systematic method, whereas Gadamer concentrates rather on the ontology of understanding.¹²

2) *The psychoanalytic analogy*

Confronted by various "*prima facie* disanalogies" between the psychoanalytic therapeutic method and Habermasian critical social theory based on the idea of emancipatory self-reflection, Thomas McCarthy concludes that perhaps we should not take the comparison to be anything more than a "broad metaphor":

Perhaps we have taken the model too literally, and there is no need to find a correlate for every feature of the

psychoanalytic situation. In any case it is evident that the model of psychoanalytic therapy is a rather broad metaphor when applied to the organization of political enlightenment (McCarthy, 1978: 212-3).

Nonetheless, two particular areas of criticism of the analogy would repay examination in terms of sharpening the Habermasian concept. These concern, firstly the "therapist-patient" relationship, and secondly the relationship between enlightenment and emancipation.

(a) *The therapist-patient relation*

In the psychoanalytic situation the relationship between therapist and patient is an asymmetrical one, the therapist being in a privileged, "superior" position *vis à vis* the patient, in the sense of being one who can "see through" the other – the patient – and enlighten him/her about things the patient cannot see about him/herself.

But at the same time the psychoanalytic therapist-patient relationship is socially institutionalized, defining role-limits that safeguard the autonomy of the patient, not least his/her right to voluntary entry into and exit from that relationship. The psychoanalytic patient is one who freely submits to the analyst's treatment, having experienced a problem from which he/she wishes to be released.

Critics point out that no analogous relationship exists between social and political theorists with enlightening aspirations on the one hand and (to-be-enlightened, ideologically self-deluded) members or sections of society on the other. Such a situation cannot be assumed to be generally prevalent in society, nor is it institutionalized for any circumscribed area of social life. And if any would-be enlightener-emancipators attempted to assume such an asymmetrical role towards

society, they would, at best, be rejected, and thus unsuccessful. Gadamer makes this point quite eloquently:

But what happens when [the psychoanalyst] uses the same kind of reflection in a situation in which he is not the doctor but a partner in a game? Then he will fall out of his social role! A game partner who is always "seeing through" his game partner, who does not take seriously what they are standing for, is a spoil-sport whom one shuns (Gadamer, 1967: 41).

At worst, the aspiring enlightener-emancipators would slide in the direction of an authoritarian, oppressive attitude towards those who resist their enlightening-emancipatory efforts. In Habermas's own words, there is a danger of "exploitation and deception", or even of "uncontrolled exercise of force on the part of self-appointed elites". (Habermas, 1974: 33, 16)

Habermas makes it clear that the "therapeutic" relationship he envisages is not primarily – as Gadamer, agreeing with Giegel, seems to assume – one between the oppressed and their oppressors, between the exploited and the exploiting class, where the former tries to enlighten the latter in order to bring an end to the oppressive or exploitative relationship. It is, rather, first and foremost, a relationship between theorists concerned with enlightening and emancipating the oppressed and exploited classes, and those classes themselves. In terms of the Marxist paradigm, which Giegel himself refers to,

[t]his model can only be used for normatively structuring the relationship between the Communist Party and the masses who let themselves be enlightened by the Party concerning their own situation (*ibid.*: 30).

Habermas, however, does not rule out processes of enlightenment aimed toward broader sections of society, including even the (equally self-deluded) ruling classes. But, whether, in any given situation, any social group would "let themselves" enter into a "patient-therapist" relationship with an aspiring enlightener-emancipator, is a contingent, empirical matter. There is no guarantee that this would happen. It depends largely on whether any social group has a sufficiently strong feeling of suffering to motivate them to seek help, and at the same time sufficient confidence in a potential "therapist".

To be sure the development of the explanatory, causal hypotheses intended for enlightenment is, for Habermas, a process that can be carried out by critical social theorists without the participation or the cooperation of those for the sake of whose enlightenment the hypotheses are developed. It is a scientific process that takes place prior to, and is distinct from, the process of enlightenment, in which the active role of the ones to be enlightened is absolutely decisive (see, *ibid.*: 31).

Concerning the asymmetrical nature of the relationship between the social enlightener and those to be enlightened, Habermas stresses that this is initially inevitable, due to the systematically distorted communicative situation, rooted in self-deception on the part of the latter; but the aim of the enlightening reflection process is precisely to remove the self-deception and the consequent distortions, and thus to establish a purely symmetrical communicative relationship. After all, the truth of the causal hypotheses proposed by the enlightener can be confirmed only by the one to be enlightened:

the vindicating superiority of those who do the enlightening over those who are to be enlightened is theoretically unavoidable, but at the same time it is fictive and requires

self-correction: in a process of enlightenment there can only be participants (*ibid.*: 40).

Of course, Habermas is not naive as to the dangers arising out of the element of asymmetry which is involved in the psychoanalytic model of social and political emancipation; and he has no illusions about the fact that, in the absence of institutionalized safeguards, much reliance must be placed on the "commitment" of the "self-appointed" enlighteners not to abuse their position:

processes of enlightenment (if they are to avoid exploitation and deception) can only be organized under the precondition that those who carry out the active work of enlightenment commit themselves wholly to the proper precautions and assure scope for communications on the model of therapeutic "discourses" (1974: 33-4).

It is worth exploring the idea of applying the model of the psychoanalytic therapist-patient relationship in a reciprocal manner: in such a way that each of the two sides in the relationship would have the opportunity to take on both the role of therapist and the role of patient *vis à vis* the other, in what we might call a situation of mutual asymmetry (to be transformed into one of dialogical symmetry).

Such a concept would mitigate somewhat the dangers of one-sided asymmetry, but, more importantly, it would expand the scope of application of emancipatory self-reflection. It would cover situations of social interaction where no one individual or group would be exclusively characterized as "being an enlightener" or as "being in need of enlightenment", while it could still be recognized that the interactions are far from meeting the conditions of undistorted communication. Such situations are, arguably, more prevalent in late modernity than in Marx's time, or even in the time of the young Habermas.

(b) The relationship between enlightenment and emancipation

In the case of psychoanalysis, enlightenment and emancipation are fused into one. A successful psycho-analytic process, in which the patient comes to confirm the validity of an explanatory hypothesis offered by the therapist, concerning the causal power of certain unconscious intentions, is supposed to release the patient from the causal power in question. Once reflectively brought to consciousness, the so far unconscious connection is dissolved.

In the case of critical social reflection, as proposed by Habermas, this, evidently, cannot be so. A process of social enlightenment might be successful in changing the minds of those it addresses, but would not, by itself, remove the social-institutional aspects of the unintended social causes revealed through it. A marxist critique of the ideology of equal exchange, for example, no matter how successful in convincing its addressees, would not automatically lead to the abolition of the legal and institutional bases of capitalist commodity exchange. To that end, political action would be necessary, over and above the process of enlightenment.

Responding to critics on this point – including his close associate Karl-Otto Apel, who criticizes him for "simple identification of reflection and practical engagement" (McCarthy, 1978:96) – Habermas admits that he neglected organizational-political issues and offers clarifications that dispel any suspicions of idealism or political passivity.

Habermas (1974: 32) distinguishes the psycho-analytically modelled process of critical self-reflection by social groups from two other phases of the enlightenment-emancipation continuum: On the one hand he distinguishes it, as we have already seen, from the development of explanatory hypotheses, or "critical theorems", which precede self-reflection. But he also distinguishes it from the political

activities that must follow the self-reflection and aim to bring about emancipatory social change, according to the insights gained in the preceding self-reflective process. This third phase, that of political organization and activity, is absolutely essential. Without it any enlightening insights that might have been gained in the self-reflective phase would remain politically sterile. At the same time Habermas emphasizes (and criticizes Lukács, Lenin and much of the working class movement for not being clear on this point) that these three phases consist of three different kinds of activity, based on different methodological principles, and different kinds of relationship between the aspiring enlightener-emancipators and the ones to be enlightened-emancipated.

In the first phase, the former group develop and discuss, according to the standards of scientific method, explanatory hypotheses about the latter group, without the latter group's participation. In the second phase the two groups engage in an asymmetrical, psychoanalytic-type relationship, where the former – in the role of "therapist" – communicates to the latter the potentially enlightening explanatory hypotheses developed in the previous phase, and the latter – in the role of "patient" – by self-reflecting on the hypotheses may or may not gain the desired self-enlightenment, subject to their own judgement. In the cases where such enlightenment is attained, the asymmetrical "therapist-patient" relationship between the two sides ceases to be necessary, and the two become potentially equal partners in – ordinary, symmetrical – dialogue.

In the third phase, the enlighteners and the newly enlightened, now as one group sharing the same emancipatory interests, must take common decisions, by means of ordinary dialogue, about emancipatory political action *vis à vis* their political opponents. In this action, those opposing the changes necessary for social emancipation can be treated in an instrumental-strategic manner, in which case

the action will include "the selection of appropriate strategies, the solution of tactical questions and the conduct of political struggle" (*ibid.*: 32). But Habermas, making approving references to Marx, does not rule out the possibility that the opponent can also be treated, depending on contingent circumstances, as another partner in dialogue, as in situations where

the weapon of critique will have greater effect than the critique of weapons. These are situations in which the initiatives of radical reformism, which seek to persuade not only within the group, but also externally, are more promising than the revolutionary struggle (*ibid.*: 31).

To be sure, other things being equal, the latter, communicative approach is in principle preferable. Any treatment of others in an instrumental-strategic way would require strong justification, and such justification cannot be given until after the event, in terms of the self-acknowledged emancipation of those so treated. And of course, any such *ex post facto* consensual justification of the emancipating methods – just as the justification of enlightenment through a consensus between the oppressed and their vanguard – is always fallible, subject to illusion (see Wellmer, 1971: 49).

Furthermore, the very possibility of bringing about the emancipatory changes attempted is also in question, being theoretically impossible to decide until confirmed in historical practice. For a nomological connection claimed to be a case of "causality of fate" – one depending entirely on ideological self-delusions on the part of social agents – and therefore in principle dissolvable through social action, can eventually prove to be in fact a case of "causality of nature", socially impossible to dissolve.

Attempts at emancipation ... are... tests; they test the limits within which human nature can be changed and above all, the limits of the historically variable structure of motivation, limits about which we possess no theoretical knowledge, and in my view, cannot in principle possess any (Habermas, 1974: 37).

3) *Critical self-reflection as praxial rationality*

We can identify three distinct ways in which rationality can move us to action: Firstly, it can do so by changing our factual/theoretical beliefs about the nature of reality in such a way as to reveal a hitherto unrecognized divergence between reality and those already held values that determine our view about how reality should be. In this case, it is theoretical rationality that makes the difference that leads to action, though, strictly speaking, the motivating power is provided by the values. Secondly, rationality can make us undertake a specific action by demonstrating that the action in question would be an effective way of attaining a certain – already desired – end. This, of course, is the work of technical rationality, which, though classified by Habermas as a form of practical rationality – the other form being praxial rationality – is ultimately reducible to a form of theoretical rationality. And thirdly, rationality may bring about a change in our values themselves, thereby creating a new imbalance between our ideas about what is and our beliefs about what should be, thus moving us to action which aims to correct the imbalance.

It is only in this, third capacity that rationality is praxial in the Habermasian sense. It is evident that with regard to the theory of rationality Habermas's efforts in both his early and later work concentrated on developing a satisfactory concept of such a praxial rationality: a rationality that is practical, in the sense that – unlike theoretical rationality – it gives us directions about how to act, and at the same

time it is not of the technical/instrumental/strategic kind, which is merely concerned with the effective realization of pre-given ends by means of technically manipulating objects. He is concerned with establishing a form of rationality which is capable of giving direction to action through its having implications for values and norms themselves.

Of course, Habermas rules out a praxial rationality akin to Natural Law theory or other forms of value naturalism, while not (yet), on the other hand, considering a Kantian form of practical reason. To be sure, Gadamerian hermeneutics might be considered as praxial rationality of the kind sought by Habermas for it contributes not only to the clarification but also to the modification of identities and evaluative-normative attitudes; and Habermas himself does recognize this action-orienting quality of hermeneutics. However, as we have seen, he believes that Gadamerian hermeneutics keeps too close to existing reality to be able to establish evaluative-normative standards capable of judging reality critically, which is the cornerstone of his kind of Critical Theory. For this reason, he would consider Gadamerian hermeneutics at best a weak form of praxial rationality, one that would need to be supplemented by a stronger form, capable of establishing values and norms in a more radical way, on the basis of a sufficient critical distance from existing social reality.

The depth-hermeneutical method of critical self-reflection is proposed by the early Habermas as precisely such a strong form of praxial rationality, to supplement the weaker form of "ordinary" hermeneutics. In this, Habermas differs from previous forms of Critical Theory, beginning with Marx. Marx's critique mainly turns upon (and against) certain factual beliefs about the nature of capitalist economic exchange, in particular the belief that the capitalist market is operating according to the principle of free and equal exchange, in a manner ordained by nature. Marx

did not intend to challenge the dominant bourgeois ideology of his time, which emphasized the values of freedom and equality; rather, his intention was to ride the wave of the prevailing ideology, showing firstly the gap – the "contradiction" – that existed between the values proclaimed by that ideology, and the capitalist reality of unfree and unequal exchange, and secondly the fetishistic/reified character of that reality. Awareness by social agents, above all by the proletariat, of the "contradictions" between ideals and reality, as well as of their own role in fetishistically sustaining that reality, would lead to action that would change social reality so as to bring it into harmony with the predominant values of freedom and equality. Thus, the rationality of Marx's critique is, strictly speaking, theoretical rather than praxial in the Habermasian sense.

Similarly, the Lukács of *History and Class Consciousness* – still believing in the working class as a potential oppositional social force espousing ideals, such as freedom and equality, that contradicted and transcended the capitalist reality of the time – concentrated mainly on the theoretical tasks of interpreting social reality and showing its reified character. The very different post-1930's Frankfurt School thinking paradigmatically articulated in Horkheimer and Adorno's *Dialectic of Enlightenment* is not a form of praxial rationality either – at least in Habermas's strong sense. For the post-1930's Horkheimer and Adorno as well as the Marcuse of *One-Dimensional Man*, while believing that there is no longer a gap between the values actually held by a – totally integrated – working class and the reality of advanced capitalism dominated by the ideology of instrumental reason, are at the same time deeply pessimistic about reason's ability to help by establishing critical standards transcending currently prevalent societal norms, or engage in value-changing discourse. They are indeed profoundly sceptical about reason's ability to do anything but play an instrumental role.

The early Habermas dissents to some extent from the sociological interpretation of advanced capitalism put forward by the post-1930's Frankfurt School. He reworks the theme of the ideology of instrumental reason into the closely related theme of technocratic ideology, and more importantly replaces the claim that advanced capitalist society is seamlessly and inescapably engulfed in (instrumentalist) ideology with the more moderate view that (technocratic) ideology is ominously spreading but has not yet succeeded in pervading the whole society – a view that leaves a (sociological) window of opportunity for social emancipation. But where Habermas most sharply differentiates himself from his Frankfurt mentors is in rejecting their profound (philosophical) pessimism concerning reason's ability to be anything but instrumental. He emphatically maintains that reason can be praxial in the strong sense explained above, through changing values by means of critical emancipatory self-reflection, in addition to being action-orienting/will-determining in a theoretical way through changing entrenched factual beliefs.

In light of the above, what is philosophically crucial (and most controversial) in the three-phase process of enlightenment and emancipation proposed by Habermas is neither the formation of explanatory hypotheses by the "therapist"-theorist, nor the undertaking of emancipatory action following the enlightenment that ensues from a successful self-reflective hypothesis testing. The crucial transition is, rather, the formation of the "patient's" will to emancipatory action *via a change in values effected in a process of enlightening self-reflection*. Here lies the key to Habermas's claim that the method of critical-emancipatory self-reflection is an adequate form of (strong) praxial rationality.

The question is, in what precise way and to what precise extent can critical self-reflection affect the values of those subject to self-reflective enlightenment? In

what precise way does it constitute praxial rationality, and is it as strong a form of praxial rationality as Habermas would have wished?

[S]elf-reflection ... traces back the formative course of an ego's identity through all involutions of systematically distorted communications and brings this analytically to this ego's awareness ... (Habermas, 1974: 16)

Self-reflection brings to consciousness those determinants of a self-formative process of cultivation and spiritual formation which ideologically determine a contemporary *praxis* of action and the conception of the world (*ibid.*: 22).

Thus, in Habermas's view, self-reflection makes the subject conscious of the process through which its own identity was formed. Self-reflective hypothesis testing brings to light compulsions, repressions and other distortions in that process, and reveals how these have unconsciously marked the subject's identity.

Furthermore, the self-awareness generated by self-reflection releases the subject from the spell of the powers that have unconsciously and falsely determined significant aspects of its identity such as behaviours, perceptions and assumptions.

[Psychoanalytic] self-reflection means the disclosure and analytical *negation* of unconsciously motivated compulsive behaviour and of limits to perceptions tending to function like false *a priori* assumptions (Habermas, 1972: 359).

This self-reflective process, by "negating" unconscious falsehood concerning one's identity, clears the way for the development of an *authentic* self:

successful self-reflection results in insight which satisfies not only the conditions of the discursive realization of a claim to truth (or correctness) but in addition satisfies the condition of the realization of a claim to authenticity... In the patient's acceptance of the "worked out" interpretations which the doctor suggests to him and his confirming that these are applicable, he at the same time sees through a self-deception. The true interpretation at the same time makes possible the authentic intention of the subject ... (1974: 23-4).

No doubt this move to authenticity, via the removal of self-deception, applies to such important aspects of self-identity as interests, values and normative principles. Critical-emancipatory self-reflection is, thus, a form of praxial rationality in the sense that it leads – through a process of rational discourse – to a change in the subject's values and norms.

But is this, for Habermas, a strong enough praxial rationality? The values authentically adopted by one subject are certainly not "objective", in the sense of being *ipso facto* applicable to all subjects. Each subject can, in principle, authentically espouse a different set of values. But what if a whole social class, or better still the entire society, or even humanity as a whole, could be treated as a single, collective, self-reflecting subject, authentically adopting certain values? Such values would be "subjective", in a sense, yet general, or even universal. Such an achievement of self-reflection might satisfy Habermas's criteria for praxial rationality. However, in his (self) critical writings following the initial publication of *Knowledge and Human Interests*, Habermas warns against considering collectivities of individuals as "large-scale subjects". He charges Marx explicitly,

and himself implicitly, with such a mistake, explaining that attributes of collectivities of multiple subjects must be derived in an intersubjective way, that is, not through the subject-centred method of critical self-reflection.

Thus to the objective structures within which socialized individuals encounter each other and act communicatively, large-scale subjects are assigned. The projective generation of higher-order subjects has a long tradition. Marx too did not always make clear that the attributes ascribed to social classes (such as class consciousness, class interest, class action) did not represent a simple transference from the level of individual consciousness to that of a collective. These are rather designations for something that can only be arrived at intersubjectively, in the consultation or the cooperation of individuals living together (1974: 13).

In "A *Postscript to Knowledge and Human Interests*" Habermas maintains that to have "a *rational* human will", that is, *praxial* rationality in his strong sense of the term, it must be possible to rationally determine "generalizable interests and norms"; and he claims that this can be done through the intersubjective processes he now refers to as "practical discourses":

practical discourses are capable of testing which norms manifest generalizable interests and which are merely based on particular interests (1972: 372).

He is not, however, able at this stage to spell out the nature of these practical discourses. He would do that, in the form of his discourse theory of morality, only

after he had altogether given up the subjectivist, consciousness-based paradigm of his early work and developed the intersubjectivist, communicative paradigm.

In the latter paradigm Habermas entirely abandons the idea of critical-emancipatory self-reflection. As I shall suggest in Part Three, he did not have to do that. Such a form of self-reflection, when developed in a communicative/dialogical direction and purged of the weaknesses acknowledged in Habermas's own self-criticism, could prove valuable in the new paradigm too, being linked to the notion of authenticity, which remains an integral and important part of the new complex concept of practical rationality.

4) *Critical-emancipatory self-reflection vs rational reconstruction*

Thus, critical-emancipatory self-reflection can, at best, enable a subject to adopt (substantive) values and norms which are authentically its own, but cannot establish objective/universal values and norms. These can be determined only by means of "practical discourses", a procedure that lies beyond the paradigm of *Knowledge and Human Interests* and remains, at this transitional stage in Habermas's thought, still largely unclarified.

There is, however, a certain range of universal values and norms which cannot be determined even through practical discourses, but nonetheless are claimed to be established in *Knowledge and Human Interests* through a method other than critical self-reflection. These are the universal interests in technical control, communicative understanding and emancipation, which are constitutive of the knowledge provided by the empirical-analytic, practical and critical-emancipatory sciences, respectively. What is the method by which these fundamental "cognitive" interests are determined?

[K]nowledge-constitutive interests... are not susceptible to justification in *practical discourses*. They cannot be recognized as generalizable through the mechanism of discursive decision-making, but can only be *found to exist* as *general* interests through a process of rationally reconstructing the conditions of how experience can be objective. The *universality* of cognitive interests implies that the constitution of object domains is determined by conditions governing the reproduction of the species, i.e. by the socio-cultural form of life *as such*. To speak in this context of *a basis of interests* is justified precisely because the cognitive strategies serving the creation of technical, practical and emancipatory (true) knowledge are related to general classes of problems pertaining to the reproduction of human life... (Habermas, 1972: 372).

The universal cognitive interests are, therefore, established by means of the method of rational reconstruction, in this case the rational reconstruction of the universal conditions governing the reproduction of the human species, which determine the constitution of different object-domains for human knowledge and make possible different kinds of science.

Rational reconstruction is a Kantian, transcendental form of self-reflection, different from critical self-reflection, yet also present in *Knowledge and Human Interests*. Moreover, it is more fundamental than – indeed it is a precondition for – the critical-emancipatory form of self-reflection. For it is only through rational reconstruction that the architectonic of the theory of cognitive interests, including the idea of emancipation through critical self-reflection, is developed. Critical self-

reflection on distorted modes of identity-formation (self-deceptions, distorted communication, etc.) must presuppose the idea of a "normal", undistorted self-formation process, and such an idea, according to Habermas, can only be provided by reconstructive reflection.

These two types of self-reflection differ in some important respects. Above all, whereas critical self-reflection aims to reveal to the subject, and as a consequence dissolve, the "pseudo-objectivity" of certain objects of the subject's experience, rational reconstruction, by contrast, aims to "explicate" "objective data" – unquestionable givens – "*without involving practical consequences*" (*ibid.*: 378).

Despite the differences, however, Habermas failed to distinguish clearly between these two types of self-reflection, and reproaches himself for this failure:

The studies I published in *Knowledge and Human Interests* suffer from the lack of a precise distinction ... between reconstruction and "self-reflection" in a critical sense... [T]he traditional use of the term "reflection", which goes back to German Idealism, covers (and confuses) two things: on the one hand, it denotes the reflection upon the conditions of potential abilities of a knowing, speaking and acting subject as such; on the other hand, it denotes the reflection upon unconsciously produced constraints to which a determinate subject ... succumbs in its process of self-formation (*ibid.*: 377).

Following his realization of the distinctiveness and unique power of reconstructive reflection – it was, after all, fundamental to the paradigm expounded in *Knowledge and Human Interests*, albeit without him being fully aware of the fact – and given the problems he came to recognize with regard to the critical-emancipatory form of

self-reflection, Habermas resolved to make the method of rational reconstruction, in a more self-conscious, explicit and systematic way, the cornerstone of his new theoretical departure. But instead of using it, within the framework of what he thought to be a flawed and limiting subjectivist philosophy of consciousness, to inquire into the conditions of knowledge, he would now employ it within an intersubjectivist, communicative paradigm of thought to reveal the "universal pragmatics" of linguistic interaction. A prominent place in this new paradigm would be occupied by the discourse theory of morality, itself a part of a more differentiated and complex conception of practical rationality.

PART TWO

PRACTICAL RATIONALITY IN HABERMAS'S COMMUNICATIVE PARADIGM

2.1 INTRODUCTION

2.1.1 Taking Stock

In working out his early paradigm, which was presented in its most developed form in *Knowledge and Human Interests*, Habermas was attempting to take Critical Theory out of the cul-de-sac into which he believed it had been driven by the first generation of Frankfurt School theorists, Adorno above all. To that end, he sets out to develop an adequate conceptual framework that would include both forms of rationality, instrumental and praxial, in a proper relationship to each other. Such a framework would be the basis for a renewed Critical Theory of society.

This aspiration was realized in terms of a theory of knowledge in which three "quasi-transcendental" and anthropologically fundamental cognitive interests underpin three forms of knowledge and, in the last analysis, three kinds of sciences: the technical interest is linked in this way to the empirical-analytic sciences, the practical interest to the historical-hermeneutic sciences, and, most importantly, the emancipatory interest to the critical social sciences, modelled on Freud's psychoanalysis and Marx's critique of ideology. Critical social sciences are based on the method of critical self-reflection, which for the early Habermas represents the highest form of praxial rationality.

As explained in the previous Part, this early Habermasian paradigm for Critical Theory, put forward in the late 1960's, soon came to be seen – not least by Habermas himself – as having important defects. The main shortcomings of the paradigm are the following: 1) There is a serious ambiguity concerning the

status of cognitive interests: are they universal anthropological constants, that is, are they part of nature, or are they transcendental preconditions of types of knowledge? The term "quasi-transcendental", used by Habermas, obfuscates rather than illuminates the situation (see Habermas, 1971: 14).

2) There is no clear distinction between two very different senses of critical reflection, namely (a) (emancipatory) critical self-reflection on the psychoanalytic model and (b) rational reconstruction, with emphasis placed on the former. The lack of clarity about the nature of rational reconstruction – the logically more fundamental of the two – is not unrelated to the ambiguity identified in (1) above.

3) Questions about the truth or validity of statements are not clearly distinguished from questions concerning the constitution of possible objects of experience, about which statements are made. More attention to the question of assessing truth claims would have pointed in the direction of intersubjective, discursive processes.

4) Societies, even humanity as a whole, as well as more limited social collectivities such as social classes, tend to be misconstrued as single, undifferentiated macro-subjects capable of reflecting on themselves, rather than being treated as the complex, predominantly intersubjective realities that they are. And this despite Habermas's explicit emphasis on intersubjective interaction and rational debate in several essays of roughly the same period as well as in the slightly earlier *Structural Transformation of the Public Sphere*.

5) The power asymmetry, in the process of emancipatory-critical self-reflection, between enlightener and enlightened-to-be, has been criticized for opening possibilities of abuse by the former, more powerful party. And although Habermas might have dealt with this problem by developing the concept of self-reflection in a

dialogical/communicative direction, in later work he largely ignored its considerable rationalizing potential.

6) In the concept of critical-emancipatory self-reflection two different kinds of issues which ought to be distinguished and dealt with separately are collapsed together: the question of (Critical) *theory* and the question of political organisation and *action*.

7) As a process of praxial rationality, emancipatory self-reflection goes only as far as bringing up the authentic values of the subject of reflection, failing to establish universally valid substantive values or norms. The paradigm did of course claim to have established, on the evaluative-normative side, the three cognitive interests. But there again the endeavour has been problematic, as pointed out in (1) above.

8) The dualistic typology of action and rationality in terms of *techne* and *praxis*, work and interaction, seems to exclude categorically actions and instances of rationality which are at once interactive and oriented to the solution of instrumental, means-ends problems of the interacting agents.

2.1.2 New Departure

Under the weight of criticism, especially self-criticism, Habermas, rather than seeking to mend his early paradigm, abandoned it altogether (some believe wrongly so) and embarked – from the early 1970's onwards – on developing an entirely new one. At this juncture his aims as well as his *prima facie* assessment of the socio-historical situation and its challenges for a critical thinker remained much the same as before. He was still both acutely aware of multiple post-liberal forms of domination and reification in society, and a firm believer in the possibility – though not the inevitability – of emancipatory social transformations. This latter

belief was in part based on what he saw as evidence of new emancipatory movements on the one hand, and of significant elements of genuine social progress on the other. He saw modernity and its Enlightenment foundations as a project still unfinished, not as one that had simply failed to live up to its promise.

Despite the difficulties he experienced with his earlier attempt, as well as the new radical questioning of rationality, most notably by poststructuralism, Habermas remained at bottom convinced of the philosophical possibility and social efficacy of a critical, yet normatively constructive theory of society; one based on a model of rationality that encompasses a strong praxial dimension while still avoiding the dogmatic side of Enlightenment rationalism.

Given his critical assessment of his earlier attempt, Habermas felt that the new paradigm had to meet the following requirements:

- 1) It had to be based on the self-conscious and consistent employment of a rational-reconstructive methodology.
- 2) Questions of validity and truth, as distinct from questions of object constitution, should figure prominently in it.
- 3) It had to be thoroughly intersubjectivist. The Cartesian philosophy of the subject and of consciousness came to be considered by Habermas as responsible for the exclusively instrumentalist conception of reason which, together with its reifying and dominating social repercussions, has prevailed in modern times.¹
- 4) Intersubjectivity should be consistently extended beyond the philosophical aspects of the paradigm, to the substantive theory of society, where any lingering temptation to treat society as a macro-subject should be strongly resisted.

5) It should establish sufficiently strong forms of praxial rationality, capable of providing non-relativistic (normative/evaluative) standards of critique and guidance for action.

6) It had to respect the distance between social theory and political action.

To these requirements I should add a further one: that the new paradigm should successfully address the problems identified with regard to the dualistic nature of the typology of action and rationality.

As it was presented in the 1980's and 1990's – first comprehensively stated in the two-volume *Theory of Communicative Action* and then further developed and elaborated in works such as *Moral Consciousness and Communicative Action*, *Justification and Application* and *Between Facts and Norms* – the new, "mature" paradigm consists of four main dimensions:

(1) A theory of Universal, or Formal, Pragmatics, derived from the application of the rational-reconstructive method to the intersubjective realm of linguistic communication.

(2) Typologies (a) of social action, and rationality, and (b) of forms of societal integration.

(3) A model of practical rationality that includes a discourse theory of morality and in its later developments, discourse theories of law and democracy.

(4) A substantive Critical Theory of modern society constructed around the concepts of lifeworld and system and the thesis of the colonization of the lifeworld by media-steered sub-systems of strategic action.

The rest of thisPart constitutes a presentation and critical discussion of these four dimensions of the paradigm, with inevitable emphasis on practical rationality, particularly the themes that will bemy focus of attention inPart Three.

2.2 FUNDAMENTALS OF HABERMAS'S UNIVERSAL PRAGMATICS OF LINGUISTIC COMMUNICATION

2.2.1 A synoptic presentation

I would never have tackled a formal-pragmatic reconstruction of the rational potential of speech if I had not harboured the hope that by means of this approach I would be able to generate a concept of communicative rationality from the normative contents of the universal and ineluctable presuppositions of a non-circumventable practice of everyday process of reaching understanding (Habermas, 1991: 243)

The starting point of Habermas's communicative paradigm is the application of the method of *rational reconstruction* to what is at once a universal and an intersubjective feature of human existence: linguistic communication. The idea is to develop a theory of Universal, or Formal, Pragmatics (Habermas, 1979: 1–68) which would “[reconstruct] universal and necessary conditions of communicative action” (Habermas, 1984: 139). Such a theory would provide certain normative standards for social action – “standards for normal, that is, undisturbed communication” (*ibid.*) – and, hence, yardsticks for a critical theory of society.²

The method of rational reconstruction employed by Habermas is akin to Kant's Transcendental Deduction, albeit not applied to the consciousness of a subject, and without aprioristic, transcendental claims (Habermas, 1982: 233–4). Just as Kant tried to reconstruct, or “deduce”, from the inescapable facts of ordinary human experience, those features of the consciousness of the individual human being without which that experience would be inconceivable, so Habermas sets

out to reconstruct those features of language – more precisely of the intersubjective realm of linguistic communication – which are universal and necessary, i.e. unavoidable, and without which that communication would be impossible. In particular, he tries to unearth speaking agents' "communicative competence", i.e. those intuitive, pretheoretical presuppositions, or that know how (here Habermas refers to Ryle's distinction between *know how* and *know that* – see Habermas, 1979: 12) without which agents, even if they are not fully aware of it, could not communicate in the way they in fact do: the competence which one cannot deny having – the presuppositions one cannot deny making – while performing an act of communication; for in that case he or she would be involved in a "*performative contradiction*" (Habermas, 1990: 80).

Habermas considers his Universal Pragmatics as an example of a "reconstructive science", on a par with other such sciences as Chomsky's structural linguistics and Piaget's developmental psychology. In his view, reconstructive sciences are disciplines that combine formal-conceptual and empirical elements, and produce falsifiable knowledge – knowledge open to revision in the light of further reflection or empirical evidence. It is in this way that he understands his own Universal Pragmatics (Habermas, 1979: 21–5), in contrast to his close collaborator Karl-Otto Apel, who insists on the foundational, justificatory character of rational reconstruction and retains the predicate "Transcendental" for his own Pragmatics of language (see Apel's critical discussion of Popper's and Albert's fallibilism, in Apel, 1987).

Both Apel and Habermas see the method of rational reconstruction as a way out of the "Munchhausen Trilemma", which has led Popper and his followers (most prominently Hans Albert) to their version of fallibilism: the idea that it is impossible to (deductively) provide foundations for any statement. For according to the

Trilemma, any attempt to provide foundations would involve one either in (1) arbitrarily breaking off the chain of reasons, or (2) an infinite regress, or (3) a circular argument. However, whereas Apel maintains that the method of rational reconstruction provides ultimate justification, Habermas – defending a softer version of fallibilism than Popper’s and Albert’s – rejects justificationism (Habermas, 1991: 215). To begin with, any pretheoretical knowledge of agents that is reconstructively revealed cannot be said to be *justified* for the agent; it is simply *presupposed*, as a “fact of reason”. But more importantly, Habermas argues, the philosophical *reconstruction* of this pretheoretical knowledge cannot, itself, be considered ultimately justified and indubitable; reconstructions are “hypotheses” (Habermas, 1982: 234): “[W]e have to put our reconstructions up for discussion in the same way in which the logician or the linguist for example, presents his theoretical descriptions” (Habermas, 1990: 97).

In reconstructing linguistic communication, Habermas’s starting point is a theory of meaning – one that unifies three pre-existing theories, each of which, in his view, captures only one aspect of a complex picture (1992: 57–64). These are (a) the intentionalist theory, which emphasizes what speakers intend to express, (b) the formal-semantic theory which focuses on speakers’ reference to the external world, and (c) the use-theory of meaning, that brings out the social aspect of language. Building also on Austin’s and Searle’s speech act theories (which highlight the performative/illocutionary functions of language over and above its propositional content), as well as on Bühler’s classification of linguistic expressions on the basis of their “relations to the speaker, the world and the hearer” (*ibid.*: 57), Habermas formulates his own pragmatic theory, not of sentences, or propositions, but of utterances, or speech acts.

The core guiding idea concerning the meaning of an utterance is that of a “*speaker* coming to an understanding with *another person* about *something*” (*ibid.*: 73). Accordingly, for Habermas, every utterance is a complex (speech) act which involves (a) making a statement about the (*external*) *objective world*, (b) expressing some aspect of the (*internal*) *subjective world* of the speaker (subjective experiences such as intentions, beliefs, preferences, feelings, desires, etc) and (c) making an appeal to the (*legitimate*) *social world* shared between the speaker and the hearer (Habermas, 1984: 100).

Implicit in every speech act are what Habermas calls *validity claims*. There are, more specifically, three types of validity claim, namely, (a) that the assertion made in the utterance about the objective world is factually *true*, (b) that the speaker’s expression of (some of) his/her subjective states (to which he/she has privileged access, *ibid.*: 307) is *sincere* or *truthful*, and (c) that the utterance is *right* with respect to the (legitimate) social world that speaker and hearer share (*ibid.*: 99). In some formulations of his Universal Pragmatics, Habermas – in addition to the above validity claims to truth, truthfulness and rightness – refers also to a fourth validity claim, that to the utterance’s comprehensibility, corresponding to a (fourth) world, that of language itself. This fourth dimension of speech acts, however, is normally left out of his further analyses (Habermas, 1979: 68).

In each speech act – whether that takes the form of an assertion, a command, a question, an exclamation, or any other – all three world relations and corresponding validity claims are present, though each utterance may explicitly thematize only one of them, by taking the form of a *constative*, *expressive*, or *regulative* speech act, respectively (Habermas, 1984: 309). Thus, when X says to Y that the lawn is turning yellow, in addition to explicitly making a factual statement about the colour of the lawn, she is also implying that she means what she says

and that, given the social norms that X and Y share or ought to share, she has the right to make that comment in the given context.

Also, when X says to Y “You ought to have declared all your income to the Tax Office”, she explicitly makes a claim to rightness and at the same time implies that she is expressing her sincere opinion on the matter, and also that it is factually true that there is such an institution as a Tax Office and that Y did not declare all his income to it. Similarly, when one directly expresses an intention, in addition to making an explicit claim to truthfulness, he is, simultaneously, also making some implicit claims to truth and rightness.

Furthermore, Habermas argues that implicit in every speech act is the speaker’s commitment, or “promise”, to offer, if challenged, justification, or “redemption”, of any of the (in principle criticizable) validity claims made, and to do so by means of a discursive, dialogical process of argumentation aimed at universal, rationally motivated agreement, or consensus (Habermas, 1984: 26). Habermas insists on an internal, conceptual link between the validity and the very meaning of speech acts: “To understand what a speaker wants to say with such an act, the hearer has to know the conditions under which it can be accepted” (*ibid.*: 307; see also, *ibid.*: 115). Thus, to understand, the hearer has to know how the validity claims made in the speech act can be redeemed. He has to understand and *judge* the reasons given, though, of course, he does not have to agree with them (Habermas, 1991: 230). This point is crucial for Habermas’s derivation of *normativity* from the *description* of the process of reaching understanding. It is also a strongly disputed point, albeit one vehemently defended by Habermas.³

The argumentational process through which validity claims must be redeemed is what Habermas calls *communicative rationality*. Communicative rationality itself presupposes, for its existence, certain “general symmetry conditions”,

encapsulated by Habermas initially in the term “ideal speech situation”, and later (following Apel) also in terms of the Peircean “unrestricted communication community” (Habermas, 1990: 88) and “ideal community of communication” (ibid.: 202). The most fundamental, general requirements of communicative rationality are summed up by Habermas – in the context of his discussion of a theory of argumentation – as follows:

Participants in argumentation have to presuppose in general that the structure of their communication... excludes all force – whether it arises from within the process of reaching understanding itself or influences it from the outside – except the force of the better argument (and thus that it also excludes, on their part, all motives except for the truth).
(Habermas, 1984: 25)

In such a process of argumentation participants intersubjectively acknowledge each other's validity claims and respond to any problematic ones with reasons only.⁴

Thus, communicative rationality/discourse/argumentation aimed at consensus is presupposed in every utterance, as the proper way of settling disputes concerning validity claims explicitly raised or implied in speech acts. It is, in Habermas's words, “a reflective continuation with different means, of action oriented to reaching understanding” (Habermas, 1984: 25).

Communicative rationality, with all the normative force associated with it – the normative force embodied in the rules of argumentation and, above all, in the notion of the ideal speech situation – is in this way reconstructively established as a universal and necessary implication of everyday language use. Anyone

engaging in linguistic communication – i.e. anyone participating in social life, thus any human being whatsoever – who denies this normative concept of communicative rationality is committing a performative contradiction.

Being applicable to all the different types of validity claims – and thus extending over both the theoretical and practical domains – this concept of rationality secures the unity of reason, albeit in its purely procedural dimensions. To be sure, as soon as communicative rationality addresses validity claims of different types, differentiations begin to emerge. Although speech acts call for universal consensus in all three aspects – the truth of the factual/existential claims, the truthfulness of the expressive claims and the rightness of the regulative claims made in them – what are admissible as reasons in argumentation (empirical evidence, norms, interests, etc) vary from one type of validity claim to another. Differences between various types of discourse will be considered in later sections, in the context of discussing forms of practical rationality.

2.2.2 Some important objections and clarifications

Habermas is well aware that language is very often used in non-communicative ways, i.e. in ways not primarily oriented to reaching understanding in the above sense, but oriented mainly to producing what Austin called “perlocutionary” effects. It is used, most importantly, in a strategic way, i.e. with an intention to satisfy aims other than reaching understanding about something, and do so by influencing the other causally, albeit through the use of language; in Habermas’s words, “by [inducing] him to behave in a desired way by manipulatively employing linguistic means and thereby [instrumentalizing] him for his [that is, the speaker’s] own success” (Habermas, 1984: 288). Yet, he insists on considering what, strictly speaking, is a reconstructive analysis of language in its communicative mode, to

be a “Universal” Pragmatics of language as such, not only of language insofar as it is used communicatively. The apparent contradiction is dealt with by Habermas in terms of an argument to the effect that the strategic use of language is “parasitic” on what is the logically prior, “original mode” of language use (*ibid.*, and Habermas, 1998a: 122). The success of a strategic use of language is dependent on the communicative success of the utterances employed:

If the hearer failed to understand what the speaker was saying, a strategically acting speaker would not be able to bring the hearer, by means of communicative acts, to behave in the desired way (Habermas, 1984: 293).

This logical priority of the communicative over the strategic mode of interacting can be extended from linguistically mediated interaction to all social actions (non-linguistic symbolic gestures, non-symbolic actions such as acts of physical violence) insofar as these can exist only in a linguistically structured social context.

Some essential additional clarifications of the above most fundamental and general aspects of Universal Pragmatics can be made through a brief consideration of important Habermasian responses to major lines of criticism.

1) The concept of communicative rationality, including the notion of the ideal speech situation which is inseparable from it, has been criticized for excessive intellectualism. Too much emphasis is claimed to be placed on the cognitive side of human beings, at the expense of their a-rational, affective, instinctual, sensuous, aspects. There is, in Agnes Heller’s words, neglect of “the human being as a whole, as a needing, wanting, feeling being” (Heller, 1982: 31). The ideals of autonomy and self-control inherent in Habermasian rationality are similarly

criticized – for instance by feminists – for treating women and men as disembodied beings (Cooke, 1999: 179).

Habermas, in response, explains that the concept of communicative rationality should not be (mis)understood as specifying a concrete form of individual or collective life in its fullness; that we must not be misled into “*inferring* an idea of the good life from the formal concept of reason” (Habermas, 1982: 262). Communicative rationality is just one dimension – albeit an important one – of a very rich reality. And even as one such dimension, it is simply a skeletal procedure, or form, that, in its different manifestations (constative, expressive, regulative), is given substance precisely by the experiences, needs, affections, intuitions, and habits, of real, total human beings; and that allows – indeed enables – them to choose and lead diverse ways of life, both private and public.

2) The idea of consensus as the aim of discourse also comes under attack, notably from Jean-Francois Lyotard, for promoting uniformity and suppressing diversity and difference. In Lyotard’s words, “consensus does violence to the heterogeneity of language games” (Lyotard, 1984: xxv, 66).

That consensus implies uniformity is a fair assessment (though not obviously valid as a stricture) with regard to questions of truth and rightness (moral rightness as we shall see later). It is not so, however, with regard to the third dimension of communicative rationality, the one concerning the expression of subjective states, understood by Habermas, as we shall see, to extend to questions of clarifying individual and collective identities and adopting preferred forms of life. Here, communicative rationality in fact protects difference against pressures for communal conformity and uniformity.

3) The concept of communicative rationality also attracts criticism for being inherently biased in favour of modern Western culture. For it is in the modern European philosophical tradition, and in the institutions of modern western societies, that we come across a clear distinction between the objective world, the subjective world and the social world, as well as the institutional separation of issues regarding truth, truthfulness and rightness, and it is only here that these different types of issues are treated reflectively in terms of argumentative, discursive practices (McCarthy, 1982: 64–5).

Furthermore, the concept of communicative rationality is criticized for being gender-biased, in a male-oriented way. This claim is closely associated with the intellectualist one considered above. The emphasis on thinking, according to this argument, devalues the allegedly more feminine qualities of affection and intuition; on the plane of ethical life the (typically male) perspective of abstract *justice*, based on universal principles, is given priority over the (typically female) perspective of concrete, situational *care* (Gilligan, 1982; Benhabib, 1987).

In the spirit of his fallibilistic, non-transcendental approach, Habermas aims for coherence with other sciences (Habermas, 1991: 231). He thus searches for interdisciplinary corroboration for his universalist claims regarding communicative rationality. In particular, he turns to the developmental psychology of Piaget and Kohlberg, pointing out the close affinity of his own as well as other similar concepts of rationality with what in these developmental models is considered the highest, most mature stage of cognitive and moral development, in Kohlberg's terminology the postconventional stage (Habermas, 1979: 69–94). In addition, Habermas worked on an analogous developmental model of social evolution (*ibid.*: 95–129, 130–177).

Nonetheless, the same critics also challenged the idea of a superior post-conventional stage of development in precisely the same way as they did the Habermasian concept of communicative rationality: with respect to its universality and cultural and gender neutrality (Gilligan, 1982).

4) Communitarians – in the tradition of Aristotle’s ethics, Hegel’s *Sittlichkeit* and Gadamer’s hermeneutics – have, like Gadamer himself before them, reproached Habermas, along with liberalism, for assuming Cartesian, “unencumbered”, “non-situated” agents engaging in purely rational discourses; and for ignoring the necessary embeddedness of all communication and discourse in background knowledge (see, for example, Taylor, 1991: 23–35; and Benhabib, 1992: 39–59).

In his response to Charles Taylor, just as in his earlier exchange with Gadamer, Habermas fully acknowledges the historical-cultural situatedness of human beings, and the “massive preunderstanding” on the basis of which participants in communication carry out their everyday interactions and discourses (Habermas, 1991: 244). Accordingly, he recognises the world-disclosing function of language. But at the same time, he argues that language has also a problem-solving function, making it possible for communicative rationality to “[trigger] off learning processes that may have a retroactive effect on the previous understanding of the world”, and criticizes Taylor for a tendency “to totalize [the] world-disclosing function of language” and for “[allowing the] problem-solving capacity of language to disappear behind its capacity for world disclosure” (*ibid.*: 221, 222).

5) Sympathetic critics have pointed out that actual discourses more often than not result in “false consensus”; and they are referring not to agreements secured through overtly strategic methods such as deliberate deception or the forcible exclusion from a discourse of eligible participants, but to agreements resulting from discursive processes that are entered into by all participants with a genuine

communicative attitude, and believed by them to be (communicatively) rational. Such unwittingly false consensus can come about due to contingent, random distortions in communication arising from insufficient linguistic or other communicative or discursive skills on the part of at least some participants; or due to practically insurmountable obstacles such as insufficient time, or unavailability of some important information. Critics, however, tend to concentrate on what, following Habermas, they refer to as “systematically distorted communication”, as the major cause of false consensus (Young, 2003, 102–120; Bohman, 2000). Over and above the quality of agreements reached in discourses, this type of communication distortion affects also the very process of development of communicative competences themselves, i.e. the contingent historical dynamics through which, what Habermas considers as cognitive, moral and social-historical developmental logics actualize themselves.

Iris Young, for example, identifies two major sources of systematically distorted communication in modern societies. The first is “structural inequalities”:

In a society structured by deep social and economic inequalities... formally inclusive deliberative processes nevertheless enact structural biases in which more powerful and socially advantaged actors have greater access to the deliberative process, and therefore are able to dominate the proceedings with their interests and perspectives (Young, 2002: 110).

The systematic biases referred to here are contingent upon the existence of social and economic inequalities among those participating (formally as equals) in discourse. They are evidently different from, and additional to, the eurocentric and gender biases which, as we have seen, are claimed to be inherent in the very

concept of communicative rationality, and thus reflected in agreements produced even by the most perfectly rational discourses.

The second main source of systematically distorted communication, for Young, is discursive “hegemony” (the latter term used in Gramsci’s sense). The argument is that there are in society entrenched ideological ways of thinking which, unbeknown to them, prevent society’s members – and, now, they prevent all members equally – from conducting fully rational discourses:

The conceptual and normative framework of the members of a society is deeply influenced by premises and terms of discourse that make it difficult to think critically about aspects of their social relations or alternative possibilities of institutionalization and action (*ibid.*: 116).

This point has similarities with the communitarian, contextualist argument presented above, concerning the embeddedness of communicative action and discourse in linguistically encoded background knowledge and preunderstandings. However, whereas communitarians consider such an embeddedness to be an inevitable and acceptable fact of social life, the critics considered here refer to unnecessary ideological distortions which stand in the way of rationally instituted, just social relations, and, therefore, ought to be removed. Furthermore, these critics reproach Habermas for abandoning in his later works (above all in *Between Facts and Norms*) the idea of systematically distorted communication (*ibid.*: 120, n11), while Bohman, responding to the reality of systematic distortion, undertakes his own analysis of it and suggests ways of overcoming it (Bohman, 1996: 107–149).

For his part, Habermas, like Apel and Peirce, has never denied that there is, and will always be, a distance – at best an “asymptotic” relationship – between ideal and actual consensus. For “the consensus required of all concerned transcends the limits of any actual community” (Habermas, 1990: 202). Thus, actual agreements will always have to be open to revision. Habermas makes clear that the conditions of discourse under which ideal consensus can be achieved – i.e. the ideal speech situation – are not part of an actual state of affairs but a counterfactual presupposition that is nonetheless operative in every actual instance of linguistic communication, and without which communication is impossible:

It is part of the structure of possible speech that in performing speech acts (and actions) we act counterfactually as though the ideal speech situation (or the model of pure communicative action) were not merely fictitious but real – precisely this is what we call a presupposition (Habermas, 2001: 102; see also Habermas, 1996: 322–3).

We make this presupposition on entering a communicative interaction, and sustain it during it; but the moment we drop it – which we may do if it becomes obvious that the actual conditions of the interaction diverge significantly from the presupposed conditions – then, the communicative relationship breaks down.

But, by the same token, the ideal speech situation is a normative standard which actual conditions must always strive to approximate and according to which they must always be assessed. As such, it is “the point of departure for a critical theory of society” (Habermas, 2001: 103).

Consequently, for Habermas it is Critical Theory's task to identify and develop ways of removing distortions in communication and the conditions that produce them; for these are the factors determining the degree of falsehood that actual agreements made in society have with regard to questions of truth, truthfulness and rightness. A great deal of communicative distortion, to be sure, is such that can be effectively removed through the skillful use of "ordinary" hermeneutics. Much actual distortion, however, might be more intractable, because it is "systematic", as the critics maintain.

It is true that having developed, in the early 1970's, the important and promising idea of systematically distorted communication – understood as a condition of mutual unconscious deception (and self-deception) – Habermas makes less and less use of it in his later work (though, as we shall see, in *The Theory of Communicative Action* he still includes it in his typology of social actions, as a form of *concealed strategic action*, (see Habermas, 1984: 333). Arguably this neglect is a result of Habermas's dissatisfaction, after *Knowledge and Human Interests* (as we have seen in the previous Part) with what could be an effective antidote – perhaps the only effective one – to systematically distorted communication: the depth-hermeneutical method of critical-emancipatory self-reflection, inspired mainly by psychoanalysis. This method, with its combination of causal and dialogical elements and the asymmetrical therapist-patient relationship it involves, seemed to Habermas too strongly dependent on the (rejected) philosophy of consciousness, and thus ill-suited to the new intersubjectivist paradigm.

6) A final major concern about Habermas's Universal Pragmatics – over and above those relating to the status and nature of communicative action and rationality and the associated notions of consensus and ideal speech situation – takes the following form: Assuming that the normative commitments that are part

and parcel of the concept of communicative rationality are convincingly shown, through the rational reconstruction of linguistic communication, to be logically binding, to what extent can we realistically expect them to become also psychologically binding and thus efficacious in real social practice? What motivational force can the presuppositions presumed to be inherent in every speech act be expected to exert on agents so that the latter do actually relate to each other in communicative ways when their self-interest inclines them to do so with a strategic attitude instead? This question of social efficacy is an important one for an intellectual undertaking centrally concerned with relating theory to practice and acutely aware of the dangers of “utopianism”, in the sense used by Marx with reference to the socialists and other radical reformers of his day.

In this respect, Habermas points out that the normative aspects of his Universal Pragmatics, far from being unrealistic, are in fact met by social reality “halfway”. On the one hand, he once more appeals – albeit controversially, as indicated above – to developmental theories. He does so in order to demonstrate that the universal, inherent logics of development of individual personalities and societies – despite the dependence of the actualization of these logics on contingent historical circumstances – are much in the direction of his own (and other similar, e.g. Kant’s and Rawls’s) concepts. On the other hand, he directs attention to empirical evidence in modern societies of actual rationalization of the communicative – as opposed to the Weberian, purposive – type. He finds such evidence, for example, in new social movements, including feminism, environmentalism and the peace movement (Habermas, 1987: 393–4), as well as in modern democratic institutions, particularly as interpreted in *Between Facts and Norms* (1996), and more generally in what he refers to as the rationalization of the modern lifeworld (see section 2.5.1, below).

But in Habermas's work, the challenge concerning the realistic character of Universal Pragmatics, at a societal level, is met also in another way: by relieving communicative action and rationality from some of the burden of social actualization placed upon it. This is done by way of shifting two huge domains of social life – namely the economy and state administration – from the field of applicability of communicative interaction and discourse to that of systems of strategic action (Habermas, 1984 and 1987; see also section 2.5, below).

In more recent work (Habermas, 1996) Habermas pays more attention, in this respect, to questions concerning discursive processes of public opinion and will-formation, and the connection of the “communicative power” attached to these processes with the “administrative power” of the state, based on strategic action (questions which will also be further discussed in section 2.5).

No doubt the above review hardly does justice to the important and difficult issues regarding the fundamental and general aspects of Habermas's Universal Pragmatics. Some of these issues will be taken up again and discussed further in subsequent sections. These include questions concerning (a) the types of validity claims (b) the relation of communicative rationality to difference and diversity, (c) the possibility of using emancipatory reflection in the context of an intersubjective, communicative paradigm, (d) the motivation (and more generally, the practical requirements) for acting communicatively, and, above all, (e) the dichotomy between communicative and strategic action and its significance for social theory. Some other issues, however, will not be reopened. These include (a) the alleged intrinsic intellectualism and eurocentric and male biases of communicative rationality, (b) the linguistic and cultural embeddedness of discourses and, in connection with it, the world-disclosing and problem-solving functions of language, and, not least, (c) the parasitic relationship of strategic to communicative action,

and (d) the derivation of normative content from the reconstruction of linguistic communication, upon both of which (b and c), arguably, the whole edifice of the Habermasian communicative paradigm rests.

Nonetheless, we shall assume that the fundamentals of Habermas's Communicative Paradigm constitute a sufficiently sound and fruitful framework of thought, that can form the basis on which to explore further the issues that are of central importance to this work: those of practical rationality and its social-theoretical implications.

2.3 TYPOLOGY OF ACTION IN THE NEW PARADIGM

2.3.1 Presentation of main ideas

In the work prior to his communicative paradigm, as we have seen, Habermas works with a dualistic typology of action: *Techne vs Praxis*. *Techne* (or work, or instrumental action, or purposive-rational action) is associated with instrumental or technical rationality – the cognitive and reflective side of *techne* – culminating in modern science-based technology. *Praxis* (or interaction, or communicative action) is associated with what we have termed praxial rationality – the reflective form of *praxis* – combining hermeneutics and critical self-reflection. *Techne* and technical rationality are concerned with effectively realizing already decided goals and values, by means of causally manipulating objects and objective or objectified processes. *Praxis* and praxial rationality, on the other hand, are concerned with the interactive, or communicative (i.e., non-causal-manipulative) clarification and determination of goals, values and norms.

Habermas's early work met with serious difficulties with regard to developing an adequate concept of praxial rationality, due mainly to problems concerning the idea of critical self reflection. Judging these problems to be insurmountable within the philosophy of consciousness, Habermas, as we know, shifted to an intersubjectivist, communicative paradigm, within which he worked out an alternative conceptualization of action and rationality. But in addition to the difficulties with praxial rationality, it was pointed out that the early conceptualization suffered from another important drawback: It seemed to categorially exclude the possibility of action, and of a related rationality of action, concerned with the effective realization of ends and values by way of communicative interaction with other subjects, rather than by causally influencing objects or objectified human beings and social situations. It might be argued that

the existence of that categorial gap, and Habermas's relative indifference to it, are in a relationship of reciprocal influence with his apparent belief that human relations in organizations involved in, or being close to, the highly goal-directed area of material production – i.e., roughly, in work organizations – will, for practical reasons, always have to be largely causal-manipulative, rather than based on communicative interactions; a belief that in the context of his later paradigm takes a much clearer, firmer, and more explicit shape, in his view that, in conditions of modernity, the fields of the economy and state administration in their entirety are best allowed to operate as networks, or systems, of causal-manipulative social relations.

In the new paradigm, Habermas also begins with a distinction between *communicative* and *purposive-rational* action, much along the lines of the early distinction between *techne* and *praxis*, work and interaction. Thus, in purposive rational action

the actor is primarily oriented to attaining an end... he selects means that seem to him appropriate in the given situation, and... 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, 1984: 285).

By contrast, in communicative action, which is at the same time communicative action coordination,

the actions of the agents involved are coordinated not through egocentric calculations of success but through acts of reaching understanding... Reaching understanding [Verständigung] is considered to be a process of reaching agreement [Einigung] among speaking and acting subjects (ibid.: 285–7).

Thus, the dichotomy between communicative and purposive-rational action hinges upon the distinction between, on the one hand, an orientation to reaching understanding, in the sense of agreement, and on the other hand, an orientation to success, in the sense of employing appropriate means to attain desired ends through causally influencing a given state of affairs.

Action oriented to reaching understanding, as we have seen, is reconstructively analysed into three different aspects – the constative, the expressive and the regulative – each associated with a corresponding “world relation”, to the objective world, to the subjective world of the speaking actor and to the legitimate social world shared by speaker and hearer – and a respective mutually recognized validity claim – to truth, truthfulness and rightness. Each type of validity claim, in turn, requires universal consensual “redemption” in terms of communicative rationality – a process of argumentation between equal, mutually respecting agents, and under “ideal” speech conditions.

Purposive-rational action is also subdivided. Habermas firstly distinguishes between *instrumental* and *strategic* forms of purposive-rational action. Instrumental actions are purposive, causal interventions into the physical world or, in general, into any “complex of circumstances and events” (which is not itself capable of acting purposefully). Strategic actions, on the other hand, are actions in which the agents aim to realize their own goals through causally influencing the

decisions of “rational opponents”, i.e. other agents, who are also capable of purposeful action (see Habermas, 1984: 285). Strategic actions are, by definition, social actions in themselves, whereas instrumental actions can be linked to social actions – strategic or communicative – by being their “task elements” (e.g. pruning the trees can be the instrumental task element of the – strategic or communicative – social action of engaging a gardener).

About instrumental action, particularly instrumental relations to nature, Habermas has little to say in the new paradigm.⁵ It can be safely assumed that he has retained his early views on the internal relation between the structure of our technological-instrumental interventions into nature and the predictive character of our nomological knowledge of nature, as well as his belief in the practical indispensability and irreplaceability of such technologically structured *practical* relations to the natural world. Strategic action, by contrast, being itself social action, remains always at the centre of attention, in the context of a paradigm of thought created primarily as a framework for developing a social and political theory.

On the level of typology, Habermas identifies different sub-categories of strategic action. To begin with, he distinguishes between *open* and *concealed* strategic action. Open strategic action is *overtly* oriented to success through causally influencing (the decisions of) other human beings. In concealed strategic action, on the other hand, the orientation to success through causally influencing others is *covert*. Here, a further distinction is made: between concealed strategic action that involves *conscious deception* and similar action involving *unconscious deception*. In the case of interactions involving conscious deception, which Habermas calls cases of *manipulation* – either *one* of the interacting agents deliberately deceives the other – i.e. acts with a strategic orientation – pretending to act

communicatively while the other acts communicatively believing that the other one does so as well, that is, that the interaction is communicative; or, *both* parties knowingly deceive each other, while pretending that they act communicatively and believing that the other one actually does so: i.e. each is at once a deceiver and a victim of deception.

In situations of unconscious deception it is also possible for one or both parties in an interaction to act deceptively towards the other without knowing it. If only one does so, the other one, while being deceived, acts either in a communicative way or in a consciously strategic/deceptive one. What is important in the case of unconscious deception is that whoever unconsciously deceives, that is, deceives while believing that he or she does not do so, is *ipso facto* also involved in *self-deception*. Habermas characterizes interactions involving one-sided or mutual unconscious deceptions/self-deceptions as cases of systematically distorted communication. They are cases in which “false consensus” is typically produced (the discussion of “false consensus” in section 2, above, is relevant in this respect). Figure 1 shows Habermas’s schematic presentation of his typology of social actions (Habermas, 1984: 333).

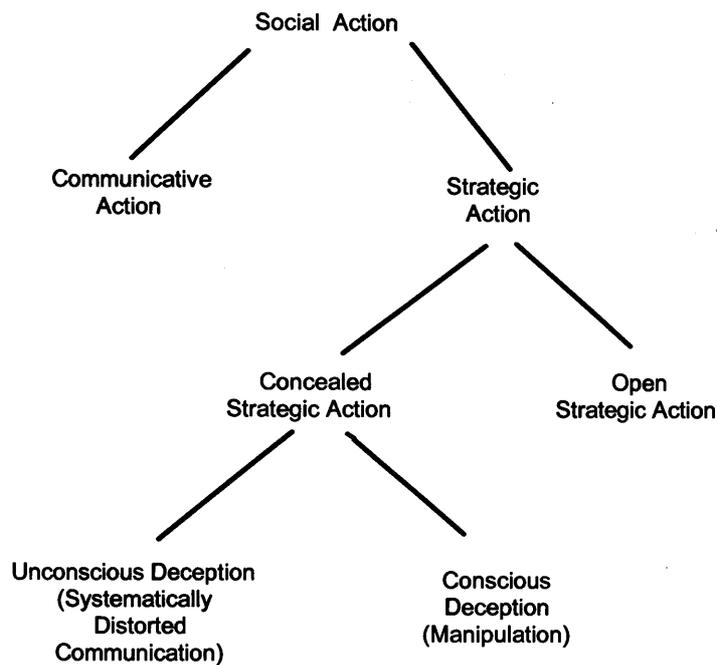


Figure 1

2.3.2 Some objections and clarifications

1) It has been argued that the distinction between action oriented to success and action oriented to reaching understanding cannot be sustained because communicative action is as success-oriented – as oriented to employing means to attain ends – as purposive-rational action is (“One can reach a linguistic understanding successfully” in Berger’s words – see Berger, 1991: 172). Every communicative act aims to succeed in reaching understanding about something and is a means to such an aim. Habermas agrees that both types of action are teleological in a certain sense, but explains that in the case of communicative actions the “end” of reaching understanding is conceptually inseparable from the linguistic “means” used, whereas in purposive-rational action the means and the ends are external to each other.

The illocutionary “ends” of reaching understanding cannot be defined without referring to the linguistic means of

reaching understanding. The medium of language and the telos of reaching understanding intrinsic to it reciprocally constitute one another. The relation between these is not one of means and ends (ibid.: 241; see also Habermas, 1984: 287).

It is on this fundamental difference between an internal and an external relationship of means to ends, that Habermas's differentiation between communicative and purposive-rational action rests.

2) Hans Joas objects that the distinction between communicative and purposive-rational action misses non-teleological forms of dealing with objects in non-social situations such as the "playful and artistic commerce with objects" or the "[p]layful self-development and creative solution of problems" (Joas, 1991: 100, 102). Without disputing the point, Habermas explains that, in developing his typology, he is "concerned with an explanation of social action, not with constructing an anthropology of action as a whole" (Habermas, 1991: 249). However, it might be counter-argued that dealing with objects, though not in itself "socially" interactive – thus not constituting social action in that sense – nonetheless takes place within a social setting, and is for that reason by no means irrelevant to social theory. And it is true that Habermas's discussions of action leave one with the strong impression – perhaps misleadingly – that the only way of dealing with objects – the only form of "non-social action" in his terms – is purposive-rational (instrumental) action oriented to the successful attainment of pre-selected ends through causal influence. Yet, over and above play and art, there are many important ways of dealing with objects which, even though ends-means-oriented, cannot be conceptualized as involving influence of the usual causal variety, nor of course as involving a communicative relationship. Such are, for example, ways of relating to

numbers and other mathematical objects, computer software, scientific concepts, and texts – perhaps to all objects belonging to Popper’s “Third World”.

3) As we shall see in section 5 below, several commentators have criticized Habermas for dividing society sharply into areas governed by communicative action and areas governed by strategic action – namely into lifeworld and system areas – not recognizing that all realms of social life do in fact include a variety of types of action; a criticism which Habermas claims to be based to a great extent on misunderstandings, for which, though, he accepts a large share of responsibility.

A more basic objection, however – linked to that concerning the division of society into two starkly different realms – concerns the even sharper dichotomy at the level of action types: an action has to be either (a) oriented to success and causally influential, or (b) oriented to norms and values and communicative, i.e. oriented to reaching understanding. An action that is oriented at once to success and to reaching understanding is inconceivable. Habermas’s new typology – despite its increased sophistication over the early one – is equally dualistic, also categorially excluding the possibility of pursuing success (in attaining agent-relative, or “egocentric” goals) in a communicative manner, i.e. by means of rationally motivated agreement with another agent. This exclusion can be traced to Habermas’s identification of “success” with “causal influence”. Characteristically, in describing the distinction between communicative and purposive-rational action, he uses, on the same page, “success” and “causal influence” interchangeably:

- (a) Social actions can be distinguished according to whether the participants adopt either a success-oriented attitude or one oriented to reaching understanding (Habermas, 1984: 286).

- (b) distinguish situations in which [competent speakers] are causally exerting an influence *upon* others from those in which they are coming to an understanding *with* them (*ibid.*).

If this identification between success and causal influence does indeed underly Habermas's model of action, the only way of pursuing egocentric goals which is coherently available to that model is through strategic action. Hence, fields of social life primarily involving the pursuit of (individual or collective) agent-relative goals – arguably, areas such as the economy and politics – cannot *in principle* be conceived by Habermas as areas governed by communicative action.

This question of a possible conceptual gap with serious social-theoretic implications for Habermas, will be the focus of the final Part of this study.

2.4 HABERMAS'S MATURE MODEL OF PRACTICAL RATIONALITY

Habermas's post-communicative-turn views on the different forms of practical rationality are directly based on the universal-pragmatically reconstructed concept of communicative rationality and its threefold internal differentiation that corresponds to the three types of validity claim. These views are initially presented briefly in *The Theory of Communicative Action* (see, for example, 1984: 333–4), and in certain respects even earlier, mainly in *Legitimation Crisis* (1976: 102–110); but they become the focus of attention and are given full-scale treatment in later works, above all in *Moral Consciousness and Communicative Action* (1990) and *Justification and Application* (1993), and in certain respects in such works as in *Between Facts and Norms* (1996) and *The Inclusion of the Other* (1998b).

In these later, more definitive formulations, Habermas distinguishes three forms of Practical Rationality, or three “Employments of Practical Reason”: the “pragmatic”, the “ethical” and the “moral (Habermas, 1993: 1). Each of these forms – presented

below – approaches the question “what should I (or we) do” in a different way, and involves a different relationship between reason and volition.

In addition, in some of these later works, above all in *Between Facts and Norms*, Habermas develops the concept of *legal rationality*, based on a combination of the above three forms of practical rationality with what Habermas calls “fair bargaining”, a form of interaction which he considers to be strategic rather than communicative. Legal rationality, and in particular the notion of “fair bargaining”, will be discussed in later sections.

2.4.1 The Pragmatic Employment of Practical Rationality

As used in this context, the term “pragmatic” must not be confused with its usage in connection with philosophical pragmatism or with the (universal) pragmatics of language. It is rather used in the sense of everyday pragmatism, referring to expedience and effectiveness. More precisely, it is used in the sense of Habermas’s terms “purposive” (or “purposive-rational”), “instrumental” and “strategic”, as well of his earlier term “technical” (as opposed to “praxial”). It refers to relations between means and ends, which for Habermas necessarily involve causal, subject-object relations.

Thus, the pragmatic form of practical rationality deals with the question “What should I/we do?” with a purposive orientation.

Practical reflection here proceeds within the horizon of purposive rationality, its goal being to discover appropriate techniques, strategies or programmes (Habermas, 1993: 3).

These techniques, strategies or programmes are rationally chosen as means to “fixed purposes” which are themselves capable of being assessed in a pragmatically-rational way in terms of existing value preferences. The end-results of processes of pragmatic practical rationality, in other words, are recommendations expressible as what Kant referred to as “hypothetical imperatives” (see Appendix II, below); i.e. as “task[s] that we *must* accomplish if we *want* to achieve a certain goal” (Habermas, 1993: 2). Hypothetical recommendations are concerned with “how we must [causally] intervene in the objective world in order to bring about a desired state of affairs” (*ibid.*: 8). They are based on empirical knowledge about the objective world in which the intervention is made, and their validity depends on – and is judged according to – the truth of that knowledge. Therefore, the pragmatic employment of practical

rationality relies on theoretical rationality, i.e. on discourses concerning factual truth, and is thus universal-pragmatically grounded by Habermas in terms of the constative aspect of linguistic communication and the validity claim to factual truth implicit in it.

The discussion of Habermas's views on (factual) truth is beyond the scope of this study; certain clarifications on this issue, however, would seem to be in order. From the early 1970's up until recently Habermas (under the influence of Peirce and in essential agreement with Apel and Hilary Putnam) conceived of truth as "ideal ascertainability" or "rational acceptability", a view known initially as the "consensus theory of truth" (see Habermas, 1996: 13–15, and 2003: 36). This is a discursive concept of truth, according to which a proposition is true if it is justifiable in a discourse held under ideal speech conditions. In the last decade, however, Habermas has come to revise this view, due to the realization that truth is "a property of propositions that they 'cannot lose' " (2003b: 38) (i.e., if a proposition is true it will always be true); whereas a proposition accepted as true even on the basis of the best of reasons and as a result of the most exhaustive argumentation process that has taken place under conditions as close to ideal ones as possible, could always turn out to be false in the light of new evidence. In other words, the discursive concept of truth does not do justice either to the communication-transcendent reference to an objective, external world, implicitly made by every speech act, or to the absolute certainty with which we deal with what, in our everyday activities, we undoubtedly take to be a common objective world, on the basis of non-thematized factual beliefs non-reflectively accepted as true:

We don't walk onto a bridge whose stability we doubt. To the realism of everyday practice, there corresponds a concept of unconditional truth, truth that is not epistemically indexed (*ibid.*: 39).

But unlike in everyday non-reflective practice, where we take the acceptability of a factual belief to be identical with the belief's truth with respect to objective reality, once we adopt a reflective, discursive mode of thinking and communicating, we know that no discursive justifiability can ever be identical to truth in the sense of a discourse-transcendent relation to the objective world: "the gap between rational acceptability and truth cannot be bridged" (*ibid.*: 40).

Nonetheless, agents reflectively seeking to test claims to truth still have to rely on rational discourse; for there is no direct access to the objects truth claims refer to, whereby the gap between accepted beliefs and truth would disappear, and no way exists that can reduce the (inevitable) gap as much as rational discourse.

Argumentation remains the only *available* medium of ascertaining truth since truth claims that have been problematized cannot be tested in any other way. There is no unmediated, discursively unfiltered access to the truth conditions of empirical beliefs (*ibid.*: 38).

Thus, the revision in Habermas's theory of truth concerned only what truth *means*, not how it is *ascertained*. The way to redeem truth claims pertaining either to pure theory or to questions of pragmatic practical rationality remains the respective discursive process established in the context of the Universal Pragmatics of linguistic communication.

The final point on the pragmatic dimension of practical rationality concerns the relationship between reason and volition that corresponds to that dimension.

It is clear that the (technical or strategic) recommendations of pragmatic discourses are not binding on the will, for they are conditional upon the will's commitment to the ends for whose sake the recommendations are made. And the

making of this volitional commitment – the choice of ends – is, in this conception, external to pragmatic discourses. Thus, for Habermas, pragmatic discourses have no determining role with regard to volition. In fact their function is to serve a volition that is entirely independent of them. They are the “slaves of passion”, in the words of Hume, to whose instrumentalist conception of practical rationality (see Appendix I below) Habermas’s understanding of practical reason’s pragmatic dimension corresponds.

In the next Part we shall question the thesis of the independence of volition from pragmatic practical rationality, arguing that the choice of ends and values is *conditioned* by pragmatic considerations, that is, considerations concerning the available means to their realization.

2.4.2 The Ethical Employment of Practical Rationality

When the goals of action, rather than the means to their realization, cease to be taken for granted and are called into question, pragmatic rationality is obviously not appropriate for dealing with the problem. For now the issue is not how to devise effective means to given ends, but how to select the ends themselves. The question “what should I/we do”, now means what ends/values I/we should choose for myself/ourselves, thus, what is good for me/us. Practical rationality applied to this question is for Habermas *ethical rationality*.

The exercise of practical reason directed... to the good and not merely to the possible and expedient belongs, following classical usage, to the sphere of ethics (Habermas, 1993: 4).

The ethical employment of practical rationality, or “ethical discourse”, is not about the arbitrary choice of specific goals. It is concerned with placing specific goals in

the context of a subject's broader choices and interests, such as one's preferred career, and those in turn, in the context of the subject's conception of the good life – itself seen in the light of the subject's culturally bound life history. In their more radical forms, ethical discourses deal with questions concerning one's self-identity and ego-ideals: "who am I, and who would I like to be?" (*ibid.*). For this reason, Habermas refers to ethical discourses also as *ethical-existential* discourses.

Thus, ethical discourses are hermeneutic processes involving a reflective appropriation of an – individual or collective – subject's cultural traditions and historical development and leading to the subject's self-understanding and self-clarification.

Such processes become critical and self-critical insofar as they dissolve a subject's illusions and self-deceptions and improve a subject's chances of leading an *authentic* and happy life. Habermas often speaks in this respect of a *clinical* or *therapeutic* use of rationality.

Evidently, in an ethical discourse, the subject's perspective is not that of an observer, who value-neutrally serves one's interests and self-identity. It is rather that of a subject (evaluatively) *adopting* interests and a conception of the self as one's own preferred ones: "... existential self-understanding is evaluative in its core" (*ibid.*). Thus, unlike pragmatic rationality, in ethical discourse the relationship between volition and rationality is an internal one: the rational process of ethical self-understanding cannot be conceptually divorced from evaluation, and thus from volition, anymore than ethical evaluation/volition – though dependent on cultural context – can be conceptually divorced from the rational process of ethical self-understanding: "In ethical-existential discourses, reason and the will condition one another reciprocally, though the latter remains embedded in the life-historical context thematized" (*ibid.*: 12).

Habermas's ethical dimension of practical rationality can be said to correspond in orientation and scope, though not in structure, with the Aristotelian conception of practical reason as *phronesis* – a faculty of judgement equally context-bound and oriented to self-realization and the good life.

Ethical discourses (and sometimes also aesthetic critique – see Habermas, 1984: 334) are taken by Habermas to be philosophically rooted in the Universal Pragmatics of the expressive dimension of linguistic communication – of linguistic communication expressing the subjective world of the speaker – thus in the inescapable communicative fact of issuing, and promising to argumentatively redeem, validity claims to truthfulness. This Universal Pragmatic grounding of ethical rationality raises a number of questions:

1) When Habermas speaks of “truthfulness” in the context of developing his Universal Pragmatics, he seems to refer to a correspondence between what speakers know their subjective states to be (having privileged access to them) and what they claim them to be through their utterances. He seems, in other words, to be referring to sincerity, or honesty, to whether the speaker tells the truth to the hearer(s). Ethical discourse, by contrast, emphasizes subjects' coming to a genuine understanding/determination of their subjective states – values, interests, conceptions of the good, ego-ideals, etc. Thus, the truthfulness with which ethical discourses are concerned is *authenticity* rather than sincerity vis a vis others; absence of self-deception rather than of deception of others. If this is so, there seems to be a problem with regard to grounding the ethical employment of practical rationality. How can a discursive practice concerned with authenticity be justified in terms of the rationally reconstructible validity claim to sincerity?

This difficulty could be resolved if the universal validity claim to truthfulness were interpreted as referring not simply to sincerity but to correspondence between what

speakers, through their utterances, express as their subjective states, and what their *genuine* subjective states are, even if these are not immediately accessible to the speaker's consciousness but need to be identified through self-reflection. In fact, on closer inspection, this is how Habermas does understand "truthfulness", his frequent, misleading emphasis on mere sincerity notwithstanding: as the sincere expression of one's authentic subjective states, as opposed to the sincere expression of the subjective states one is conscious of at any given moment, even if as a result of self-deception or insufficient self-knowledge. This understanding is apparent, for example, in the following statement: "Dramaturgical actions embody a knowledge of the agent's own subjectivity. These expressions can be criticized as untruthful, that is, rejected as deceptions or self-deceptions" (Habermas, 1984: 334). Thus, expressions of one's subjectivity (in some formulations, Habermas refers to them as "dramaturgical actions") are truthful if they are neither deceptive nor self-deceptive, that is if they are both sincere and authentic.

2) Like all types of validity claims, validity claims to truthfulness (in the sense of both sincerity and authenticity) ought to be, for Habermas, redeemable by argumentative means. That is, they ought to be settled through universal rational agreement. For,

the fundamental intuition connected with argumentation can best be characterized from the process perspective by the intention of convincing a universal audience and gaining general assent for an utterance; from the procedural perspective, by the intention of ending a dispute about hypothetical validity claims with a *rationally motivated agreement* (Habermas, 1984: 26).

Yet, Habermas maintains that subjects have *privileged access* to their inner, subjective states (see, for example, *ibid.*: 86), and considers ethical discourses as processes, of *self-description*, *self-understanding* and *self-clarification* (see Habermas, 1993: 4-5 and 11, emphasis added).

Surely, on the face of it, there seems to be an internal tension in the idea of one's subjective experiences, and idea of the self, requiring validation in terms of universal consensus. In fact in some important parts of his work Habermas differentiates between validity claims to truthfulness on the one hand, and validity claims to truth, rightness (and comprehensibility) on the other, arguing that only the latter three are "universal validity claims that can be tested in discourse" (Habermas, 1984: 42). However, on closer look, his own arguments do not permit such a sharp differentiation. Claims to sincerity and authenticity, though certainly different from claims to truth and normative rightness in some important respects, do also call for, and are amenable to, validation through the use of reasons.

To begin with, we must distinguish between claims to the truthfulness of utterances expressing one's subjective states, such as ethical choices, and claims to the validity of the content of one's subjective states, e.g. claims to the correctness of one's ethical choices themselves. The latter types of claim are often made in utterances, but are so made contingently; they are not among the rationally reconstructible, ineluctable validity claims which are universally present in all speech acts. And it is argumentation concerning such claims – claims "to justify important value decisions and to gain assurance concerning *his* [the subject's] identity" (Habermas, 1993: 11) – that Habermas has in mind when he confines "ethical-existential discourses" to contextual forms of argumentation – rather than universal ones as in the case of truth and rightness discourses – in which only "[t]hose who belong to a shared lifeworld are potential participants who can

assume the catalyzing role of impartial critics in processes of self-clarification” (ibid.). In such contextual forms of argumentation participants aim at justifying particular (ethical) value judgements in terms of value standards acceptable within their own cultural context: “[V]alues can be made plausible only in the context of a particular form of life” (Habermas, 1984: 42). The value judgements justified in this way are binding not only for the subjects who make them initially but also for those who share the subject’s “form of life” and are (potential) participants in the discourses for the contextual justification of those judgements.

However, it can be argued – although it isn’t by Habermas – that universal discourses, with the (potential) participation of all human beings, also have a role in the *contextual* justification of value judgements. On the assumption that contextual value judgements and contextual reasons and arguments relevant for their justification are *comprehensible* to (though not necessarily acceptable by) all subjects who do not share the particular cultural context in question – which is a reasonable assumption to make – all human beings can have (in principle) a role (indeed they *must* have a role) in deciding whether a contextual value judgement is justified within its own context. Such a judgement would thus be justified in terms of universal consensus but binding only for those who share the form of life in which it is made.

Moreover, it can be said – and this is a point Habermas does make – that, although rational discussions on value judgements cannot lead to *rationally* binding conclusions for those who do not share the context of those judgements, can nonetheless have an *illuminative* effect on them, as in the case of aesthetic critique:

The situation with discussions of value standards, for which aesthetic criticism provides a model, is somewhat different.

Even in disputes about questions of taste, we rely upon the rationally motivating force of the better argument, although a dispute of this kind diverges in a characteristic way from controversies concerning questions of truth and justice... [T]he peculiar role of arguments in this case is to open the eyes of participants, that is, to *lead* them to an authenticating aesthetic experience (Habermas, 1984: 42).

When their eyes are opened in this way, that person may adopt the other's evaluation as his or her own. This outcome – that is, their coming to be committed to an evaluation held and proposed by another and been the object of arguments – is the result of the relevant arguments, but not their rational conclusion.

But let us now come to (universal) validity claims to truthfulness proper – that is claims not to the justifiability of value judgements but to the sincerity and authenticity of expressions of evaluations and other subjective states. Questions concerning the truthfulness of such expressions – unlike those concerning the justifiability of value judgements – are amenable to discursive processes carried out with (potential) universal participation and aiming at universal consensus. In particular, reasons and arguments are considered by Habermas to be pertinent to claims to truthfulness in two ways:

Firstly, the sincerity, or honesty, of an expression of a subject's inner world can be rationally tested – in potentially universal discourse – in terms of its consistency with the subject's behaviour: “[I]nsincerity can be *revealed* by the lack of consistency between an utterance and the past or future actions internally connected with it” (*ibid.*: 41). Habermas tends to differentiate considerations regarding the consistency between expressions of subjective states and actions as rational grounds, from the reasons that are appealed to for the justification of

factual propositions or normative statements: “In the case of claims to truth or rightness, the speaker can redeem his guarantee discursively, that is, by adducing reasons; in the case of claims to truthfulness he does so through consistent behaviour” (Habermas, 1990: 59). However, the facts regarding the consistency between expressions and actions are just as universally applicable reasons in argumentation as facts regarding the objective world. The difference between the subject of expressive utterances and third parties with regard to their role in the discursive testing of the sincerity of those utterances is this: whereas third parties have an equal say as the subject of the expressive utterance in arguing about the sincerity or insincerity of that utterance – that is about the utterance's consistency with the subject's actions – only subjects themselves can put forward for discursive testing expressive utterances representing their own subjective states. In other words, expressive utterances can in principle be accepted as valid or rejected as invalid in universal discourses, but if such an utterance is thus rejected, others cannot without the consent of the subject propose – let alone consensually decide on – an alternative subjective state as valid for the subject. This seems to be the core meaning of the intuitively plausible idea that subjects have a privileged role in the discursive testing of utterances that claim to express their own subjective states.

Secondly, the validity of a claim to the authenticity of a subjective state depends on whether the subject is free from illusion, including self-deception. In particular, authenticity is tested through securing full awareness on the part of subjects of their context-bound life histories – the histories of their cultural development in the case of collective subjects – within which their present subjective states (desires, feelings, beliefs, evaluative choices) have emerged. Once a subjective state is – through such awareness – freed from all illusion and has survived, it is *ipso facto* recognized as authentic.

Now, provided that all potential participants, including the subject, are capable of being motivated only by the force of the better argument, claims to authenticity can in principle be examined in universal discourses, in terms of (public) reasons and arguments concerning facts in the subject's life history. And this, despite the fact that such a discursive process is essentially a process of self-clarification and self-understanding for the subject; a process at the end of which – when all the discursive efforts at eliminating the subject's possible illusions are exhausted – it is, again the subject's privilege to determine the authentic contents of his/her/their subjective world.

However, it cannot be assumed that the subject, even when intent on following the better argument just like all the other participants in the discourse, is always capable of doing so. As Habermas makes clear, the subject can be motivated by unconscious forces, in which case, instead of a rational hermeneutic process, we have a process of "systematically distorted communication". This situation, according to Habermas, calls for self-reflective therapeutic dialogue, or "critique", on the depth-hermeneutic, psychoanalytic model (Habermas, 1984: 41–2). This form of dialogue aims to restore the subject's capacity for non-systematically-distorted discourse, through helping the subject recognize the influence of hitherto unconscious causes on his/her present subjective states and thus behaviours.⁶ Yet, Habermas admits that, even in such a non-fully-discursive dialogue, argumentation plays a role: "[I]n the psychoanalytic view, the healing power of analytic dialogue owes *something* to the convincing force of the arguments employed in it" (1984: 41–2).

Nonetheless, he largely abandons depth-hermeneutic self-reflection, together with the related concept of ideology critique, as an integral part of his post-linguistic-turn paradigm, emphasizing instead ordinary "hermeneutic self-clarification"

(Habermas, 1993: 5).⁷ He does so mainly due to the danger of authoritarianism which is inherent in the asymmetry between the psychoanalytic patient and therapist roles (see Part One, above), and more generally because the depth-hermeneutical model does not fulfill the necessary presuppositions of discourse.

3) The validity claim to truthfulness, on which ethical discourses are based, not only is not a claim to the universal validity of any ethical judgements or beliefs expressed, but it is not a claim that such judgements or beliefs *deserve universal respect, or recognition* either. It is simply a claim that it should be universally agreed that the utterance in question sincerely expresses certain aspects of the speaker's inner world and, furthermore, that those aspects are authentically – i.e. without illusion – the speaker's "own" states. Yet, Habermas clearly advocates the equal respect of all individual and collective ethical choices:

the universalism of equal respect for all is first put to the test by radical freedom in the choice of individual life histories and particular forms of life (Habermas, 1993: 15).

The Universal-Pragmatic grounding of the equal respect of ethical perspectives, therefore, (since it cannot be effected via the validity claim to truthfulness, nor, of course, via the validity claim to factual truth) must be established by Habermas in terms of the validity claim to rightness. We shall look into this question below, in the context of the discussion of the moral aspect of practical rationality.

4) Habermas speaks of individuals as well as groups as being possible ethical subjects, i.e. as subjects of values, identities, conceptions of the good. At the same time he is concerned that social collectivities are not treated as mere macro-subjects, insisting that we must "do justice to the intersubjective nature of collective will formation, which cannot be correctly construed as individual will formation writ

large” (Habermas, 1993: 16). This concern, in fact, has been one of the reasons why he abandoned his early paradigm – which he considered as being based on the subjectivist philosophy of consciousness – for the intersubjectivist, language-based paradigm.

Consequently, Habermas needs to explain how collective ethical discourses derive collective ethical choices from the goals, values, identities and conceptions of the good of individuals. There is, of course, no difficulty in that with regard to individual ethical choices which coincide with each other, or can be brought into line with each other through hermeneutic dialogue – which is largely the case, for groups whose members share the same cultural tradition, in which their individual ethical choices are rooted. It is, however, highly challenging when the sets of values, identities and conceptions of the good of the individuals comprising a collectivity differ, and much more so when they conflict with each other. And it must be said that this is a challenge which Habermas does not meet. For in order to communicatively synthesize, or aggregate the diverse sets of ethical choices of several individual subjects into a single, composite set of ethical choices adopted by the whole group collectively, there must be, firstly recognition by each individual subject of the (authentic) ethical choices of all other individuals involved, and secondly, a communicative method for synthesizing the mutually respected choices, both recognition and method being justified in terms of Universal Pragmatics. We have already pointed out that mutual recognition cannot be grounded in terms of the validity claim to truthfulness, which forms the basis of ethical discourse, and that the question of its grounding needs to be examined in connection with the only remaining relevant aspect of practical rationality, the moral aspect. On the other hand, Habermas neither provides nor Universal-Pragmatically justifies in the context of ethical discourse, any communicative procedure for aggregating diverse values, identities and conceptions of the good.

Compromise as a result of bargaining, which is increasingly relied upon by Habermas (especially in *Between Facts and Norms*) as a way of dealing with conflicting goals, is seen by him as a strategic, not a communicative, approach, as we shall see in the following Part (section 3.3.1).

2.4.3 The Moral Employment of Practical Rationality

Here we are dealing with the question “what should I/we do?” in the deontological, *moral* sense of what I/we *ought* to do, or what is my/our *duty* to do, unconditionally (Habermas, 1993: 8). Whereas the pragmatic and ethical employments of practical rationality are oriented to the purposive and the good, respectively, the moral employment of practical rationality is oriented to what is universally right, or just, or equally in the interest of everyone. The shift from the perspective of “what is good for me” to the moral attitude of “what is good for all” occurs, according to Habermas, “as soon as my actions affect the interests of others and lead to conflicts that should be regulated in an impartial manner” (*ibid.*: 5). The impartial resolution of conflicts of interests is contrasted to their settlement through a strategic approach, where “the participants assume that each decides egocentrically in accordance with his interests” (*ibid.*: 6).

The moral employment of practical rationality primarily takes the form of discourses oriented to the justification of moral norms, or maxims. Moral norms are understood, in line with the Kantian, deontological tradition, as commands or “ought” statements, applicable universally, i.e. in all (relevant) situations, regardless of social, historical or other context, and for all potential addressees, that is, all human beings, or – even more broadly – for all beings capable of speech and action.⁸

Moral norms prescribe duties, or obligations, either of a positive kind (Do X) or of a negative kind, that is prohibitions (Do not do Y). Often one and the same norm can take both a positive and a negative form (e.g. “You must always keep your promises”, or “Never break your promises”). Furthermore, one’s moral duty implies a corresponding moral right of others. My duty not to lie corresponds to others’ right not to be lied to by me. Unlike Kant, however, Habermas does not consider the criterion “whether I can will that a maxim should be followed by everyone as a general law” as the test for valid moral norms (*ibid.*: 7). For such a test – such a formulation of the *principle of universalization* – “remains bound to the personal perspective of a particular individual” (*ibid.*), and could, counter-intuitively, justify as moral, for example, “a casual attitude towards deception”, if only one personally happened not to mind others having a similar attitude in comparable situations (*ibid.*).

Habermas’s *Universalization Principle* is formulated as follows:

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 (and these consequences are preferred to those of known alternative possibilities for regulation) (Habermas, 1990: 65).

Habermas stresses that this universalization principle must not be interpreted as a test that can be applied to norms *monologically*. It is not up to individuals to judge, by putting themselves into the position of others affected, whether everyone else would approve the norm in question. Rather, the decision must be taken, in a cooperative manner, by *all* concerned, a point encapsulated in what Habermas calls the *Discourse Principle*:

Only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity as *participants in a practical discourse* (*ibid.*: 66).

Thus, contrary to those whom he considers “monological” moral theorists, Habermas denies that moral philosophers can go beyond clarifying moral concepts and the nature of moral discourse, to passing (philosophically authoritative) judgements on substantive moral norms. Such norms can be determined only at the level of universal moral discourses, in which of course philosophers must take part as ordinary participants, and, perhaps, also play the role of process experts (see Habermas, 1990: 94).⁹ His version of the universalization principle, Habermas maintains, captures more adequately than rival ideas (such as those of R. M. Hare, Kurt Baier, Bernard Gert, Marcus Singer and Rawls) important common intuitions regarding the nature of morality, as distinct from ethics, above all those of “impartiality” and “equality of treatment” (*ibid.*: 64–66; Habermas, 1990: 43). As already implied in the preceding discussion, moral discourses in which U is applied are argumentative deliberations aiming not to *produce* universal moral norms but to *test* moral norms whose validity has been questioned.

Following this initial presentation of some central tenets of Habermas’s conception of moral rationality, a number of additional issues need to be discussed.

1) The process of testing moral norms, through the application of the Principle of Universalization takes place in abstraction (i.e., it is detached) from any particular concrete situation or context; and this is what, for Habermas, secures their universal status: “Valid norms owe their abstract universality to the fact that they withstand the universalization test only in a decontextualized form” (Habermas, 1993: 13). Participants in such moral discourses apply the Principle – i.e. examine whether they would accept the consequences of the general observance of the

norm under consideration – by envisaging as wide a variety of possible applications of the norms being tested as they can. However, the perspective of such participants cannot but be historically situated and finite, making it impossible to take into consideration the relevant features of all the possible concrete situations that might present themselves to moral agents, especially future ones. Thus, any set of valid moral norms – i.e. norms that have passed the test of universalization in the above sense – might not provide sufficient guidance to a moral agent confronted with a new concrete situation. An agent, for example, being asked by the ill-intentioned pursuers of an innocent person to give information (that the agent possesses) about the whereabouts of the person being pursued, might not be able to derive sufficient guidance as to the morally right thing to do, from the two arguably relevant moral norms, (1) “Do not harm innocent persons” and (2) “Do not tell lies”. For it is unlikely that the processes of discursive validation of the two norms could have endowed these norms with sufficient qualifications regarding their conditions of application so that they might fully cover the very specific conditions in which the agent in question is called upon to apply the norms. In such a situation, how could the agent decide what to do? Which of the two norms is more appropriate in the given situation? Is it morally right for the agent to lie to the pursuers, making sure that no harm would be done to an innocent person, or to tell the truth, and as a result cause such harm?

In addition to cognitive, decision-making deficiencies, decontextualized, universalist moral discourses are also responsible for a motivational deficit¹⁰ (and this, despite the fact that, unlike Kant’s universalization principle that totally divorces duty from inclination, Habermas’s principle incorporates interests into moral norms). For such discourses involve “the uncoupling of moral judgement from the concrete motives that inform action” (*ibid.*: 14), thereby depriving, so to

speak, the moral will of the (strong) motivating force of interests and leaving it only with the (weaker) force of reason:

under a strict deontological point of view, moral questions are taken out of their contexts in such a way that moral solutions retain only the rationally motivating force of insights (Habermas, 1990: 178).

Thus, even if one knows what, according to universalist morality, the right thing to do is, one may not “summon the resolve” to act accordingly. It is Habermas’s view that unlike the imperative “*should*” in its pragmatic and ethical senses, which have, respectively, an external (hypothetical) or reciprocal relation to the will, the moral “*should*” – the unconditional, categorical ought – has an obligatory function: “the rightness of moral commands... is intended to bind the will rationally from within” (ibid.: 15).

Sometimes the will obeys the imperatives of moral rationality, and in these cases Habermas, following Kant, characterizes it as “autonomous” will. But often volition does not adhere to rational moral imperatives. This happens when the “rationally motivating force of insights”, or the “motivational force of goods reasons” is outweighed by other, interest-oriented motives, concretely operating in specific situations (ibid.: 10). In such cases we speak of “weakness of will”.

Habermas rejects neo-Aristotelian responses to the above cognitive and motivational difficulties of universalist, deontological morality (See Habermas, 1993: 113–132). Neo-Aristotelians such as Alastair MacIntyre, Bernard Williams, Charles Taylor and others, typically reject the sharp separation between norm-based morality and evaluative ethics and propose tackling concrete ethical/moral decision-making challenges in terms of some form of context-dependent practical

deliberation – akin to Aristotle's *phronesis* – restricted within the bounds of a particular form of life. Such deliberation, being carried out by agents socialized in, and belonging to the form of life in question, is claimed to possess both the situational cognitive sensitivity and the motivational force necessary for good practical judgement and corresponding action.

Habermas's central objections to this approach run as follows: Aristotle's *phronesis* is metaphysically underpinned by a teleological theory of human nature which privileges life in a Greek *polis*. Modern philosophers, by contrast, no longer have at their disposal the metaphysical tools with which to establish a hierarchy among the vast plurality of contemporary forms of life, and thus to provide support for any particular context-based set of practical judgements. Consequently, they are left with two options. The first one is to reduce practical reason to the clarification of the parochial ethical/moral common sense of one's own culture. This relativistic interpretation, with its unmistakable conservative overtones, would be objectionable to most neo-Aristotelians, who would like to reserve for practical reason a socially critical role. The second option is to open up the boundaries of particular cultural life-forms from within and extend the application of practical reason to what is common to all life-forms. But this would not be far from the Kantian-type narrowly moral point of view, including Habermas's own.

The horizon of every form of life is fluid, its boundaries permeable... If each individual community can achieve "knowledge"... concerning what is good for it, it is far from obvious why this practical knowledge should not be extended in an intercultural direction and become so thoroughly emancipated from provincial limitations that it orients itself to what is *equally good for all*. Without

metaphysical backing, what Aristotle called *phronesis* must either dissolve into mere common sense or be developed into a concept of practical reason that satisfies the criteria of procedural rationality (Habermas, 1993: 124–5).

Hence, Habermas insists that the above mentioned cognitive and motivational problems facing universal norm-based morality should not lead us to give up either the specifically modern distinction between such a conception of morality and evaluative ethics, or the effort to develop an adequate form of non-contextual rationality for the former. His suggestion for overcoming the cognitive problem of situational moral decision making is to supplement *Discourses of Justification* – that is the discursive processes concerned with the validation of universal moral norms – with *Discourses of Application*. The latter would be concerned not with the testing of norms but with the application, in specific situations, of already validated norms. Though situation-sensitive and requiring some form of “hermeneutic prudence”, such discourses would not be contextual in a relativist sense. They would “rely on reasons that are in principle valid for everyone” (Habermas, 1993: 172).

In his initial major statement of his discourse theory of morality, in *Moral Consciousness and Communicative Action*, Habermas refers to “the mature capacity for moral judgement” as a faculty necessary for both justifying and applying norms. This faculty is said to be characterized by a combination of “cognition, empathy and *agape*” (Habermas, 1990: 182). In his important subsequent elaborations of the theory, however, in *Justification and Application* a few years later, he adopts for the discourses of application the more precise model developed by Klaus Günther in *The Sense of Appropriateness* (1993). Whereas discourses of justification are based on the principle of universalization, discourses

of application, according to this model, rely on the *principle of appropriateness*. Together, the two principles provide the foundations for a fully rational process of moral decision making:

In discourses of application, the principle of appropriateness takes on the role played by the principle of universalization in justificatory discourses. Only the two principles taken together exhaust the idea of impartiality (Habermas, 1993: 37).

The principle of appropriateness is said to specify a method for impartially deciding which of the *prima facie* valid but conflicting moral norms which are found to be applicable in a given situation (e.g. “Do not tell lies” or “Do not harm innocent persons”, in our earlier example) is the appropriate one to apply in that particular situation. This method is based on what Günther calls “the criterion of coherence” (Günther, 1993: 243). Here we shall not discuss this criterion of appropriateness, or more generally expand any further on the issue of discourses of application.¹¹

The resolution of the cognitive problem of moral norm application leaves, of course, intact the motivational deficit of universalist morality. In this respect, Habermas— in line with Hegel as well as the neo-Aristotelians — recognizes the dependence of moral action on ethical life. Rational moral insight does carry a certain motivational force, evidence of which are “the pangs of conscience that plague us when we act against our better judgement” (Habermas, 1993: 14). This force, however is weak. It needs to be supplemented with social (cultural as well as institutional) motivational force. It is thus no accident, for Habermas, that universalist morality historically arose in conditions of modernity; For it is only under such conditions — with their cultivation and institutional support of postconventional forms of consciousness — that it can be effective in practice.

Any universalistic morality is dependent upon a form of life that *meets it halfway*. There has to be a modicum of congruence between morality and the practices of socialization and education... Morality thrives only in an environment in which postconventional ideas about law and morality have already been institutionalized to a certain extent (Habermas, 1990: 207–8; see also *ibid.*: 108–109).

2) Application issues notwithstanding, the Universalization Principle remains, for Habermasian norm-based discourse theory of morality, the central rule of moral argumentation. As such, of course, it itself stands in need of justification. Habermas's strategy, in this regard is to ground the principle in the structure of communicative rationality, thus, in the last analysis, in the universal pragmatics of speech oriented to reaching understanding.

[T]he principle of universalization, which acts as a rule of argumentation, is implied by the presuppositions of argumentation in general... 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 implicitly presupposes as valid the principle of universalization, whether in the form I gave it above or in an equivalent form (Habermas, 1990: 86).

Among the necessary presuppositions of argumentation referred to by Habermas, particularly relevant to morality are those concerning (a) the exclusion of "all external or internal coercion other than the force of the better argument", (b) "relations of mutual recognition" and (c) participants' equal opportunity to

contribute to the argumentation and to put forth their own arguments”, especially “to express [their] attitudes, desires and needs” (*ibid.*: 88, 89).¹² Thus, the universal-pragmatically established presuppositions of argumentation include the recognition and equal treatment of every participant’s interests. On this basis Habermas argues that the universalization principle is derivable from the structure of communicative rationality and action. In an alternative formulation of this point to the one quoted above he traces the universalization principle specifically to normative argumentation (as distinct from argumentation concerning truth or truthfulness):

[E]veryone who seriously tries to *discursively* redeem normative claims to validity intuitively accepts procedural conditions that amount to implicitly acknowledging [the universalization principle] (*ibid.*: 92–3).

The testing of a moral norm by means of the universalization principle is certainly in line with the presuppositions of argumentation, including that of the (recognition and) equal treatment of the interests of those concerned, thus with communicative action as such. However, this does not amount to the derivation from communicative action of the necessity to apply the principle. What can reasonably be argued is that *if* a universal moral norm needs to be argumentatively tested, such a testing – in order to respect presuppositions such as that of the equal treatment of the interests of those concerned, but also in order to deal with the universality of the norms being considered – must be done in terms of the universalization principle. Consequently, the necessary application of the principle would be derivable from, or presupposed by, the structure of communicative action itself, only if the latter necessitated the (discursive) justification of *universal, moral norms*. Equally, to say that the structure of communicative action does not

presuppose the necessary employment of the universalization principle, would mean that it (the structure of communicative action) does not presuppose the need to (discursively) justify moral norms. The two presuppositions – that of moral norms and that of the universalization principle – stand or fall together.

Evidently, Habermas defends both these presuppositions in terms of the idea of the validity claim to rightness. According to that idea, in every speech act there is an implicit claim that the act in question is right with regard to legitimately shared social norms, which, directly or indirectly, include valid *moral* norms. Thus, communicative action is thereby said to imply the need to discursively justify universal moral norms and, by the same token, the need to apply the principle of universalization. In addition, the validity claim to rightness seems to imply that communicative action entails the existence of valid moral norms, i.e., that some of the applications of the universalization principle will yield positive results, that is approve the candidate moral norm being tested.

However, all these three aspects of Habermas's validity claim to rightness are highly questionable. We shall return to this issue in the nextPart (section 3.5), where I shall propose an alternative to the validity claim to rightness and look further into the role of the universalization principle.

3) In the section on the Ethical Employment of Practical Rationality above, we referred the issue of ethical choices made by collective agents to Habermas's treatment of moral rationality. It was pointed out that in order to give a communicative (as opposed to a strategic) account of the aggregation of diverse individual ethical choices into composite ones, Habermas had to provide (a) a universal-pragmatic justification of the recognition and equal treatment of the interests of the individuals concerned and (b) a communicatively rational method of aggregation.

In the light of the preceding discussion of his views on moral rationality, we can maintain that Habermas largely succeeds with regard to requirement (a), but not with regard to (b). Concerning the latter, even if the universalization principle were (plausibly) construed as a method for synthesizing diverse interests, it could not serve the purpose, for that principle can deal only with *norms* of action of the *universal* collective of all thinking and acting beings, not with *substantive* ethical choices (goals, values, etc.) of multifarious, particular, contingently formed groups.

In the following Part (section 3.6.2) we shall return to the pragmatic, ethical and moral/normative forms of practical discourse in order to explore relations of mutual dependence between them.

2.4.4 A conceptual lacuna

In Part One it has been argued that Habermas's early paradigm categorially excludes the possibility of human beings using other human beings as means to their ends in "praxial", or communicative, rather than instrumental-manipulative, ways. An analogous conceptual gap seems to reappear in the later paradigm, as already indicated at the end of the previous section. There, it was pointed out that the dichotomy between success-oriented and communicative action categorially rules out the pursuit of success, in the sense of seeking to attain egocentric goals, in other than strategic ways, that is, communicatively.

The same lacuna is evident, now with regard to the rationality of action, in the universal-pragmatically grounded tri-partite model of practical rationality just considered. In this model, communicative rationality can only deal on the one hand with the adoption and sincere expression of a subject's goals, values, ethical choices in general, and on the other with the testing of moral norms. No place is reserved for determining how to attain the realization of particular goals, values, etc., by using certain other subjects as means, through a communicative discourse

in which those other subjects participate. The model allows only for the “pragmatic”, thus strategic, pursuit of ends by means of others. In such a pursuit other human beings are treated as objects to be causally influenced. On the level of speech act theory, the heart of the problem seems to be the inability to give a universal-pragmatic account of *requests* – that is of utterances through which speakers typically ask addressees to be means to their ends – as speech acts that can conceivably be communicative, i.e. oriented to understanding/agreement, rather than being necessarily strategic.

Thus, utterances such as “Please give me a glass of water” and “Can I have a discount on that laptop?” – insofar as they are addressed to the hearer not as commands but as mere requests, which are not in any way impermissible to make, but whose satisfaction by the hearer is understood to be optional – cannot be validated in terms of any of the forms of discourse that correspond to the three validity claims grounded by Habermas in communicative action. This lack of a communicative interpretation of requests makes it impossible for Habermas to conceive of bargaining as a form of interaction that can possibly be communicative. Consequently, it also rules out the development of a communicative method of aggregating diverse ethical choices into composite ones; for such an aggregation could only be achieved on the basis of some form of compromise-seeking bargaining process.

As we shall see in Part Three (section 3.3), in some later texts Habermas develops an interpretation of requests, or, in his terms, of “naked imperatives”, as utterances which can take the form of somehow inferior, “weak” communicative acts. However as will be shown, this interpretation cannot account for all requests. Habermas also distinguishes between ordinary bargaining and “fair” bargaining, the latter seen as a form of interaction linked to morality. Nonetheless, he considers even that type of bargaining as in the last analysis power-based, strategic action.

2.5. HABERMAS'S SUBSTANTIVE THEORY OF MODERN SOCIETY: LIFEWORLD AND SYSTEM

2.5.1 Main Theses

Habermas distinguishes two forms of societal integration, that is, two ways in which the multiplicity of social actions are coordinated with each other, so that societies hold together and reproduce themselves: *social integration* and *system integration*.

We have drawn a distinction between mechanisms of social integration, which attach to action orientations, and mechanisms of system integration, which reach right through action orientations. In one case the agents' actions are coordinated through a harmonizing of action orientations that is present to them; in the other, through a functional intermeshing of action consequences that remain latent, that is, that can go beyond the participants' horizon of orientation (Habermas, 1987: 202).

Social integration, that is, societal integration in terms of "action orientations" is understood by Habermas as integration effected by means of *communicative* interactions: It is achieved through "normatively secured or communicatively achieved consensus" (*ibid.*: 117); or, by means of "consensus-forming mechanisms (which bring about agreement on values, norms and linguistic communication)" (Habermas, 1991: 252). System integration of society, on the other hand, takes place not in terms of the coordination of action *orientations* but "behind the backs" of social actors, as a result of "a non-normative regulation of individual decisions that extends beyond the actors' consciousness" (1987: 117).

On the basis of the distinction between social and system integration, Habermas draws a corresponding distinction between two possible perspectives on social phenomena, those of *lifeworld* and *system*: To the extent that a society is seen as socially integrated – an internal, participant perspective – it is conceived as “*the lifeworld of a social group*” (1987: 117). To the extent that it is seen as holding together through system integration – an external, observer perspective – society is considered as a self-regulating system; as “*a system of actions* such that each action has a functional significance according to its contribution to the maintenance of the system” (*ibid.*). Social actors themselves neither intend (or anticipate) the systemic effects of their actions, nor do they – in their capacity as actors as opposed to their possible capacity as observers, e.g. social scientists – have an overall view, or even awareness of the system which their own actions, through their consequences, are creating.

The lifeworld is the inter-subjectively shared field of interactions between the members of society. It consists of common traditions and cooperative processes of situation interpretation and action coordination. As such, the lifeworld is internally linked to action oriented to reaching understanding (Habermas, 1987: 119; 1991: 252). The reproduction of the lifeworld takes place through two complementary types of consensus, corresponding, respectively, to preservation and continuity on the one hand and learning and change on the other: The first is *taken-for-granted consensus*, based on the “massive preunderstanding of participants” (Habermas, 1991: 244) who share the same traditions; the second is *communicatively achieved consensus*, based on an orientation to argumentatively testing criticizable validity claims (see Habermas, 1996: 322–4).

In this connection, Habermas speaks of a polarity between taken-for-granted and achieved consensus, and provides an analysis of the historical movement from

the predominance of the former to the predominance of the latter. This movement is described as a “linguistification of the sacred”, a process in which language ceases to be used merely in an instrumental way, as the medium for the ritual application of traditional social-religious (“sacred”) norms, and plays the primary role of questioning and justifying the validity and forms of application of norms (Habermas, 1987: 77ff). Thus, “linguistification”, in this sense, results in the “communicative thawing of traditionally solid institutions” (*ibid.*: 91). Acting social agents always have knowledge of the lifeworld as the order of society created by their own communicatively oriented actions. This knowledge has the form of explicit awareness insofar as that order is the product of their deliberately pursued and achieved consensus, and is intuitive insofar as the lifeworld order is a result of unreflective, taken-for-granted consensus.

A lifeworld is said to be *rationalized* to the extent that it is reproduced by means of communicatively achieved – rather than taken-for-granted – consensus. For to that extent, the potential for rationality which is inherent in the communicative action that lies at the heart of the lifeworld is actualized (Habermas, 1987: 145–6; 1991: 223–4). The rationalization of the lifeworld, therefore, is for Habermas intrinsically linked to the “linguistification of the sacred”. For the latter is precisely conceived as a process of “release of the rationality potential in communicative action” (*ibid.*: 77).

To be sure – and this is a lesson Habermas has learnt from Mead as well as from the phenomenological-hermeneutic tradition, above all Husserl, Schutz and Gadamer – at any given time we can “thematize”, or “linguistify” and argue about only a small segment of the lifeworld; a segment that is thrown into relief as a problematical situation which calls for our communicative attention aiming at an agreed interpretation – a “shared definition” – and coordinated action. A

situation, thematized in this manner, has a horizon beyond which lies the rest of the lifeworld, taken for granted from the perspective of the situation in question. A thematized segment of the lifeworld can expand to include more of what had hitherto been taken for granted, but beyond its new horizon there will always be the remaining taken-for-granted lifeworld “as the horizon within which communicative actions are ‘always already’ moving” (Habermas, 1987: 119). It is thus not possible to thematize (and consequently make transparent) the entire lifeworld – the whole of tradition – at once (Habermas, 1987: 122–135). What can be done in the way of lifeworld rationalization – and can be said to characterize modernity to a considerable degree – is thematized and subject to argumentative deliberation tradition piece by piece, each time (unavoidably) taking the rest for granted:

It is distinctive of the modern understanding of the world that the cultural tradition can be exposed to testing... across its entire spectrum and in a methodical manner (*ibid.*: 133).

In modern societies the rationalization of the lifeworld can be observed in all three of what Habermas, following Parsons, takes to be its constitutive spheres, whose differentiation itself, in his view, has come about in the wake of lifeworld rationalization. These differentiated and yet complementary and mutually supportive spheres, or “structural components”, of the lifeworld are (a) *culture* (meaning essentially society’s stock of knowledge), (b) “*society*” (the term now being used in a narrower sense meaning society’s institutional order) and (c) *personality* (referring basically to subjects’ speaking and acting competences).¹⁵

The sphere of culture is increasingly characterized by a reflexivity of traditions, value-generalization (or “abstraction”) and a professionalization of cultural

transmission, especially in the fields of science, morality and law (jurisprudence), and art and art criticism. In the sphere of "society", there has been a progressive institutionalization of formal, and at the same time democratic, procedures for establishing legal and other principles and norms. Rationalization in the sphere of personality has, most importantly, moved in the directions of creating formal systems of child rearing and education, and developing individuated, autonomous self-identities and personalities equipped with abstract moral and cognitive competences, in contrast to earlier concrete thinking (*ibid.*: 145–8).

Habermas maintains that societies – with the significant exception of modern ones, as we shall see – can be conceived *simultaneously* as both socially integrated lifeworlds and self-regulating systems functionally integrated through the consequences of members' actions (Habermas, 1987: 118, 120).

In non-modern societies (more so in the simpler tribal societies organized around kinship relations than in the more differentiated traditional societies that have a state organization), system integration, although by and large beyond the consciousness of members, is nonetheless largely effected through precisely those members' communicative action orientations, which are at the same time (intentionally) socially integrative. For this reason social scientists (for example, functionalist social anthropologists) who study those societies from the observer perspective are able to trace observed systemic features to specific aspects of the lifeworld. Thus, the two orders are, in a sense, "coextensive". But of course, the reverse process – that of tracing socially integrative lifeworld features to systemic functions – is not possible. In such cases, for Habermas, this gives the lifeworld and the internalist, hermeneutic perspective primacy over the still "parasitic" (*ibid.*: 181) system and systems theory.

In modern societies, however, lifeworld and system cease to be coextensive (and hierarchically related) in the above manner. In the course of social evolution certain domains of social life, namely the market economy and state administration, have developed, which are integrated systemically through the functional interconnections of the consequences not of actions that are *at the same time* communicatively oriented and intended for social integration, and thus constitutive of the lifeworld, but of strategic actions that are merely “steered” by non-symbolic, or “delinguistified”, media of communication, namely money and power, respectively. Thus, in modern societies, the market economy and state administration have not just formed themselves into epistemologically distinct domains of social life amenable to scientific study from the observer perspective of systems theory (in which case they would remain the obverse, coextensive side of a lifeworld simultaneously comprehensible intuitively from the internal perspective of participants). Rather, they have, ontologically, so to speak, virtually disconnected themselves – they have been “uncoupled”, in Habermas’s terminology – from the lifeworld; they are now self-regulated subsystems amenable to analysis exclusively from the counter-intuitive, externalist perspective of systems theory.

The transfer of action coordination from language over to steering media means an uncoupling of interaction from lifeworld contexts. Media such as money and power... encode a purposive-rational attitude toward calculable amounts of value and make it possible to exert generalized, strategic influence on the decisions of other participants while *bypassing* processes of consensus-oriented communication... the lifeworld is no longer needed for the coordination of action (*ibid.*: 183).¹²

From the point of view of everyday lifeworld activity, these subsystems appear increasingly as “something in the objective world”, as “second nature”, towards which social agents can relate only in an instrumental, subject-object manner (*ibid.*: 154, 173).

Habermas accounts for the uncoupling of the two subsystems from the lifeworld in terms of the process of lifeworld rationalization. It will not be necessary, for our present purposes, to go into the intricacies of his social-evolutionary and historical arguments to that effect (see, especially, Habermas, 1987: 153–197). The essential points are as follows:

In modernity, lifeworld rationalization, with the concomitant internal differentiation and complexity of the lifeworld, went so far that it overburdened the lifeworld’s coordination capacity; in other words, it overburdened society’s capacity to reproduce itself as a lifeworld, that is, purely communicatively. Thus, societal coordination and integration at modern levels of lifeworld rationality required relief mechanisms that would reduce the need for explicit communication, simplifying social coordination processes.

Historically, this requirement for relief mechanisms came to be fulfilled with the evolution and uncoupling from the lifeworld of the market economy and state administration as two self-regulated subsystems of strategic action. The core of this development was the legal institutionalization of the non-symbolic or “delinguistified” steering media of money and power, which came to replace (rather than “condense”) communicative action, as the respective integrative mechanisms in these two social action domains. This institutionalization – a process necessary for the uncoupling and at the same time one that kept the uncoupled subsystems “anchored” in the lifeworld, i.e. indirectly linked to communicative action – would be impossible in the absence of important aspects of a moderately rationalized

lifeworld, above all, modern universalist morality and law. Thus, the rationalization of the lifeworld both created the need and provided the necessary means for the modern uncoupling of the economic and administrative subsystems from the lifeworld.

The legal institutions responsible for the institutionalization, uncoupling and anchoring of the subsystems are the conduits through which any influences between the subsystems and the lifeworld are transmitted. In theory, such influences can flow in either direction:

[T]he institutions that anchor steering mechanisms such as power and money in the lifeworld could serve as a channel *either* for the influence of the lifeworld on formally organized domains of action *or*, conversely, for the influence of the system on communicatively structured contexts of action (Habermas, 1987: 185).

In *The Theory of Communicative Action* Habermas emphasizes what he believes to be an excessive influence in the direction from the subsystems to the lifeworld, an influence he characterizes as a *colonization* of the lifeworld and tends to attribute mainly to the expansionist dynamics of capitalism. Colonization occurs when core areas of the lifeworld are “mediatized”, that is, when communicative action, in areas in which it is indispensable for the very reproduction of the lifeworld, is replaced by instrumental/strategic action steered by the non-linguistic media of money and bureaucratic administrative power:

[S]ystemic mechanisms suppress forms of social integration even in those areas where a consensus-dependent coordination of action cannot be replaced,

that is, where the symbolic reproduction of the lifeworld is at stake. In these areas, the *mediatization* of the lifeworld assumes the form of *colonization* (*ibid.*: 196; see also pp. 356, 367).

Prominent examples of colonizing tendencies are the progressive monetarization and/or bureaucratization of such identity- and meaning-producing areas of the lifeworld as family life, education, culture and the political public sphere.

As “meaning can neither be bought nor coerced” (Habermas, 1991: 259), such an encroachment of the subsystems on the lifeworld – one that, as we have seen, by definition undermines the production of meaning in society – has dire social effects. According to Habermas, serious pathologies of modernity diagnosed by major social theorists and encapsulated in such notions as alienation (Marx), anomie (Durkheim) and loss of meaning and freedom (Weber), are the result of system colonization of the lifeworld, and not of lifeworld rationalization itself (expressed, for instance, in the decline of religion and traditional morals) as counter-enlightenment critics of modern culture claim (Habermas, 1987: 148). Of course, it is the rationalization of the lifeworld that initially makes possible the emergence of the subsystems which then turn back to colonize the lifeworld itself. However, though made possible by lifeworld rationalization, colonization is not made inevitable by it. It is rather due to the “independent imperatives” of the two specific subsystems themselves (*ibid.*: 155, 186). Moreover, lifeworld rationalization, while giving rise to forces that lead to the erosion of the lifeworld, at the same time, paradoxically, created the “utopian perspective” from which to criticize that erosion and the capitalist modernization responsible for it (*ibid.*: 357).¹⁵

As mentioned above, the colonizing influences of the subsystems upon the lifeworld are channelled through the legal apparatus that institutionalizes the subsystems. In particular, the vehicle of these influences is juridification, in the specific forms that it has assumed in modern times.

By the term "juridification" Habermas refers to "the tendency toward an increase in formal (or positive, written) law that can be observed in modern society" (ibid.: 357). Habermas's thesis is that in some of its "waves", modern juridification has been an ambivalent process (hence, he speaks of a "dialectic" of juridification) (ibid.: 356–373): while it guaranteed freedoms and rights, at the same time it had destructive effects on the lifeworld existence of social actors. Major examples are the early modern emancipation of wage labour from feudal bonds, which "had to be paid for with the proletarianization of the wage labourers' mode of life" (Habermas, 1987: 361), and the ambivalent character of the democratic welfare state. In the case of social-welfare law, for example, the unquestionably historically progressive "legal entitlements to monetary income in case of illness, old age and the like" are bought at a high price, in the form of their "bureaucratic implementation" (ibid.: 362). Similar ambivalences exist with regard to family and school law, where the undoubtedly desirable legal protection of the rights of the individuals involved (family members, school children, teachers, etc.) from each other's potential interest-based violations as well as from oppressive traditional practices, at the same time makes those very individuals dependent on the state and tends to transform their now formally regulated relationships into depersonalized interactions, in which "[as] legal subjects they encounter one another in an objectivizing, success-oriented attitude" (ibid.: 369).

Whereas in *The Theory of Communicative Action* Habermas one-sidedly emphasizes the (colonizing) influence of the subsystems upon the lifeworld, in

later works, above all in *Between Facts and Norms*, he pays more attention to (actual and potential) influences in the opposite direction. Key role in lifeworld's ability to control the media-steered subsystems and protect itself from their colonizing incursions is, once more, assigned to law. In contrast to his earlier focusing on the negative aspects of juridification, as a vehicle of lifeworld colonization, now Habermas views law more positively. By virtue of two complementary aspects of it, law is understood to have a foothold both in the domain of the lifeworld and in the world of the media-steered subsystems. On the one hand, as lawmaking, that is as a process primarily of generating legitimate legal norms and secondarily deciding how general legal norms should be applied in specific situations, law belongs to the lifeworld, and as such is linked to processes of opinion formation in political life and, more broadly, the public sphere. (Here, a prominent role is today played by new social movements such as those of feminism, anti-racism, environmentalism and peace activism). On the other hand, as law-administration and law-enforcement, supported by the state's legitimate instruments of coercion, law belongs to the realm of the subsystems. Due to this twin capacity, law is viewed "as the medium through which communicative power is translated into administrative power" (Habermas, 1996: 150).

Legitimate lawmaking, that is the process of generating valid – thus justified – legal norms, is conceptualized by Habermas in terms of the "Discourse Principle", which states that,

Just those action norms are valid to which all possibly affected persons could agree as participants in rational discourses (*ibid.*: 107).

This formulation of the Discourse Principle differs significantly from the earlier one, put forward in the context of his discourse theory of morality (Habermas, 1990: 66). Though both formulations “explicat[e] the meaning of impartiality in practical judgements” (1996: 107), with the previous one, as Habermas self-critically maintains, he had “not sufficiently distinguished between the discourse principle and the moral principle” (*ibid.*: 108). The older version was concerned with the rational process of justifying exclusively *moral* norms, that is, norms based on universalizable interests. That process, therefore, (potentially) involved the universal participation of all human beings. The new version of the Discourse Principle, quoted above, is understood to be one concerned with the rational justification of action norms in general, including both moral and *legal* norms. As such, the Principle is “*neutral* with respect to morality and law” (*ibid.*: 107).

When employed for the validation of moral norms, the Discourse Principle becomes a universalization principle. However, when applied to legal norms – i.e. action norms given an institutional, legal form – thus specifying a rational process of legitimate lawmaking, the Principle of Discourse takes a more complex form, that of the “Democratic Principle”. For law, Habermas insists, is not identical with, or subordinate to morality. Lawmaking does of course take into account issues of (universal)*moral justice*, but at the same time it also considers questions of another two types: *ethical* questions, regarding the substantive goals (themselves based on the identities and conceptions of the good) of those involved – i.e. those who will be affected by the legislation being considered and thus (in principle) participating in the lawmaking process; and *pragmatic*, means-ends questions, i.e. questions concerning the instrumental effectiveness of the relevant legislation. For a law such as, for example, one specifying a graduated income tax system, must, according to this conception, be at once (a) morally just (in a universalist sense),

(b) oriented to certain purposes (goals, values, interests) of the legal community concerned, and (c) effective in realizing those purposes.

The process of lawmaking in terms of the Democratic Principle, therefore, involves the employment of the three forms of practical discourse, which Habermas grounds in the theory of Universal Pragmatics: the moral (including both discourses of justification and discourses of application), the ethical, and the pragmatic (see the previous section of this study; also, Habermas, 1993: 1–17, and 1996: 159–162). In addition, however, it involves processes of bargaining, or negotiation, insofar as among the relevant purposes, or interests, pursued there are some that are not common to all the individuals or groups comprising the legal community, but differ from or conflict with each other and need to be reconciled on the basis of mutually acceptable compromises. These bargaining processes seeking to reach compromises among conflicting interests are not oriented to reaching understanding but are, for Habermas, necessarily based on participants' "factual power positions and the corresponding threat potentials" (1996: 140), i.e. strategic. This is the case, in the last analysis, even when bargaining is fair", that is, procedurally regulated in such a way that bargaining power is "disciplined by its equal distribution among the parties" (*ibid.*: 166); though, as we shall see in the next Part, Habermas does go somewhat towards establishing certain bridges between fair bargaining and argumentation oriented to reaching understanding.

Thus, Habermas distinguishes between (moral, ethical and pragmatic) communicative discourse, on the one hand and (essentially power-based, thus strategic), fair bargaining or negotiation, over interests, on the other, conceptualizing rational democratic lawmaking as a combination of the two.

The lifeworld, for Habermas, can exert effective influence on the subsystems precisely through rational lawmaking in the above sense, though not without "a

radical and broadly effective democratization” (Habermas, 1991: 261). In this way, the lifeworld acquires the potentiality to promote such forms of juridification as would transform the capitalist logic of the economic and administrative subsystems, reverse their colonizing influence and keep them under lifeworld control (ibid..).

However, despite his belief in the possibility – and indeed the desirability – of containing the media-steered subsystems and checking their colonizing effects on the lifeworld, Habermas insists that under conditions of modern complexity, the two media-steered subsystems remain indispensable for the reproduction of society, in particular for society’s material reproduction. The costs, in terms of time and energy, of the communicative coordination of social action (Habermas, 1996: 325) make it impossible for modern society in its totality to be socially integrated, as a lifeworld. The uncoupling from the lifeworld of the domains of the economy and state administration as subsystems steered by the non-symbolic media of money and power, respectively, possess an “intrinsic evolutionary value” (Habermas, 1991: 261); and, unlike the colonizing encroachment of those subsystems on the lifeworld, any attempt to reverse it would be not only unrealistic, but historically a rather regressive move. Responding to criticism of this view, Habermas states:

I believe for empirical reasons that there is no longer much prospect of the democratic reshaping from within of a differentiated economic system... by switching its steering from money and organizational power *completely* over to participation (ibid..).

Moreover, Habermas tends to consider the two media-steered subsystems by and large unproblematic, so long as they are properly “anchored” in the lifeworld and

prevented from exerting colonizing influences upon it. For, as we have seen, for Habermas it is colonization and not uncoupling itself that is responsible for the major pathologies of modern society.

2.5.2 Some Important Critical Perspectives

The main objections to Habermas's two-tier theory of modern society centre around two major aspects of it: Firstly, the distinction between a lifeworld socially integrated via communicative action orientations and systemically integrated domains of media-steered strategic action; and secondly, the normative acceptance of the economic and administrative subsystems as permanent features of modern society.

1) With regard to the distinction between Lifeworld and System, Habermas has been criticized for linking together too tightly types of action with forms of action coordination and societal integration and, ultimately, with domains of social life; namely, linking communicative action with social integration and the lifeworld, on the one hand, and strategic action with media-steered system integration and the societal domains of market economy and state administration, on the other:

Habermas identifies in a misleading fashion a typology of action with the distinction among types of coordination of action...[H]e connects too closely different types of action and different societal spheres of action, in so far as he speaks of subsystems of purposively rational or of communicative action... [A] linear relationship of correspondence among types of action, types of coordination of action, and societal domains is... established (Joas, 1991: 104).

Critics point out that in all societal domains we encounter various types of social action. In particular, the domains that belong to the lifeworld, such as those of family life, education, and the political public sphere, are not devoid of relations of power and domination (and that, of course, applies also when these domains are free from colonization by the media-steered subsystems); on the other hand, communicatively oriented action is not entirely absent from, or unrelated to, the economic and administrative spheres of action.

In view of these objections, which threaten to undermine the foundations of his two-level theory of modern society, Habermas further specifies his position as follows:

(a) He had never intended to portray the lifeworld as a purely communicative, free from domination area of social life – in his own words an area of social life corresponding to “an innocent image of power-free spheres of communication” (1991: 254). He acknowledges that, with the exception of (unlikely to occur) lifeworld conditions where non-repressive, non-deceptive and non-self-deceptive relations obtain, social integration takes place through overtly or latently strategic action, mainly in the form of adherence to “norms of domination which sublimate violence”, or in the form of distorted communication (*ibid.*). Thus, it is not so much a certain action type – namely, communicative action – that is associated with social integration and the lifeworld, as action *orientation* itself (communicative or strategic), as opposed to the unintended intermeshing of action consequences, which is responsible for system integration.

A point not brought out by critics, however, is that, although Habermas aspires to a lifeworld largely free from (overt or latent) strategic action, which is now widely present in it, he cannot adequately conceptualize – due to his dichotomous typology of communicative vs strategic action – lifeworld activity which is at once

success-oriented and communicative. The closest he comes to such a conception is with the idea of “fair bargaining”, which he introduces as a dimension of rational, democratic lawmaking and whose potential applicability could, perhaps, be broadened to cover all other areas of the lifeworld insofar as they include success-oriented interactions. Fair bargaining itself, however, remains at bottom a power-based, strategic form of interaction.

Yet, a great deal of activity in the complex, modern lifeworld is irreducibly success-oriented, requiring coordination through negotiated compromise between conflicting interests. Habermas himself recognizes this with regard to the political areas of the lifeworld:

Compromises make up the bulk of political decision-making process... Under conditions of cultural and societal pluralism, politically relevant goals often embody interests and value orientations that are by no means constitutive for the identity of the community at large... (1996: 282).

This, of course, is true of the lifeworld as a whole, and can be expected to be even more so under conditions of further rationalization of individual and group identities. But if this is so, and if (for him) compromises between conflicting interests can come about only through interactions that, even when they consist of fair bargaining, are in the last analysis strategic, Habermas cannot conceive of – and cannot hope for – an entirely, or even for the most part, communicatively integrated and rationalized lifeworld.

(b) Regarding the action domains of the market economy and state administration, Habermas keeps insisting that these are, indeed, “*in the final instance*” systemically integrated through the consequences of media-steered

strategic actions. Nonetheless, he explains that these domains are not entirely divorced from normative contexts and communicative action, and that to characterize them as areas of “norm-free sociality”, as he did in *The Theory of Communicative Action*, is misleading. For in the first place, the steering media of money and administrative power are legally institutionalized, thus anchored in the lifeworld. And secondly, the media-steered actions within the domains of the economy and administration are inseparable from socially integrative actions oriented to reaching understanding, though it is not the latter that are, in the last analysis, responsible for the systemic integration of these domains (1991: 256–8).

These responses suggest that Habermas continues to underestimate the effect that action-orientations have on the two domains which he considers to be integrated in a purely systemic manner. Concerning the sphere of the economy, for example, it is by no means evident that interventions into the market by state institutions, or by big private companies, affect economic integration – positively or negatively – through the unintended functional interconnections of their consequences, rather than by virtue of their orientations, that is, socially. And insofar as the latter is the case, Habermas’s systemic model fails to explain modern economic integration. Furthermore, to the extent that any such powerful interventions into the economy are functionally or normatively desirable, Habermas is not justified to advocate the (continued) existence of a purely systemically integrated market economy.

Equally, the systemic understanding of state administration seems to be highly questionable. Post-Weber organization theory – analyzing modern organizations, both public and private – identifies two types of organizational activities: the formal and the informal.¹⁶ Formal activities largely consist in adhering to organizational norms (rules, regulations, procedures, etc) and obeying orders from hierarchical

superiors, in accordance with Weber's ideal type of bureaucracy. Research, however, has revealed that much of the activity going on in allegedly purely bureaucratic organizations is not of this formal type.

In his penetrating critique of Habermas's systemic conception of organizations, Thomas McCarthy refers to significant informal elements:

[I]t seems obvious that there are also situations in which organizational superiors can act authoritatively only "with reservation", that is, in which they know they cannot achieve their goals without collegiality, cooperation, mutual understanding (1991: 129).

To this we can add two points. Firstly, informal organizational activities are not limited to communicative ones, as the above passage implies. Strategically oriented informal action also takes place. Secondly, informal organizational activity, communicative and strategic, is not sporadic and inconsequential. It has a substantial presence, and plays a significant role with regard to the organization's results as well as its functioning. In particular, organization theory regards many of the informal elements of organizations as an answer to disfunctionalities inherent in the purely formal-bureaucratic model. Though all modern organizations combine formal and informal elements, the relative weight of the two aspects varies from organization to organization and from one period in the life of an organization to another.

Habermas is aware of the gap between activity in actual organizations and that anticipated by the idealized model of bureaucracy. Nonetheless, he denies that non-formally-regulated behaviour plays any significant organizational role,

considering it as largely “neutralized”. For Habermas, action in modern organizations is, in essence, formally regulated:

[T]he classical model of bureaucracy is right in one respect: action within organization falls *under the premises* of formally regulated domains of action (Habermas, 1987: 310).

On this account, it is clear that Habermas views the overall integration of organizations exclusively in terms of relations of formal authority. This is made explicit in the following passage:

The basic characteristic of the *action orientations* of members is... the fact that all their actions fall under the conditions of organizational membership, that is to say, under the premises of a legally regulated domain of action. When we understand business concerns as *self-regulating systems*, it is the aspect of legal organization that comes to the fore (*ibid.*, emphasis added).

Leaving aside Habermas’s virtual dismissal of the possibility that informal elements – whose existence he acknowledges – might have an impact on an organization’s functioning, his account of organizations as *self-regulated systems* in terms of members’ *action orientations*, as clearly stated above, comes as a surprise. For as McCarthy aptly points out (1991: 129–30), this account is at odds with Habermas’s conception of system integration as taking place not through *action orientations* but via the functional interconnections of *action consequences*.

Evidently, the conception that emerges from Habermas’s analysis of how modern organizations integrate is one of social integration, essentially through formally-

regulated action orientations. This formal regulation of intra-organizational actions is, to be sure, thought to be power-driven, making these actions strategic in nature.

Viewing modern private and state organizations as societal units integrated socially – rather than systemically – through a varying combination of formally-regulated and informal, strategic and communicative action, raises the question whether it is possible to conceive of, and design, organizations that are integrated exclusively – or at least predominantly – by means of communicative action orientations. This question will be revisited in Part Three.

2) With regard to Habermas's favourable attitude towards the continued existence in modern society of media-steered subsystems of the economy and state administration – albeit, when these are relieved of their colonizing dynamic – critics reproach him on two main counts: (a) for not paying sufficient attention to the problematic character of the two subsystems even if they were kept in check by the lifeworld and prevented from making colonizing inroads into it; and (b) for considering the economic and administrative subsystems indispensable for modern, complex society, and, by implication, maintaining that systems theory is of enduring value as a scientific approach to modern social phenomena.

(a) Both money-steered market economy and power-steered, rule-based administration are beset by serious "intra-systemic" problems and contradictions (Joas, 1991: 117). These problems threaten the very self-integrative and goal-attainment capacity that so much recommends the subsystems to Habermas, and call for additional efforts to secure the subsystems' stability and efficiency.

With regard to the market economy, the obvious example (which, needless to say, is by no means unknown to Habermas) is the internal crisis tendencies initially identified and analyzed by Marx and later studied also by mainstream – above all

Keynesian – economics (Berger, 1991: 177). From a non-systemic, normative point of view, of course, the capitalist market economy – even when kept to its proper domain, away from meaning- and identity-producing areas of the lifeworld – can still be criticized for exploitation and injustice, as well as for instrumentalizing and commodifying human beings.

The “dysfunctions” of the pure, Weberian model of hierarchical, rule-governed (bureaucratic) organizations, when that model is applied in complex and rapidly changing situations (in contrast to being applied to simple, routine, stable tasks) have also been illuminated by several studies in the fields of the sociology of organizations and management theory. There are two core causes of these dysfunctions. Firstly, in the complex, fast changing and unpredictable conditions of modernity, strictly formal organizations do not possess the flexibility needed to respond in a timely and effective way to emerging situations; for it is not possible to devise a system of formal rules and hierarchical decision making that, in those conditions, could anticipate and cater in advance for all the challenges that an organization may face, or to give formally-determined solutions to unanticipated challenges when they emerge. Secondly, the more complex and fluid the conditions in which they operate, the less able formal organizations are to impose control over the actions of their members, often having to do so in the face of strong resistance.

These factors tend to give rise to informal practices, either in an effort to promote organizational goals more effectively (and this despite the rigidifying effect which the formal system has on members’ personalities), or in order to serve particularistic interests, including members’ desire to enlarge their control over their work situation. When they arise, these informal practices in some cases establish themselves as useful and acceptable elements of the organization, and

to that extent the resulting organizational integration ceases to be systemic in the rule-governed manner claimed by Habermas; whereas in other cases they are treated by the organizational authorities as unacceptable acts of disobedience, and responded to by further tightening of the formal system, which, in turn, only exacerbates the initial disfunctionalities, leading thus to what has been called the “vicious circle” of bureaucracy (see, for example, the classic studies, Gouldner, 1954; Merton, 1957; Burns and Stalker, 1961; Blau, 1963; Crozier, 1964).

Furthermore, strictly bureaucratic organizations and their Tayloristic managerial complements have been stigmatized for their dehumanizing effects on their members. Indeed, one of the most profound critics of bureaucracy in this respect was none other than Max Weber himself, who, while praising it for what he considered to be its technical superiority, at the same time spoke movingly of bureaucracy as an iron cage that deprives those working in it of freedom and meaning (see Appendix III; also Braverman, 1974). Here it should be added that this dehumanizing side of bureaucratic-Tayloristic economic and state administrative organizations cannot but have a corrupting effect on those areas of social life which Habermas strictly reserves for the lifeworld, such as the family and the political public sphere. This corrupting effect is over and above any incidence of what Habermas describes as the colonizing bureaucratization of those lifeworld domains. Moreover, it is intrinsic to bureaucracy, and thus inextricable from it, unlike its tendencies to colonize the lifeworld by means of bureaucratization, which – at least in Habermas’s view – can be kept in check.

b) Habermas does not show convincingly how the domains of the economy and state administration can operate as well-functioning, media-steered systems which nonetheless make no colonizing inroads into the lifeworld. In Peter Dews’ words, “*The Theory of Communicative Action* bequeathed us the problem of how the

life-world can 'steer'... systems, and limit their intrusions, without disrupting their functioning" (Dews, 1993).

But even on the assumption that this problem could be tackled successfully, critics, including generally sympathetic ones, reproach Habermas for over-readily surrendering the economy and state administration to media-steered subsystems of strategic action and for consequently "ced[ing] too much territory to systems theory" (McCarthy, 1991: 120). It is argued that in this way he ultimately gives up Critical Theory's – including his own earlier – aspirations for an entirely non-reified, emancipated and cooperative social life, and seriously compromises his emphatic advocacy of participatory and deliberative forms of democracy, especially in his more recent work. In addition, his embracing of systems theory, even if partial, is at odds with his earlier polemical characterization of it as "*the Hochform of technocratic consciousness that enters the lists against any tendencies toward democratization and promotes a depoliticization of the public sphere by defining practical questions from the start as technical questions*" (*ibid.*: 133–4). Thus, by being assigned to systems theory, the domains of the economy and state administration are said to be technocratically insulated from moral, ethical and political critique. In the same vein, several commentators deny that systems theory is best suited, let alone necessary, for the analysis of the market economy and state administration even in their present forms. Both Hans Joas (1991) and Thomas McCarthy, among others, argue that purely action-theoretic accounts of modern society are not impossible.

To be fair, in contrast to the exclusively systems-theoretic approach to society, Habermas's hybrid theory does not reify what it takes to be media-steered subsystems. It does not treat them as given, inevitable parts of social reality – as second nature. Habermas, rather, considers them as social arrangements which

can be retained and kept under control (in a form relieved of any colonizing capacity) through deliberate lifeworld choices, particularly political-legislative ones. Thus, in the final analysis, he subordinates the media-steered subsystems to a rationalized lifeworld, and systems theory to the theory of communicative action. Habermas's recommendation of retaining the media-steered economic and administrative subsystems constitutes, in a way, a self-limitation of communicative rationality for the sake of greater benefit – a move analogous in certain respects with Odysseus asking his comrades to tie him on the mast in order to protect himself from the anticipated, destructive lure of the Sirens.

With some risk of oversimplification, it can be argued that Habermas's favouring the continued operation of the economy and state administration as systems of media-steered strategic action is based on a comparison between two conceptions of modern society: the first is one of a society consisting of lifeworld and system domains and attaining high effectivity regarding its "material reproduction", in the sense of both material well being and societal stability and integration; the second conception is one of a society coextensive with a lifeworld structured and reproduced by means of highly rationalized communicative action, but which, due to a heavy cognitive and motivational burden, would fail to achieve adequate levels of productive and integrative effectivity. The choice between the two would thus involve a trade-off between productive and integrative effectivity, on the one hand, and human emancipation from objectification and strategic instrumentalization, on the other; or, put differently, a trade-off between society's overall purposive rationality (with respect to the goals of material well being and integration) and its communicative rationality. For a critical/communicative theorist such as Habermas to opt for the former type of society, he must consider the gains in effectivity/purposive rationality involved in that choice as far outweighing

whatever losses in emancipation/communicative rationality are thought to be suffered because of it.

This judgement, however, is based on shaky ground. On the one hand, Habermas, as we have seen, underestimates the dysfunctionalities and other serious shortcomings of the two favoured subsystems, even when considered in their pure form; he thus overrates their contribution to society's purposive rationality. On the other hand – and more importantly – he fails to put forward a sufficiently worked out concept of the rejected option of a thoroughly communicative, yet complex and modern society, on the basis of which to assess the likely level of that society's purposive rationality. More particularly, he has not conceptually envisaged a communicative economy or a communicative state administration; nor has he constructed other, intermediate options, such as of societies that include more limited systemic "pockets" of economic and administrative action. All such concepts would have to incorporate possible mechanisms whose function would be to condense, without negating, communicative action and thus lessen, as far as possible, society's communicative – cognitive and motivational – load and correspondingly enhance its productive and integrative effectivity.

In fact, as has already been pointed out, Habermas lacks a concept of action that is at once success-oriented and communicative, which would be necessary for meeting this challenge. For without a concept of success-oriented communicative action he would not be able to extend the application of the communicative approach over any part of the thoroughly success-pervaded economic and administrative fields of action, or indeed to any competitive areas of what he now identifies as the lifeworld.

A final, major point of concern regarding Habermas's favourable stance towards retaining the subsystems – a point that would evidently be valid also if he had

chosen the other way – is that his choice is based on his own personal evaluative assessment of the relevant gains and losses, over and above the estimates of these gains and losses themselves. There would be no misgivings, of course, if he intended the judgement under discussion as the contribution to a public dialogue on the issue, made by one of the participants. But to present this judgement as the authoritative, epistemologically privileged view of a social theorist, or philosopher, contradicts Habermas's own proceduralist conception of practical rationality – indeed his postmetaphysical conception of reason as such. In accordance with that conception, the question of what societal arrangements should be made, including whether to institutionalize any media-steered subsystem of any particular kind or scope – a question, to be sure, that becomes possible only once reifying attitudes towards social structure are seen through – ought to be left to processes of practical discourse in which all members of society can participate. Such discourses should not reach conclusions *a priori*, on the basis of cognitively constructed possibilities, but should be linked to social experimentation aimed at testing the realizability of imaginative visions in practice.

In the following, final Part of the thesis I shall be concerned with the development and philosophical grounding – still largely within the Habermasian communicative paradigm – of the missing concepts of *success-oriented communicative action* and of the rationality of that action. On the basis of these concepts, it will be possible to work towards developing ideas of a society coordinated and integrated exclusively or predominantly in communicative and communicatively rational ways.

PART THREE

TOWARDS A THOROUGHLY COMMUNICATIVE CONCEPTION OF SOCIAL ACTION AND ITS RATIONALITY

3.1 A CONCEPTUAL GAP IN HABERMAS: BRINGING TOGETHER CRITICAL THEMES DEVELOPED IN PARTS ONE AND TWO.

The dichotomous classification of action and practical rationality which we find in both the early work of Habermas (culminating in *Knowledge and Human Interests*, 1972) and in the later, post-linguistic-turn paradigm (mainly expounded in *The Theory of Communicative Action*, 1984, 1987) creates a serious conceptual lacuna.

In the early work Habermas distinguishes between *techne* and *praxis* (or, respectively, between work and interaction); in certain early texts he even uses, correspondingly, the terms purposive and communicative action which predominate in the later paradigm. In *techne* a subject seeks to realize a pre-given/previously adopted goal by means of technically manipulating an object, including possibly an objectified human being, social entity or social process. Thus, in *techne*, and consequently in technical rationality – the type of rationality involved in *techne* – the employment of certain means to an end and the technical manipulation of an object by a subject are inextricably bound together.

In *praxis*, on the other hand, two mutually respectful subjects cooperatively try to agree on common values or norms, or on how to act in accordance with already agreed values or norms. Thus, *praxis*, and hence the corresponding form of rationality, which I have termed praxial rationality, also involve two inextricably

bound elements: the concern with determining or following common values or norms, on the one hand, and a relationship between mutually respectful subjects, on the other.

An analogous situation exists in the later paradigm. Here, again, action is either oriented to success or oriented to reaching understanding. In the first case it is purposive, or purposive-rational (instrumental, if it involves a relationship between human and natural objects; strategic, if it involves a social interaction). In the second case it is communicative. Practical rationality is, accordingly, either purposive (instrumental or strategic) or communicative. Purposive action and rationality concern, just like *techne* and technical rationality, the employment of means for attaining given ends through technically manipulating or, what is a somewhat more precise, equivalent expression, *causally influencing*, objects. Communicative action and (practical, for there is also theoretical) communicative rationality, on the other hand, like *praxis* and praxial rationality, concern the mutually respectful effort to determine, clarify or apply common values and norms.

A significant innovation in the later work is a clear internal differentiation within (practical) communicative rationality. This differentiation is between: (a) ethical rationality, dealing with what is "good for me or for us", more precisely with the clarification and expression of a particular individual or collective subject's authentic substantive values, intentions, identity; and (b) moral rationality, concerned with what is unconditionally right, or just, for all, in other words with specifying and applying universal moral norms by means of the Universalization and Appropriateness Principles, respectively. The differentiation within praxial rationality, in the earlier work, between (ordinary) hermeneutics and the depth hermeneutics of critical-emancipatory self-reflection is not equivalent to the new differentiation. For although self-reflection is considered to be a higher form of

practical rationality than hermeneutics (as morality is taken to be *vis à vis* ethics), it nonetheless remains within the bounds of (Aristotelian/Hegelian) contextual immanence, not aspiring, as communicative morality does, to (Kantian) transcendent universality.

Both the early and the later forms of the dichotomous classification of action, and the rationality of action, seem to categorially exclude the possibility of success-oriented yet communicative action and corresponding rationality. They exclude, in other words, the possibility of such cases as that of one person seeking to pursue a goal by means of another person, and doing so through communicatively coming to an understanding with that person.

In the early work Habermas, on a philosophical plane, rejects all objectification/reification of human/social reality, not least that involved in a social science modelled on the natural sciences and the social engineering based on it. Yet, as we have seen, in some of his social-theoretical discussions of that period he seems to be resigned to the prospect that certain areas of social action that are close to the productive process, such as the management of work organizations, will always be of an instrumental/purposive, thus objectifying, nature.

In his later work, while still espousing the initial emancipatory philosophy of Critical Theory, Habermas, much more explicitly and systematically than before, argues that two major areas of social life, namely the market economy and state administration, cannot but be organized as media-steered subsystems of strategic action. The main thrust of the argument is that in conditions of modern complexity it would be extremely impracticable and socially costly to coordinate and integrate action in these areas in communicative ways. This line of argument implies that a communicative coordination and integration of such actions would not be impossible in principle. Yet the conceptual tools for such an – even in principle –

possibility are not available to Habermas. For the (in theory) possibility of communicative coordination and integration of such thoroughly success-oriented areas of social life as the economy and state administration would require the missing category of success-oriented communicative action.

Habermas came under severe criticism, especially from fellow Critical Theorists who shared his anti-reification, emancipatory aspirations, for surrendering the economy and state administration to media-steered subsystems of strategic action, and thus handing these fields of social activity to the jurisdiction of Systems Theory rather than keeping them within the scope of Critical Theory. Under the weight of this criticism, and in the light of his own subsequent inquiries concerned with establishing a communicative basis for democratic politics and law – areas of social life that inescapably contain a strong success-orientation component – Habermas has come more recently to add certain new elements to his analysis of action and practical rationality which acknowledge and in some ways address the problem of the categorial gap identified above. These new elements appear mainly in the works *Between Facts and Norms* (1996) and "Some Further Clarifications of the Concept of Communicative Rationality", in *On the Pragmatics of Communication* (1998a). Before examining them, however, we need to take a closer look at Habermas's dichotomous typology of action and practical rationality, presented in its most mature form in *The Theory of Communicative Action* (1984, 1987), with a view to specifying with more precision the category claimed to be missing from it.

3.2 A MORE PRECISE SPECIFICATION OF THE CONCEPTUAL GAP

In the concept of purposive action and rationality, as we have seen, Habermas binds inextricably together an orientation to success in employing means to attain goals with a causal-manipulative relationship between subject and object. Similarly, in the concept of communicative action and rationality he binds together an orientation to values and norms with a mutually respectful relationship between subjects. In binding these two pairs of aspects together, Habermas fuses two issues which, at first sight at least, seem to be distinct. First is the issue concerning the *orientation* of an action (and of the rationality of the action), which offers two options: orientation to means for the attainment of ends (which we can also refer to as orientation to success, using Habermas's term, though without associating it necessarily with a subject-object relationship as he does); and an orientation to values and norms. Second is the issue concerning the *relationship* between the acting agent and the other agents, or the objects, involved in the action (and in the corresponding rationality of action). This issue also gives two options: a technically manipulative/causal, subject-object relationship and a mutually respectful subject-subject relationship. I shall refer to the latter as *dialogical*, or *communicative*, relationship, (though using the term communicative strictly in the sense of mutually respectful, or dialogical, without associating it necessarily with an orientation to values and norms as Habermas does).

If indeed we are dealing with two different issues each offering two options, our classification of types of action and the rationality of action becomes fourfold, as shown in the grid below (Figure 2).

| | | Orientation of action and rationality | |
|---|---|--|--|
| | | Orientation to success, to using means to attain ends | Orientation to determining and applying values and norms |
| Relationship involved in Action and rationality | Technical, causal, subject-object relationship | (1) Success (means-ends) oriented <i>and</i> causal action and rationality | (2) Value and norm oriented <i>and</i> causal action and rationality |
| | Dialogical, Communicative, subject-subject relationship | (3) Success (means-ends) oriented <i>and</i> dialogical/communicative action and rationality | (4) Value and norm oriented <i>and</i> dialogical/communicative action and rationality |

Figure 2

Out of the four types of action and practical rationality corresponding to the four areas in the grid above, Habermas's dualistic classification covers at first glance only two: Habermas's *techne*, or purposive (instrumental or strategic) action and rationality, is covered by the category in Area 1 in the above grid. Habermas's *praxis* or communicative action and rationality is covered by the category in Area 4. On a closer look, the category in Area 2 might be interpreted in a way that also brings it within the scope of the Habermasian dualistic typology, more specifically, under Habermasian *techne*/purposive (instrumental or strategic) action and rationality: In a given interaction, agent A may set as an end of her action to get another agent, B, to adopt as his own, or to apply, a certain value or norm – though, perhaps, not necessarily a value or norm which A espouses herself. In other words, A's action is oriented to using B as a means to her own end. Now, if A pursues her goal in a manipulative, causally effective way (either consciously or unwittingly/self-deceptively), while the interaction in question is overtly about the determination or application of the two parties' values or norms, we have a typical

instance of action/rationality which belongs to Area 2 but nonetheless comes under the Habermasian concept of the purposive. Thus, once we include the aim of making another adopt or apply certain values or norms among the possible ends of agents, all the instances of the category in Area 2 can be grasped in terms of Habermas's concept of *techne*, or the purposive, more precisely in terms of the strategic variety of the purposive.

The category in Area 3, however, concerning the communicative employment of means to ends, or, put differently, the pursuit of success in a communicative way, is far more problematic with regard to Habermas's dualism. To be sure, Habermas's concepts have no difficulty in capturing cases where two or more agents communicatively seek to come to an understanding about how to pursue a common goal by means of technically manipulating a *third* factor – material or human/social. Such cases fall squarely under the concept of purposive action/rationality, the communication between the agents being about the truth of factual statements expressed perhaps in the form of hypothetical imperatives. But this is not what the category in Area 3 is about. According to the parameters on the basis of which the grid is constructed, the category in question concerns, rather, the dialogical/communicative/mutually respectful relationship between a success-oriented subject and the means to that subject's end(s).

What might count as an instance of this concept? Surely, we can readily rule out cases where human beings use *objects of nature* as means to their ends, for we can safely assume that, literally speaking, there can be no communicative relationship with such objects (with the possible exception of higher animals). The concept, however, need not therefore be an empty shell. For it might be possible to conceive of cases where success-oriented human beings seek communicatively

– rather than causally – to induce *other human beings* (to act so as) to become means to the realization of the former's own goals.

However, some further clarifications are in order, concerning the meaning of "success-oriented" and "own goals". Habermas's typology does not rule out entirely the communicative use of another agent in pursuing one's goals. There are certain kinds of goals, certain forms of success, whose communicative pursuit by means of (the actions of) another agent falls under the category in Area 4, and can thus be captured by Habermas's concepts of communicative action and rationality. Consider, for example, the situation where agent A tries to get back from agent B a book which B borrowed from A and failed to return as promised. This is surely a case where one person is trying to attain a goal by means of another person. Agent A can pursue her goal strategically, by trying to influence B causally, for example through the use of force, threats, or deception. Alternatively, however, she can do so communicatively, according to Habermas's concepts, by appealing to the moral norm – presumably acceptable to both A and B – that promises should always be kept. In the same way, an attempt by A to realize, by means of B, a goal that violates a certain mutually accepted moral norm, can be rejected by B also in a communicative way, through an appeal to the norm in question.

Thus, goals that constitute applications or violations of moral norms – let us refer to them as *morally regulated* goals (in a positive or negative sense) – can, according to Habermas, be pursued (by A), through another agent (B) communicatively. Their realization is either prescribed or prohibited by the relevant norms. If prescribed, it is a moral right/entitlement of A and moral obligation/duty for B; if their realization is prohibited, it is a moral obligation/duty of A to refrain from pursuing it and a moral right/entitlement of B to refuse to become a means to

it. Morally regulated goals, in the above sense, even if the initiative for their realization comes from one of the two sides – who thus considers them as his or her “own goals” – are in a way, ultimately shared goals, determined by the two sides’ common adherence to the relevant moral norms. Consequently, actions oriented to the realization, through others, of morally regulated goals can be communicative, according to Habermas’s typology, and as such included in the category in Area 4 of our grid.

At first sight, what is argued above about actions oriented to the realization of *morally* regulated goals, might be thought to apply with regard to goals regulated by shared *non-moral* norms (e.g. laws) as well. Here, however, an important qualification needs to be made. In his analysis of law – which can be extended to non-moral norms in general – Habermas argues that the determination of legal norms depends on a combination of moral, ethical and pragmatic considerations; and insofar as the ethical grounds (i.e. the values) involved are not originally common to the members of the legal community but are shared as a result of compromises, they are inevitably based on bargaining, which, even when “fair”, is for him a strategic process (see section 2.5.1, above).¹ Therefore, for Habermas, goals regulated by non-moral norms are pursuable in an entirely communicative way – thus covered by the category in Area 4 – only if the norms in question are shared without the intervention of any bargaining. And this can be known for sure only at the end of a process of rational reflection that reaches back to the moral norm that underlies the non-moral norm in question.

Another class of (*prima facie* not common) goals that can, in a similar manner, be recognized by Habermas’s Area 4 concept as capable of being communicatively pursued through another agent are those goals which are positively or negatively regulated by a shared substantive (ethical) value, provided, again, that it is shared

without the involvement of any bargaining. These are, in other words, the goals whose realization can be shown to be desirable or undesirable, good or bad, in relation to such values, including interests, higher goals, conceptions of the good, etc. In a similar way as in the case of action involving non-bargaining-dependent *normatively* regulated goals, actions oriented to the realization, through others, of non-bargaining-dependent *value*-regulated goals can, according to Habermas's typology, be pursued and responded to communicatively, through appeals to the relevant shared values; they can thus also be included in Area 4.

Shared substantive values (e.g. social harmony, clean air, education for Third World children, environmentally sustainable development) must not be confused with shared conceptions of self-regarding or, more generally, particularistic values, such as "personal health", "family prosperity", or "national security". In these cases interacting agents may have the same idea as to what the values personal health, family prosperity and national security are, but what they value in them may not be the same; it may be personal health for *themselves* (and perhaps for certain other individuals), prosperity for their *own* families (and perhaps for certain other families) and security for their *own* nation (and perhaps for certain other nations). Shared ideas of particularistic values, unlike shared, in the sense of coincident, values, entail different, perhaps conflicting goals for interacting agents.

Certainly, in any interaction situation there might be goals which can be (shown on reflection to be, genuinely, not just at first sight), (a) unregulated by shared norms or values, or (b) regulated by norms or values which are shared but whose sharing depends on bargaining. These goals are, in the last analysis, normatively and evaluatively permissible or "optional" for the agents concerned. When adopted and pursued by an agent in the context of an interaction, any such optional goal is that agent's "own goal" in a sense stricter than that in which a genuinely shared goal is.

Strictly “own goals”, in the above sense, can be called *egocentric* goals, using this not entirely accurate term in accordance with Habermas’s usage in analogous contexts, as, for example, when he speaks of “egocentric calculations of utility” (1984: 94), or “egocentric calculations of success” (*ibid.*: 286). Egocentric goals, as defined here, are not necessarily selfish – hence the terminological oddity. They may well be pursued for the sake of a third party. They are egocentric simply in the sense that they are not shared by interacting agents, or are shared only on the basis of bargaining.

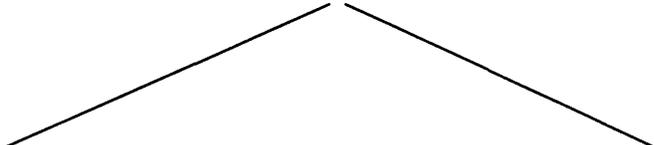
It is precisely the communicative pursuit, by means of other agents, of egocentric goals in the above sense, that belongs exclusively to Area 3 of our grid, and cannot be captured by the dualistic typology which Habermas adopts both in his early work and in *The Theory of Communicative Action* and related post-linguistic-turn writings. For according to this typology such goals can be pursued through others only in an instrumental/strategic manner. As an example, consider the following situation:

John wants Mary to lend him one of her books – he has a goal which he wants to realize by using another person as means. As in our previous example of A wanting to get her book back from B, John can pursue his goal strategically, by causally inducing Mary to lend him the book. Alternatively, John could explore the possibilities for a communicative pursuit of his goal, for example on the basis of a normative obligation Mary had already undertaken (e.g., a prior promise), or on the basis of a shared value or higher goal which generates certain shared commitments, such as a common project for which it is necessary for John to consult the book in question. But if no such normative or evaluative basis exists for a communicative pursuit of John's goal – if John wants the book for his own purposes which are in no way shared by Mary, his goal thus being egocentric, in

our sense – Habermas's dualistic typology does not give John any alternative way of pursuing his goal other than the strategic one. No possibility is available in Habermas's typology for a communicative handling – a communicative "negotiation" – by John and Mary of John's egocentric request to borrow Mary's book.

The classification of goals proposed here is summed up below, in Figure 3.

(*Prima facie* not common)
goals persuable
through other agents



Goals ultimately shared by interacting agents (in the context of a given interaction): Goals regulated (thus either prescribed or ruled out) by norms, or values shared without bargaining.

(For Habermas, these are capable of being pursued through others communicatively, thus catered for by Area 4 of the grid in Figure 2).

(Genuinely) egocentric goals, (in the context of a given interaction): Goals not regulated by shared norms or values (that is, goals that are permissible – thus not ruled out – but not prescribed by shared norms or values, including goals adopted in the absence of any shared norms or values) or regulated by norms, or values shared only on the basis of bargaining.

(For Habermas, these are not capable of being pursued through others communicatively, thus catered for by Area 3 of the grid in Figure 2, which is not recognized by Habermas's typology).

Figure 3

In contrast to traditional societies where the closely knit networks of shared norms and values tends (in a largely unreflective manner) to regulate much of social activity, in modern societies, the far more abstract normative-evaluative framework leaves unregulated a great deal of social interaction, which, as a result, consists of the reciprocal pursuit of egocentric goals. Consequently, the absence from Habermas's typology of the possibility of the communicative coordination of egocentric pursuits, *necessarily* concedes to strategic action not only realms such as the economy and state administration but also a great deal of social action in both the private and public areas of what Habermas calls the lifeworld (family life, education, the political public sphere etc.); and this, despite the fact that he considers the lifeworld as the domain of society where communicative action should – and to a substantial degree does – reign.

Having established that Habermas's early as well as post-linguistic-turn typology of action and rationality does not allow for the communicative pursuit of egocentric goals by means of other agents – let us speak, for short, of *success-oriented communicative action and rationality* – the question remains whether, and if so how, such a form of action and rationality is conceptually possible; and in what way such a concept, if possible, would be grounded in the theory of Universal Pragmatics and be made to cohere with the rest of Habermas's conceptual architectonic. Furthermore, we need to ask whether success-oriented communicative action and rationality, if conceptually possible, is practicable as a way of dealing with the whole range of the considerable coordination and integration challenges facing modern societies.

As we have seen in Part Two (section 2.5), in *The Theory of Communicative Action* Habermas explicitly expresses the view that, at least in the predominantly success-oriented areas of economic and state-administrative action, tackling the

tasks of coordination and integration in a way other than through subsystems of strategic action would be highly unrealistic in conditions of modern complexity. In doing so he appears to be implying that a communicative coordination and integration of action even in those two societal domains would not be *in principle* impossible. Nonetheless, he does not explain how it would be in principle possible, thus leaving open also the question of the possibility of communicatively dealing with the coordination and integration of success-oriented action in areas of the lifeworld. In some more recent writings, as already indicated, Habermas discusses issues closely related to the conceptual possibility of *success-oriented communicative action and rationality*. In the following section we shall take a look at these discussions.

3.3 FAIR BARGAINING AND WEAK COMMUNICATIVE RATIONALITY: SOME RECENT THINKING BY HABERMAS CONCERNING ACTION AND ITS RATIONALITY.

3.3.1 Fair Bargaining

In *Between Facts and Norms*, his major work of political and legal theory, Habermas deals with the obvious tension between the idealized claims to legal legitimacy made by modern democratic states and the far from ideal, actual social and political conditions of lawmaking. His aim is to overcome, in terms of his communicative paradigm, the serious problems which he identifies in all the major efforts to provide an adequate theory of modern politics and law. Habermas is not satisfied either with the positivist-realist approaches to law and its legitimation – which emphasize legal enactment and functionality, neglecting the rational justification of law – or with the ways in which various normative theories (of rational natural law) deal with the question of justification.

To begin with, Habermas distinguishes clearly between the essentially political process of *lawmaking* and the essentially administrative process of *law-enforcement*, locating the former within the domain of the lifeworld and the latter within the realm of media-steered subsystems of strategic action (see also section 2.5.1, above). Whereas in *The Theory of Communicative Action* he stresses the role of juristification (“the tendency towards an increase in formal law”, 1987: 357) as a vehicle of colonization of the lifeworld by media-steered subsystems, in *Between Facts and Norms* he focuses on the potential role of lawmaking (as distinct from law-enforcement) as a vehicle of delimitation and control of the subsystems, thus as an agency of opposition to, and reversal of the colonization of the lifeworld. Indeed he now self-critically explains that although in the former work he emphasized the direction of influence from systems to the lifeworld, the

architectonic of his theory does not privilege that direction of influence over its opposite, i.e. the direction from the lifeworld to the subsystems (Habermas, 1993: 170–1).

Habermas's effort to provide a satisfactory theory of the rational justification, and thus the legitimation, of law focuses on the nature of the decision making (or "will-formation") that is involved in the political process of lawmaking. In contrast on the one hand to the Kantian-liberal tradition which puts the emphasis on individual human rights and is seen as making the rational justification of law dependent on law's conformity to moral norms alone, and on the other hand to the Rousseauian-republican tradition which emphasizes popular sovereignty ("the general will") and sets conformity to society's shared values (the general interest) as the criterion for law's justification, Habermas proposes as such a criterion a more inclusive and complex, four-dimensional conception of rational political-legal will formation.

Law, though not reducible either to morality or to ethics (in the Habermasian sense of these terms) has nonetheless both moral and ethical aspects; a piece of legislation must respect moral norms and also take into consideration the values, interests, needs and preferences, of those to whom it is addressed. But at the same time, lawmaking decisions often need also to consider "pragmatic" questions concerning factual, means-ends relationships. With regard to these three aspects, a rational lawmaking process can be based on the three forms of rational discourse that Habermas had in previous work identified as the three dimensions of his tripartite conception of practical rationality: the moral (justification and application), the ethical and the pragmatic discourses, grounded in Universal Pragmatics and corresponding to the validity claims to rightness, truthfulness and truth, respectively (*ibid.*: 1–17; see also section 2.4, above).

However, over and above these three forms of discourse, Habermas recognizes a fourth dimension in the process of lawmaking and wider political decision-making, namely bargaining, or negotiation. Bargaining takes place when the parties in an interaction pursue (*prima facie*) opposing goals or interests, over which, nonetheless, they have to reach an agreement. In Habermas's formulation, it is a process of "negotiation between success-oriented parties who are willing to cooperate" (1996: 165). The introduction of the notion of bargaining in this context reflects the recognition on the part of Habermas that politics, lawmaking in particular, very often involves, among other things, the reciprocal pursuit of what we have earlier called egocentric goals – we can, in the same way, speak of egocentric interests – by means of (the consent and usually the active effort of) others.

Now, from the point of view of our inquiry concerning the missing category of success-oriented communicative action and rationality, the pertinent question is whether bargaining in the above sense could reasonably be construed as a possibly communicative process. For if it could, the concept would (in principle) have applicability far beyond lawmaking, and political will-formation in general, becoming a model for the communicative pursuit of egocentric goals across the whole range of social action.

Although, as indicated in the previous section, Habermas ultimately rules out the possibility that any form of bargaining can be communicative in a strict sense, his full position on the subject is fruitfully differentiated. On the one hand, he regards bargaining over opposing interests as a form of strategic process based on the use of "factual power" (e.g., "threats and promises"), unlike discourses, where power is entirely neutralized and only the force of the better argument counts. Whereas discourses aim at reaching "rationally motivated consensus [that] rests

on reasons that convince all the parties in the same way" (1996: 166), bargaining processes aim at interest-balancing compromise agreements "accepted by the different parties each for its own *different* reasons" (*ibid.*). i.e. in pursuit of its own (egocentric) interests. Consequently, when he refers to the four constituent elements of the democratic lawmaking process, Habermas always differentiates between (moral, ethical and pragmatic) *discourses* and *bargaining*, speaking standardly of "discourses and negotiations".

But on the other hand, Habermas discusses a special form of bargaining, namely "fair bargaining", or "fair negotiation", which he places on the strategic side of the strategic/communicative divide only with considerable qualifications. Compromises reached in the course of bargaining in general, are acceptable to participants because the latter consider that these compromises leave them better off than (or at the very least as well off as) they would be without them. Participants who feel that this condition is not fulfilled would stop cooperating and pull out of the bargaining. However, the overall gains produced in the process need not be fairly distributed among the participants. For this distribution depends on the possibly unequal power that the different participants can wield in support of their own interests. Fair bargaining, though, is bargaining that results in balanced, or fair, compromises, i.e. compromises in which the gains are fairly distributed among the participants. Such compromises – still accepted by the participants on the basis of their own different reasons – are the outcome of bargaining processes which are regulated through procedures which secure the equal distribution of power among participants – not quite the neutralization of power as is the case in discourses. More precisely, procedures of fair bargaining

provide all the interested parties with an equal opportunity
for pressure, that is an equal opportunity to influence one

another during the actual bargaining, so that all the affected interests can come into play and have equal chances of prevailing (1996: 166-7)

The procedures that secure the equal distribution of bargaining power among participants and hence the equal opportunity of their interests to count in the ensuing compromises, can, according to Habermas, be justified by means of moral discourses (*ibid.*: 167). In this way, fair bargaining processes are for Habermas linked to the Discourse Principle. It is for this reason that he considers lawmaking processes which include fair bargaining in the above sense (in addition to moral, ethical and pragmatic discourses) as on the whole “rational” processes which confer legitimacy on the laws that result from them.

Nonetheless, as explained in *Between Facts and Norms*, even fair bargaining, being only indirectly linked to the Discourse Principle, is ultimately a non-communicative, strategic, process. It is a process which, in the final analysis, relies on power, thus on causal influence, rather than on the “force” of argument; in other words, the bargaining elements, or “reasons”, on the basis of which the parties, each for itself, come to agree on a certain compromise (i.e. elements such as the interests, needs and preferences of each, as well as the threats and promises exchanged between them) do not really play the role of elements raising or aiming to redeem validity claims and thus being objects in a process of communicative argumentation; rather, they play, even if in an equitable way, the role of facts exerting causal influence upon those involved.

In this conception of fair bargaining we witness, again, the co-presence of the two aspects which, as we have seen, are merged inextricably together in Habermas’s concepts of the communicative and the strategic: on the one hand the aspect of action orientation (here we have an orientation to success) and on the other hand

the aspect of the relationship involved in the interaction (here we have a causal-manipulative relationship). However, through the concept of fair bargaining Habermas at the same time goes some way towards breaking the bond between success-orientation and causal relationship, and thereby moving towards a concept of action and rationality which is at once success-oriented and communicative. He does that by linking, though indirectly, the success-oriented process of bargaining with the communicative Principle of Discourse and the Universal Pragmatics of language more generally.

In "Some further Clarifications of the Concept of Communicative Rationality" (Chapter 7 of *On the Pragmatics of Communication*, Habermas, 1998a) – henceforth *Clarifications* – Habermas introduces certain ideas which make possible the communicative treatment of forms of bargaining, and purposive (inter)action and rationality more generally, which in *Between Facts and Norms* and all previous works he assigns to the unavoidable strategic interplay of causal-manipulative influences. To these ideas we shall now turn.

3.3.2 "Weak" communicative action and rationality

It is a basic tenet of Habermas's communicative paradigm ever since its first, full articulation in *The Theory of Communicative Action* that language, when employed in its primary, communicative mode (rather than in the "parasitical", strategic mode), i.e. when employed in order to perform communicative actions in the form of speech acts, has the (illocutionary) aim of a speaker reaching understanding with someone (an addressee) about something: more precisely, about something in the objective world, something in the speaker's subjective world and something in the social world intersubjectively shared between speaker and addressee. This aim is pursued by means of explicitly or implicitly making, and if challenged

discursively justifying, or “redeeming”, validity claims of three kinds; namely, claims concerning (a) the truth of assertions about the objective world, (b) the truthfulness of representations of the speaker’s subjective world, and (c) the normative rightness of actions or the rightness of general norms of action. A speech act is said to be communicatively rational insofar as the validity claims it raises can be redeemed (see section 2.2.1, above).

In *Clarifications*, Habermas for the first time distinguishes between a *strong* and a *weak* form of understanding, hence between a *strong* and a *weak* form of communicative action and rationality. The strong form of understanding is agreement on the basis of the same, shared, actor-independent reasons, a situation for which Habermas reserves the term *consensus*. Weak understanding, by contrast, is a non-consensual form of agreement, one based on actor-dependent, i.e. not shared, reasons (Habermas, 1998a: 320–1).

Thus, here Habermas interprets as a form of (weak) communicative interaction what in his discussion of bargaining, in *Between Facts and Norms*, he understands to be a necessarily strategic exchange, ultimately based on causally operative power, even when that power is procedurally controlled so as to be equalized. Cases in which strong understanding/consensual agreement is aimed at are either factual statements, raising primarily validity claims to truth, or norm-regulated imperatives – that is imperatives that are either prescribed or prohibited by shared norms – which raise primarily validity claims to normative rightness. An example of the latter is “By the end of the month you must give me ten thousand pounds, as required by the contract you and I have signed”. Weak understanding, on the other hand, is, according to Habermas, sought in cases of (1) declarations of intention and (2) “simple” or “naked” imperatives, i.e. imperatives not governed by norms in the above sense.

In fact, if weak understanding attaches to speech acts appealing to actor-relative – i.e. not shared – reasons, a naked imperative, in the sense of an imperative not regulated by norms, may still aim at strong understanding; for such an imperative may be regulated by normatively permissible, yet *shared interests* or *values*, and thus still pursued in terms of actor-independent reasons. In the following discussion, we shall use the term “naked imperatives” in the sense of what can be referred to as *egocentric requests* that is requests, or imperatives, that are not regulated *either* by shared norms *or* by shared values. An analogous situation exists in relation to the claims to rightness in the sense of permissibility, which, as we shall see, are implied also by declarations of intention. Here too, the claims must be understood as ones concerning permissibility with regard to both shared norms *and* shared values. In drawing the distinction between strong and weak communicative actions, Habermas puts an explicit emphasis on *normative* regulation (“authorization”, “embeddedness”, etc). Sometimes, however, he explicitly uses as criterion for the distinction, the more adequate one of regulation of either a normative or an evaluative kind:

In strong communicative action, the participants presume... that they pursue their action plans only within the boundaries of norms and values deemed to be valid (*ibid.*: 328).

3.3.2.1 Declarations of intention

Consider the declaration of intention “I shall spend the whole of next year travelling around the world”. If this declaration is uttered with an illocutionary aim, what kind of understanding does the speaker aim to reach with the addressee? The validity claims raised by such a speech act are those of (a) truthfulness and (b) viability, i.e. (a) that the speaker is sincere in expressing the intention in question (the

intention representing the speaker's genuine preferences, themselves reflecting his needs, interests, values, etc); and (b) that it is factually true that the speaker can (has the necessary means to) do what he intends. According to Habermas, a validity claim to rightness on the basis of shared norms (or, we would add, shared values) is not supposed to be raised by the announcement of an intention. Thus, if the addressee comes to accept validity claims (a) and (b), she will have agreed that the speaker has good reasons to act according to his expressed intention, but she will not thereby have adopted those reasons as her own, i.e. she will not herself have good reasons for acting in the same way.

The speaker's reasons for having the intention are his own actor-relative reasons, not ones shared with the addressee. The acceptance by the addressee of the validity claims raised in the announcement of an intention amounts, in other words, merely to "taking the announcement seriously" (*ibid.*: 322). Agreement of this kind is described by Habermas as a form of weak communicative understanding.

Two points need to be made with regard to this analysis. The first concerns Habermas's view that declarations of intention do not raise validity claims to normative (or evaluative) rightness, this being the essential reason for considering them as forms of weak communicative actions. It is certainly true that such declarations do not raise a claim to the effect that the intention in question is right in the sense of being somehow *obligatory* for speaker and addressee on the basis of shared norms (or values). It is, however, reasonable to argue that they do raise the claim that the intention is right in the sense of being normatively (and evaluatively) *permissible* to have and express. For if it were not so permissible, it could be communicatively rejected by the addressee, with the agreement of the speaker, on the basis of shared norms (or values). Thus, the communication could be settled exhaustively on the basis of the same, actor-independent reasons, and would, therefore, not be counted as weak. It is in fact in this very same sense, of

rightness as permissibility, that factual statements are thought by Habermas to raise simultaneously validity claims not only to the effect that they are true and truthful, but also that they are rightfully made. Nonetheless, even if declarations of intention are understood to raise also validity claims to rightness in the sense of permissibility, their acceptance by the addressee still amounts only to taking the intention seriously, not adopting it as one's own.

The second point concerns certain confusions regarding the meaning of terms such as "same reasons", "actor-relative reasons" and "agreement", used in Habermas's analysis of declarations of intention (as well as of naked imperatives, as we shall see). An agreement to share the intention – an agreement *with* the intention – would indeed have to be arrived at by each party on the basis of his or her own different, actor-relative reasons. An agreement to take the speaker's intention seriously, however, must be made by both parties for the *same* set of reasons, namely those of the speaker. Admittedly, these reasons have a different status than the ones appealed to by agents testing validity claims to truth or rightness: Although they must be commonly appealed to by the two sides in order to test the seriousness of the speaker's intention, they are actor-relative, in the sense that, in the last instance, they are "posited" (and espoused, or "owned") by one of them, namely the speaker. Thus, in that sense, they have an asymmetrical relationship *vis à vis* speaker and addressee; whereas reasons concerning truth and rightness relate to the two sides symmetrically, being "posited" for both by the external and intersubjective/social worlds shared by them. This difference applies, of course, not only with regard to declarations of intention but in relation to all utterances aiming to express the speaker's subjective states and raising primarily validity claims to truthfulness (see section 2.4.2, above).

Nonetheless, this asymmetry does not warrant the view that communicatively oriented declarations of intention – and expressive utterances in general, for that matter – aim at agreement on the basis of different reasons; for the main agreements aimed at through such an utterance – the one concerning the seriousness, or truthfulness, of those utterances and those concerning issues of factual truth and normative/evaluative permissibility – can be sought in terms of the same reasons, even though in the case of the former those reasons are actor-relative. Nor is the characterization of these utterances as (somehow inferior) “weak” forms of communicative actions justified; for they are speech acts oriented purely to reaching understanding with the addressees, and in no way to causally influencing them; moreover they are capable of being redeemed by entirely discursive means.

In fact, Habermas explicitly acknowledges this much, concerning both announcements of intention and imperatives, despite his classifying both these forms of utterance as weak communicative actions:

Announcements and imperatives... move within the horizon of a mutual understanding based on validity claims and thus still within the domain of communicative rationality (Habermas, 1998a: 323).

In a similar vein, he speaks of agents interacting in terms of these two forms of utterances as “acting communicatively... renounc[ing] all intentions to deceive...” (ibid.: 327).

3.3.2.2 Naked imperatives

Let us now consider communicatively intended imperatives such as “(Please) lend me ten thousand pounds”, “Can you please close the window” and “I would appreciate it if you helped me with this report”, on the assumption that they are not normatively or evaluatively governed – i.e., not imperatives claimed to be either impermissible to issue, or based on a contractual or other obligation of the addressee – but “simple” or “naked” imperatives, in Habermas’s terms. Naked imperatives, Habermas explains, are such that “[t]heir illocutionary meaning consists in the fact that the speaker wants to motivate an addressee – that is, *another person* – to bring about “p”’ (*ibid.*: 322) – “p” being the propositional content of the imperative. More simply, here the speaker aims to (communicatively) get the addressee to comply with the (naked) imperative.

In this case, therefore, unlike the case of declarations of intention, it is required that speaker and addressee both come to agree *with* the imperative. The addressee is required to adopt the imperative as his own, just as the one who issues the imperative does, and thus to agree with the latter *in that sense*. But speaker and addressee are expected to agree with the imperative for their own, different, actor-relative reasons. The speaker’s reasons for asking for a loan will be different to the addressee’s reasons for giving the loan. Normally, the one who issues the imperative “assumes the availability either of sanctions, should the desired action fail to be carried out, or of rewards, if it is carried out” (*ibid.*: 323). These will count as reasons for the addressee’s compliance. Sanctions and rewards may also be explicitly threatened or promised, while rewards can also be assumed or explicitly demanded by the addressee, as a condition for complying with the imperative.

Cases in which it is reasonable to assume that complying with the imperative would be intrinsically satisfying for the addressee or that not complying would be intrinsically unsatisfying to him (e.g. when one is requested to make a donation for a cause which the addressee considers worthy) can also be interpreted in this way. By making such a request the speaker implicitly promises the addressee the satisfaction she believes to be intrinsic to his compliance and threatens him with the dissatisfaction (e.g. feelings of guilt) which she believes would necessarily accompany his non-compliance.

As with declarations of intention, Habermas maintains that in issuing naked imperatives with a communicative intent, speakers raise, and promise if challenged to redeem discursively, (i) validity claims to *truthfulness*, concerning their own actor-relative reasons – i.e. subjective states such as interests, values etc. – for issuing the imperative (together with any accompanying promises or threats), and (ii) validity claims to *truth*, concerning the existential statements they make or imply regarding the feasibility of the actions which they asked the addressee to carry out, as well as those they promised or threatened to carry out themselves in return. It might be added – for the same reasons as in the case of declarations of intention – that with naked imperatives speakers also make claims to the normative (and evaluative) rightness, in the sense of permissibility, of their requests, promises and threats.

But over and above these claims, here – as Habermas stresses – the speaker implies also that she has a right to “expect the addressee to comply with her imperative” (*ibid.*: 322); and she has this expectation on the basis of interests, values and other actor-relative reasons which she attributes to the addressee. Thus,

[a]n imperative is rational only if... the actor has good reasons for supposing that the addressee has reasons not to oppose her imperative (Habermas, 1998a: 22–3).

The speaker's attribution of reasons to the addressee will naturally be tested, in the last analysis, in terms of the addressee's truthful expression of his or her own subjective states. For this reason Habermas apparently assumes that the speakers' expectations of addressees' compliance to their imperatives are covered by the validity claim to truthfulness. However, this does not seem to be an adequate account of the Universal Pragmatics of the situation. If the rational reconstruction of speech acts is to be carried out in terms of validity claims exclusively made by the speakers themselves, a (critical) aspect of an imperative such as the speaker's expectation of compliance by the addressee cannot be accounted for in terms of validity claims to truthfulness made not by the speaker but by the addressee.

In fact, the claim implicit in the speaker's expectation of compliance by the addressee is one concerning *the truth of the speaker's beliefs about the subjective states of the addressee*. If this is so – and given that such claims cannot be assimilated to those concerning truth about the external world – Habermas's Universal Pragmatics fails to account for an important type of speech acts such as (communicatively oriented) naked imperatives. It therefore seems necessary to supplement it, at least with regard to this type of speech acts, with an additional form of validity claim, one pertaining to the truth about the other's subjective states – let us refer to it, for the sake of brevity, as the validity claim to cross-subjective truth. With regard to “world relations”, which are said to be implicit in speech acts by virtue of the validity claims made in them, this form of validity claim would seem, on the face of it, to add the subjective world of the addressee to

Habermas's Kantian trinity of (i) the external, objective world, (ii) the shared social world and (iii) the speaker's subjective world.

These suggestions, however, will have to remain tentative, in view of further discussion of Universal Pragmatics in section 3.5 below, where they will be incorporated in a more comprehensive proposal for revising Habermas's model.

Finally, just as in the case of declarations of intention, so in that of naked imperatives, the term *weak* communication seems to be inappropriate. For here too, the interaction is oriented purely to reaching understanding, or agreement, and is amenable to an exclusively communicative form of rational treatment: with regard to its validity claims to truth and rightness, it is amenable to communicative rationality in terms of the same, actor-independent reasons; and with regard to its validity claims to truthfulness and what we have called "cross-subjective truth", it can be rationally treated in terms of actor-relative – yet still the same for all participants – sets of reasons.

3.4 COMMUNICATIVE NEGOTIATION: FULLY OR RESTRICTEDLY SUCCESS-ORIENTED COMMUNICATIVE INTERACTION

The preceding discussion has obvious implications for our central inquiry into the conceptual possibility of action (and hence of rationality of action) which is at once purposive, in the sense of success-oriented, and communicative, in the sense of oriented to understanding, as opposed to causal influence. In somewhat more precise terms, it has implications for the possibility of pursuing through others what we have earlier called egocentric goals – i.e. goals not regulated either normatively or evaluatively – by purely communicative means.

Certainly, the issuing and responding to what have earlier been referred to as naked imperatives involves the reciprocal pursuit of normatively non-regulated (i.e. normatively permissible but not obligatory) goals by means of the actions of other agents. And in so far as this pursuit is oriented to mutual understanding, the action concerned – or, better, the *interaction* concerned, given the strong element of reciprocity which is inherent in these cases – is at once success-oriented and communicative. Habermas himself is quite explicit about the necessary conjunction of what he conceives as “weak” communicative action with success-orientation and purposive rationality.

In the weak forms of the communicative use of language and communicative action, communicative rationality is entwined with the purposive rationality of actors in an attitude oriented toward success (Habermas, 1998a: 329; see also *ibid.*: 323, 326).

Thus, through his analysis – in *Clarifications* – of naked imperatives as conceivably communicative interactions, albeit weakly so, Habermas has, in effect, established the conceptual possibility of fully communicative success-oriented action and rationality. By extension, he acknowledges the possibility of communicative bargaining, or negotiation, for bargaining/negotiation processes consist precisely in the reciprocal pursuit of egocentric goals through others. By the same token, this analysis negates the exhaustive division between *purposive/strategic* (i.e. success-oriented and causally effective) and communicative (i.e. norm and value-oriented and dialogical) action and rationality, or, alternatively, between *techne* and *praxis*, which, as we have seen, has been central to Habermasian theory from its early formulations in the 1960's to the most definitive statements of the later, communicative paradigm. For the first time in the Habermasian framework, actions are conceived which cut across this divide, being at once communicative, in the (restricted) sense of dialogical, and purposive, in the (restricted) sense of success-oriented.

At this point one might be tempted to conclude that all interactions where egocentric goals are mutually pursued through others – all cases of bargaining, or negotiation, all social exchanges where coordination of success-oriented actions is necessary – could be dealt with, by the parties involved, in a communicative way, i.e. through raising validity claims and if challenged justifying them discursively. And this would mean that there is no need – no conceptual necessity – ever to resort to strategic action in order to meet any such action coordination challenges: no conceptual necessity for surrendering to (systems of) strategic action huge areas of social interaction such as the economy and state administration; or, for so treating those success-oriented elements of social, including lifeworld, activity which Habermas, in *Between Facts and Norms*, recognized as necessary aspects

of lawmaking processes and for the sake of which he introduced the notion of bargaining – fair bargaining in particular.

Such an inference, however, would be too hasty, for the following reason. What the argument so far has tried to demonstrate is that success-oriented communicative interaction is conceptually possible; not that all cases of success-oriented interaction can be subsumed under that concept – i.e. that all the efforts of agents to pursue egocentric goals by means of each other can be coordinated successfully in a purely communicative manner – or even that it is conceptually certain that there is one single case that belongs to that category (though on empirical grounds it can be readily testified that many such cases are likely to be encountered). To clarify this point, it is necessary to focus on an element of naked imperatives/egocentric requests (which is thus an element of negotiation situations in general) which, though present, remained unthematized in the preceding analysis.

In issuing a (communicatively oriented) naked imperative the speaker assumes (a) that she has her own actor-relative reasons for doing so, and (b) that the addressee has his own actor-relative reasons to comply. The crucial point here is that the two actor-relative sets of reasons are supposed to be reasons for wanting *one and the same thing*: the addressee to act in a particular way, namely, in accordance with the imperative. The speaker, therefore, in issuing the imperative is implicitly saying to the addressee, “I believe that you have reasons to do what I have reasons to want you to do”.

This is a significantly different kind of claim to (i) “I believe that you have reasons to do X, and I have reasons to want you to do Y”, and to (ii) “I believe that we both have shared reasons to do Z”. In cases (i) and (ii), even though any particular claims of these types may prove to be wrong, it is certain that for each type of

claim there is an instance on which the agents involved could in principle come to agree and thus coordinate their respective actions successfully: Some possible beliefs of the speaker concerning his own actor-relative reasons, the addressee's actor-relative reasons and reasons shared between him and the addressee, respectively, must be true, on the assumption that there exist some speaker-relative reasons, some addressee-relative reasons and some reasons shared by the two. This is not so, however, with regard to claims of the type "I have reasons to want you to do X and I believe you have reasons to want to do X", which, as we have seen are raised in negotiation situations, including naked imperatives. Similarly, assuming that there exist a set of speaker-relative reasons and a set of addressee-relative reasons, there is no guarantee that *any* claim of this type is valid. It is in principle possible that any actor-relative reasons of two or more agents would not converge on (i.e. simultaneously justify) one and the same action. For it is conceivable that in any communicatively oriented negotiation no deal is found which simultaneously satisfies the different interests, values, etc. of the parties involved, and which it is, thus, *purposively rational* for each party separately to accept. Such could be the case when at least one side believes that the best deal for itself among the deals acceptable to the other side(s) is not purposively good enough for itself. In such a case there would be no point of coincidence between what the sides find (purposive-rationally) acceptable to themselves, thus no mutually agreeable deal.

If the above argument is correct – if it is indeed the case that in any negotiation, i.e. in any situation of mutual pursuit of egocentric goals through others, it may be impossible to reach a communicative agreement, hence attain the required action coordination – then, it may be necessary to supplement success-oriented communicative action with strategic action in order to meet the multifarious coordination needs of egocentric social action. This conclusion has been drawn on

the assumption that the success-oriented action in question is carried out by the parties in what, from their point of view, is a *fully* purposively rational manner. It has been argued, in other words, that it may be impossible for agents to coordinate their actions communicatively – and thus it may be necessary for them to resort to strategic action – when they pursue their egocentric goals through each other with an orientation to gaining the *maximum benefit* possible for themselves. To be sure, as already mentioned, such maximization – thus full purposive rationality – is in fact attainable by communicative means in plenty of cases; and the scope of such a possibility can be considerably expanded with skilful negotiation, as will be seen below, in section 3.6.3. Nonetheless, the possibility of failure in this respect – and therefore of the need to seek maximization of benefit (i.e. full purposive rationality) through strategic action – cannot be eliminated. In other words, the notion of success-oriented communicative action cannot be said to cover fully the whole of Area 3 in Figure 2, above.

In order to render possible the communicative coordination of *all* mutually success-oriented action – thus generalize the communicative approach to all negotiation cases, and make it in principle unnecessary to employ strategically oriented action – we need to revise the concept of fully purposively rational communicative action into one of purposive-rationally restrained communicative action. With the latter, we stop insisting on the absolute maximization of agents' satisfaction, in favour of aiming only at the maximum satisfaction which it is possible for agents to agree on by purely communicative means.

Interactions of this kind which fail to communicatively yield the maximum possible benefit for all parties – which thus fail to attain full purposive rationality – would still be *success-oriented* communicative interactions, though imperfectly so, for they

would still pursue egocentric goals, rather than goals regulated by shared norms or values. By the term *communicative negotiation* I shall refer to *prima facie* success-oriented communicative interactions insofar as they are restrained in the above sense, regardless of whether they in fact achieve full or only imperfect purposive rationality. The qualification “prime facie” refers to the possibility that what initially appear as egocentric goals which need to be pursued by means of the success-oriented method of negotiation, may, in the course of the negotiation prove to be in part – or even fully – regulated by shared norms and/or values (see section 3.6.3, below).²

Respecting the restrictions imposed by the requirements of communicative action means that in no part of a communicative negotiation would any of the parties involved seek – through employing strategic action – satisfaction of its own egocentric goals beyond what is communicatively justifiable, i.e. beyond what is acceptable to the other side(s) on the basis of each raising, and if challenged discursively redeeming, validity claims. It means, therefore, that none of the parties, at any point in the negotiation, gives priority to the satisfaction of its own egocentric goals over those of the other(s). Thus, in order to generalize the (in principle) applicability of the notion of success-oriented communicative (inter)action over the whole of Area 3 in Figure 2, above, so as to rule out altogether the need for the strategic coordination of actions aiming to realize egocentric goals by means of other agents, it is necessary to introduce the principle of (recognition and) *equal treatment* of each other's egocentric goals in every negotiation; a principle, in other words, of not according priority to the satisfaction of one's own interests over the satisfaction of the interests of the other parties. Its role is to facilitate a communicative conclusion to a negotiation where an unrestrained, fully purposively rational orientation would lead at least one of the parties to abandon communicative interaction in favour of a strategic approach.

Of course, the introduction of the principle of equal treatment in the above sense, if it is not to be an arbitrary, ad hoc move, needs to be itself philosophically grounded. In this respect, I shall follow Habermas who, drawing on Alexy, considers a principle arguably identical to the one discussed here – one that also underlies his own (moral) universalization principle – to be a fundamental presupposition of discursive argumentation (see Part Two, section 2.4.3 and note 12, above). Indeed, a process such as communicative discourse, which repudiates all coercion and deception, relying solely on the “force” of the better argument, cannot but treat equally the *reasons* relevant/admissible to the form of discourse concerned. And, whereas in discourses concerning factual truth the perceptions of participants are among the reasons that can legitimately be put forward, in the case of communicative negotiations – that is, discursive processes concerned with the rational acceptability of negotiation proposals – among the admissible reasons are the interests, values and purposes, including the egocentric goals, of those involved.

The concept of communicative negotiation invites comparison with Habermas’s idea of fair bargaining, discussed earlier. For in both, values, interests, preferences etc, are accorded equal chance of being heard and taken into account. But at the same time the two differ in important respects: whereas in fair bargaining the “equal treatment” of these elements is effected through the equalization of the power of bargaining agents, which in no way alters the factual, causally effective nature of those elements themselves, in the case of communicative negotiation the parties treat each other’s interests as mutually respected, i.e. in a sense “shared”, reasons, which are to be appealed to in a communicative, discursive process. On these grounds, communicative negotiation can be seen as a communicative analogue of strategically conceived fair bargaining.

According to the above account, even though it is possible that a communicatively attainable negotiation deal might not be fully purposively rational for a certain participant, it is also possible that in some negotiation cases it might be. As we shall see later (section 3.7) when the concept of communicative negotiation is further elaborated, this can be the case when a communicatively attainable negotiation deal either (maximally) satisfies all the relevant aims of the participant in question, or constitutes a greater or lesser compromise on that participant's maximum relevant aims, being nonetheless as good as or better than any of the deals which that participant considers attainable by strategic means. In other words the possibility exists that for an agent it is preferable from an unrestrictedly purposive-rational point of view to approach a given negotiation *communicatively* rather than *strategically*.

Such cases, it should be pointed out, can be accounted for also in terms of Rational Choice Theory [see, for example, Raiffa et al (2002), drawn upon in section 3.7 below] and do not therefore require for their justification the complex edifice of Habermasian universal pragmatics. Within the broad Rational Choice paradigm, for example, work has been done that claims to show how “cooperation” – i.e. non-coercive, non-manipulative interaction– can arise among fully purposively rational agents (see Axelrod, 1984). A Rational Choice account would, in addition, provide a satisfactory solution to the problem of motivation for non-coercive, non-deceptive social interactions, relying on the strong motive force of needs and interests rather than on what Habermas himself has admitted is reason's “weak motivational force” (1996: 164).

As will be seen in section 3.7 below, with skill the range of cases admitting of fully purposive-rational communicative solutions, and thus the scope for the application of Rational Choice Theory, can be considerably expanded. With regard to these

cases Habermasian theory could at best be credited for elucidating the idea and methodologically enriching the practice of human interaction which is not deceptive, manipulative or coercive.

This being the case, however, a non-fully-purposively-rational justification of cooperation, thus of communicative negotiation – and Habermasian-type universal pragmatics is as good a justificatory line as can had – is still necessary; for as we have seen it is always possible to encounter negotiation cases where the actor-relative interests of the parties cannot converge on one and the same solution, and which, therefore, can be communicatively dealt with only in a restrictedly purposive-rational manner.

Furthermore, a non-Rational-Choice justification of negotiating communicatively – such as the Habermasian universal pragmatic one, encompassing, as indicated above, a principle of equal treatment of egocentric interests – is necessary to be operative as a presupposition in *all* communicative negotiation cases, not only in those which do not admit of fully purposive-rational solutions. The reason for this is that it is not possible for participants to know in advance – often it is not possible to know until the very end of the negotiation – whether a particular negotiation can be (communicatively) resolved in a fully purposive-rational way, in which case it would not require the presupposition of equal treatment.

3.5 A RECONSIDERATION, IN THE LIGHT OF THE CONCEPT OF COMMUNICATIVE NEGOTIATION, OF MORAL AND NON-MORAL NORMS AND RIGHTNESS AND THEIR UNIVERSAL-PRAGMATIC GROUNDING

3.5.1 Applications of the universalization principle as communicative negotiations

In *Moral Consciousness and Communicative Action* (1990) Habermas, as we have seen in the previous Part (section 2.4.3), offers his version of the *universalization principle* as a rule of argumentation suitable for testing the validity of moral norms. The latter are understood, along Kantian, deontological lines, as universal and unconditional prescriptions of right and wrong or, alternatively, of what is obligatory and what is impermissible, what ought and what ought not to be done by everyone in all circumstances.

According to Habermas's formulation of the universalization principle, a moral norm is valid only if "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 (and these consequences are preferred to those of known alternative possibilities for regulation)" (Habermas, 1990: 65). The application of the universalization principle for testing a candidate moral norm, therefore, involves (a) the recognition of each other's interests, and (b) the raising, and if necessary the discursive support by each participant, of validity claims to the effect that the interests of those concerned constitute sufficient reasons for each to accept the moral norm under consideration. (These claims imply further claims to what we have earlier – in section 3.4 – referred to as cross-subjective truth, that is truth about others' interests, and subjective states more generally).

Evidently, the application of the universalization principle, which Habermas describes also as “a cooperative process of argumentation” (*ibid.*: 67), has the same structure as what I have called “communicative negotiation”. In fact every such application can be construed as communicative negotiation on a grand scale: a multiple, multi-party communicative negotiation, analyzable into a great number of component, multi-party communicative negotiations. In an application of the universalization principle – in order to test a candidate universal, moral norm, e.g. the norm “Everyone should always keep their promises” – all those who would have to abide by such a norm if it is endorsed (that is all human beings, or even more generally, all thinking and acting beings) must (potentially) participate, and give their considered, discursively arrived at positive or negative verdict on the norm in question.

To decide on their verdict, participants have to imagine/hypothesize finding themselves, in turn, in a variety of situations (in an as complete as possible range of communicative interaction situations, S_1, S_2, \dots, S_n) relevant to the norm being tested. With regard to the above mentioned norm, a relevant situation, S_1 , would be whether in the context of circumstances C_1 , agent A_1 would keep promise P_1 which he had given to agent B_1 ; where circumstances C_1 might include the conditions under which A_1 had given the promise, A_1 's family responsibilities, B_1 's business and family circumstances, the importance for A_1 of his reputation in the community, etc. Another relevant situation, S_2 , might be whether agent A_2 would keep her promise P_2 to agent B_2 , in the context of a different set of circumstances, C_2 . For each such hypothetical situation, participants should, in a process of communicative argumentation, (a) calculate the likely consequences and side effects the general observance of the norm under consideration (i.e. its observance by all in all instances of the situation) would have on each of the agent-roles concerned (in our example S_1 , roles A_1, B_1, A_1 's and B_1 's family

members, business associates, etc.); (b) imagine everyone in turn in each of the different roles, and, on the basis of what are discursively established as the genuine interests of each, decide (i) whether they would all accept the observance by all, of the norm being tested, in all instances of the situation, regardless of the role they would occupy in them, and (ii) whether they would all prefer such observance to any other possible (also communicative) form of resolution of the interaction problems in question. If all participants decide affirmatively on the general observance of the norm being tested in (all instances of) all the relevant situations (S_1, S_2, \dots, S_n) the norm in question is thereby established as a valid moral norm.

Discursive consideration by participants of each hypothetical situation relevant to the candidate moral norm, as described above, can be understood as a (multi-party) communicative negotiation among all human beings, in which the proposed deal consists of the observance by everyone of a certain norm in all the instances of the specified situation. The overall application of the universalization principle can, similarly, be seen as the summation of all the particular, situational communicative negotiations. Such a summation is concerned with whether the outcome of all these component negotiations is the *same* affirmative one regarding the general observance of the norm being tested.

3.5.2 The contingency of norm-based morality

Habermas seeks to provide justification for the universalization principle in terms of his Universal-Pragmatic theory of communicative action, arguing, as explained in section 2.4.3, above, that it is presupposed by the very structure of linguistic communication, in particular by the validity claim to rightness which is said to be inherent in it. As suggested in that section, if a (universal) moral norm is to be tested, it can indeed be reasonably said that the most appropriate way of doing

that is by letting all those who would be affected by the tested norm's acceptance negotiate such an acceptance among themselves, on the assumption that the structure of their communication already commits them to treating each other's interests equally; which is precisely the test being given to moral norms by means of the universalization principle. However, to demonstrate in this way the appropriateness of the universalization principle for testing universal, moral norms – even if that appropriateness is shown to be exclusive – does not amount to deriving from the structure of communicative action (a) that the universalization principle will ever need to be applied to test any norm, or (b) that any application of the principle that might take place will lead to the endorsement of a universal, moral norm; i.e., that any candidate moral norms will ever be validated.

In fact, propositions (a) and (b), above, cannot be derived from the structure of communicative action. For the latter does not *a priori* determine the content of the communication that will actually take place between human beings. From the point of view of the structure of communicative action, it is an entirely contingent matter whether any specific speech act will ever be undertaken, and thus whether any specific negotiation will ever actually be called for or carried out. This, of course, applies equally to the universal (hypothetical) communicative negotiations that constitute – as explained – applications of the principle of universalization. It cannot thus be *a priori* determined whether any particular norm of action will ever become object of such universal negotiations, thus a candidate moral norm.

Furthermore, even if such a negotiation over a candidate moral norm is undertaken, there can be no guarantee that all its component negotiations, regarding different situations relevant to the norm in question, will yield the *same* positive result, thus establishing the norm under consideration as a valid moral norm. Again, how each component negotiation will be resolved is entirely contingent. What *can* be *a priori* guaranteed, according to our earlier analysis of

communicative negotiations, is that – since in such negotiations the effort to communicatively resolve the negotiation would never be abandoned in favour of a strategic approach – each component negotiation *will*, for sure, deliver a solution, though not necessarily an affirmative one.

Thus, the above argument shows that it is not implicit in the concept of action oriented to reaching understanding that the universalization principle would be invented, or that, if invented, it would, or should, be applied, or that if applied it would lead to universal agreement on any moral norm. It is not, in other words, guaranteed by the nature of communicative action that any moral norms – in the sense of universal norms established by means of the universalization principle – are either necessary or possible. This, of course, does not exclude the possibility that such norms are, contingently, so established.

In the above argument about moral norms, the qualification “(norms) established by means of the universalization principle” is important. For norms of this kind must be distinguished from morally loaded norms which – as Habermas and others rightly maintain – are derivable from the structure of linguistic communication as rules of argumentation, binding in contexts of discourse. One such rationally reconstructible rule, which is stressed by Habermas, concerns the exclusion, in all discourse, of “all external or internal coercion other than the force of the better argument” (1990: 89).

Evidently, the possibility of a moral norm, in the sense of a norm that has received its validation by means of the universalization principle, depends on two factors: the ethical/evaluative choices of human beings, and pragmatic/means-ends relationships between these choices and possible human actions. More precisely, it depends on the existence of such human interests that their communicative/non-strategic realization for each person can best be secured by means of the

universal observance of a single, relatively simple, rule of action. Such interests might be described as *universalizable interests*.

The existence of at least some such universalizable human interests must therefore be presupposed by any confident expectation that the application of the universalization principle will succeed in establishing some universal norms of action. Of course, the truth of such a presupposition cannot be known *a priori* through the Habermasian rational reconstructive method. It can only be established *a posteriori*, as for instance and primarily by applying the universalization principle itself, i.e. through carrying out universal, multiple negotiations in which interests are asserted and mutually recognized by participants themselves.

Nonetheless, an anthropological assumption that there are universalizable human interests is implicit in Habermas's thinking on the universalization principle. In this respect, it would be interesting to compare this assumption with that of the *uniformity of nature*, which underlies the Principle of Induction. For Habermas considers the Principle of Induction to be a "bridging principle", in the domain of empirical science, between "particular observations and general hypotheses", analogous to the principle of universalization, which he takes to be the equivalent bridging principle between particular and general in the area of moral discourse (Habermas, 1990: 63). This analogy, however, raises issues which cannot be discussed here.

3.5.3 The social functionality of norm-based morality

A norm that has passed the universalization test has two distinct functions: (1) It defines a certain class of interactions between human beings (e.g. interactions involving promises) and prescribes a particular, standardized way of resolving every interaction belonging to that class, if the optimum long-term satisfaction of

certain interests is to be achieved for all human beings (e.g. it prescribes that, if a promise is given, it must be kept). (2) It specifies the way in which such an optimum outcome should be pursued, namely by applying the norm in question.

Now, let it be pointed out that function (1) does not entail function (2). It is in principle possible that the type of (communicative) resolution specified by the norm as the one that maximizes long-term interest satisfaction for all, is achieved not by applying the norm itself but by direct communicative negotiation (in the absence of a norm) of each single relevant interaction, or perhaps by the communicative negotiation of particular clusters of such interactions; such a cluster might consist, for example, of all the promise-involving interactions between two particular agents over a specified length of time, including their entire life. But such a non-norm-based way of pursuing the optimum result specified by a moral norm is rejected in the course of the norm's discursive endorsement through the application of the universalization test, in which the method of norm-application is adopted.

An advantage of the norm-based approach is immediately apparent: It is highly improbable that a potentially infinite number of communicative negotiations would result in the same type of deal – and that, that deal would be the one specified by the relevant norm. Thus, it would be exceptionally unwise to rely on such a precarious method – rather than on consistently applying a (potentially) available norm – for securing the desired outcome. Moreover, further reflection reveals that not only is it improbable to achieve the optimum deal in every communicative negotiation relevant to a moral norm, but that it is virtually certain that such a deal would not be obtained in any negotiation of this kind. This is so for the following reason: In cases where an interaction is resolved in accordance with the requirements of a relevant moral – or indeed non-moral – norm (e.g. by keeping a promise, paying a debt, or making an honest declaration of income to the tax office) there are, typically, *immediate* “winners” and “losers”. The loser acts – and

those “adversely” affected, e.g. his or her family, business associates, etc. agree that she acts – according to the relevant norm on one understanding: that the winner(s) in the particular interaction as well as all others in similar encounters with them (the present losers) are committed to observing the norm in question, including when they are themselves in the role of the (immediate) loser. But, in any direct communicative negotiation over the same particular interaction – now in the absence of a universally binding relevant norm – the equal treatment of the immediate interests of those involved would certainly not lead to a solution identical to the *optimal* one that would be prescribed by the appropriate moral norm; for that solution would leave (immediate) winners and losers. It would inevitably lead to some other, immediately equitable, solution.

Thus, tackling each (morally relevant) situation by means of a direct communicative negotiation would, typically, result, for each agent involved, in a level of overall, long-term satisfaction of interests incomparably lower than what would be attained by approaching the same situations by way of applying to them the appropriate moral norms. Lower overall interest satisfaction would also be achieved by means of limited, communicatively negotiated agreements of reciprocity between an agent and one or more other agents. It would indeed be achieved by means of such agreements of any scope, short of the universal one resulting from a universal negotiation process tantamount to an application of the universalization principle: an agreement committing everyone to universal observance of a (moral) norm.

In addition to being preferable in terms of overall interest satisfaction, moral norm-following is also a more socially efficient way of dealing with the relevant classes of interactions. For – even taking into account the (potential) requirements of moral norm-validation, and, more importantly, the (real) everyday difficulties regarding the application of moral norms (see section 2.4.3, above) – moral norm-following is

a relatively time- and effort-economical method of meeting the coordination needs arising in the relevant classes of interactions, in comparison to (communicatively) negotiating equitability in every specific case or for clusters of cases.³

On the above understanding, morality concerns interacting human beings undertaking towards each other commitments to act in certain standardized ways in specified classes of interactive situations; and this, regardless of whether such actions would immediately be less beneficial than the outcome they would achieve through the equitable, communicative resolution of the same encounters, either singly or in clusters. These universal, standardized solutions to whole classes of interactions – i.e. moral norms – are adopted for the sake of communicatively optimum, long-term interest-satisfaction for each, and are available only contingently and only with regard to certain possible (universalizable) human interests. Insofar as such (normative, moral) solutions happen to be available, they also make a valuable contribution to society's procedural efficiency.

3.5.4 Moral and non-moral normative frameworks

The (possible) set of valid moral norms – i.e. of the norms that pass the Universalization test, as explained above – can be said to form a framework that specifies the set of actions/goals and by implication the set of interests or values that are morally regulated, i.e. whose realization is morally either obligatory or prohibited, and separates these from the actions/goals/interests/values whose pursuit, including their pursuit through others, is morally permissible, or optional, and thus open to possible negotiation between agents. The qualification “morally” (with regard to the obligatory, the prohibited and the permissible) signifies the context in which the normative framework in question applies: that context in which agents interact with each other purely in their capacity as human (or more generally thinking and acting) beings, in abstraction from any other determinations.

This is also the capacity in which human beings can participate in applications of the principle of universalization, that is, in discourses that determine moral norms. In Figure 4, the moral normative framework is graphically represented by a circle. The area inside the circle comprises what is morally regulated, i.e. obligatory or impermissible. The area outside the circle comprises what is optional, from a moral point of view, i.e. morally permissible but not obligatory.

The moral normative framework

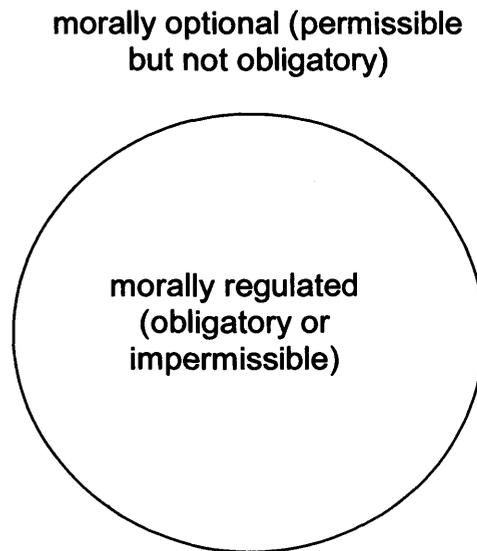


Figure 4

Within the area of what is morally permissible to pursue, there may be elements which, for specific, small or large groups of agents, constitute shared goals (including complex projects), or broader positive or negative shared interests or values, i.e. desirable or undesirable states of affairs, which themselves specify further shared goals, or rule out others, for the corresponding groups. Such goals, interests or values can be shared by a group of agents either spontaneously or following negotiation between its members.

Morally permissible goals are what we have earlier referred to as egocentric goals for any agent who adopts them, *vis à vis* those who do not share them with him or do not share with him values that entail or rule them out. A morally acting agent can pursue such goals by means of those who do not share them with him, in either of two ways: through communicative negotiation, or through causally-effective, strategic action, so long as such action is compatible with the moral normative framework.

For the sake of communicatively maximizing overall interest satisfaction for each, and doing so in a socially efficient way, agents interacting in capacities different to that of abstractly conceived human being can, similarly, determine respective normative frameworks. For example, in their capacity as citizens of a particular state, or as members of a particular political, administrative, economic, religious, sports or other organization, including a criminal gang, they can, respectively, construct a constitutional-legal framework for their state, or a normative framework for their organization. The same can be said of collective agents, such as states or other organizations, who interact with each other in those capacities, possibly as members of international conglomerations of states such as, for example, the European Union or the United Nations, or as members of associations of organizations, respectively. These non-moral normative frameworks also specify what is obligatory, impermissible and permissible in their context of application. They similarly allow for the differentiation, within the normatively permissible, between goals that are, for any set of interacting agents, shared either directly or by being regulated by shared values, and goals that are egocentric and thus only pursuable either through communicative negotiation, or through strategic action insofar as the latter is permissible within their context. To be sure, normative frameworks can interlock with each other in complex ways.

The communicative determination of non-moral normative frameworks should take place by processes analogous to applications of the universalization principle, that is multiple communicative negotiations among all those involved in, or otherwise affected by, a certain interactive context (e.g. all the inhabitants of a certain state, all the members and other stakeholders of a certain organization, etc.). The possibility and scope of such frameworks depends on the existence of interests (values, needs, desires, goals) which are, within the given context generalizable by analogy to the universalizable interests that make normativity possible in the abstractly human, moral context.

3.5.5 Inadequacy of the instrumental justification of moral and non-moral norms

The conception of morality and non-moral normativity that emerges from the above analysis, is, therefore, that of a concern with the maximum satisfaction, on the basis of equal treatment, of non-negotiable interests generally embraced within any particular context. In this conception, the aspect of equal treatment is rationally reconstructible from the nature of linguistic communication, and its denial constitutes a performative contradiction; whereas the existence of generally embraced non-negotiable interests in any given context, remains a communicatively contingent matter, depending – according to context – on ethical choices as well as pragmatic considerations.

As explained in section 3.4, in negotiations it is conceivable, and often achieved in practice, that the best deal attainable communicatively – i.e. relying on the force of the better argument and respecting the principle of equal treatment of interests – is, from the point of view of purposive rationality, better than, or at least as good as, any deal attainable by purely strategic means. Insofar as any given

communicative negotiation proves to be fully purposively rational, in the above sense, the communicative approach to that negotiation is justifiable instrumentally – in line with Rational Choice Theory – and a Habermasian, rational-reconstructive justification of it is in that case strictly speaking superfluous. However, the question whether any particular communicative negotiation can have a fully purposive-rational resolution cannot be answered ahead of the negotiation; for it depends on a complex combination of factors, such as the interests, relations of power and communicative and negotiation skills of participants, which can only unfold in the course of the negotiation, and fully reveal themselves at the end of it. For this reason, a non-instrumental justification of the communicative approach – such as the Habermasian justification – is still necessary for all negotiations.

This argument applies of course also to the negotiations that constitute applications of the principle of universalization/generalization for testing moral or other norms of action. Thus, although any particular, communicatively validated norm may contingently prove to be justifiable instrumentally in the sense that its general observance happens to satisfy the relevant interests better than any alternative form of regulation, including strategic ones, morality and non-moral normativity cannot be *a priori* so justified. Their communicative nature requires a non-instrumental justification, such as the Habermasian one, based on the rational reconstruction of linguistic communication.

3.5.6 From the validity claim to normative rightness to the validity claim to intersubjective acceptability

The concept of communicative negotiation and the interpretation of moral and non-moral norms as conclusions of complex, multi-party communicative negotiations,

makes possible a certain critical perspective on Habermas's idea of the validity claim to rightness and eventually leads to its extensive revision.

It has already been argued, earlier in this section, that the existence of moral or non-moral norms communicatively justifiable in terms of the application of the universalization or other generalization principles – as distinct from morally loaded rules of rational argumentation – cannot be *a priori* guaranteed, nor of course ruled out, by the structure of linguistic communication; for it cannot be determined *a priori* whether any particular norm-justifying communicative negotiation will take place, and if it does, whether it will lead to the endorsement of the norm in question. If this is so, a validity claim to rightness, in the sense of justifiability in terms of legitimate social *norms*, cannot be implicit in communicative action as Habermas maintains; it cannot be the case that every speech act implicitly raises and offers to discursively redeem such a claim. However, the intuition that led Habermas to postulate the validity claim to norm-based rightness, namely that one of the roles of every speech act is to coordinate action, or more generally to regulate social relationships, is fundamentally sound.

Thus, every speech act, in addition to raising validity claims (a) to the truth of the factual statement it explicitly or implicitly makes and (b) to the truthfulness of its explicit or implicit self-expressive content, must also raise, and offer to redeem discursively, some claim (c) about the “appropriateness” (in a sense of the term that needs to be explained) of the intersubjective relationship regulation it implicitly or explicitly proposes or assumes. Factual statements, such as (i) “(our tests show that) you have a gall-bladder stone”, and expressive statements, such as (ii) “I feel bored”, which thematize claims to truth and truthfulness, respectively, also raise, implicitly, claims to the (contextual) appropriateness of their being made. An analogous situation exists with regard to explicitly regulatory utterances, such as

(iii) "What he is planning to do is both wrong and illegal"; (iv) "If you buy two of these, I'll give you a discount of 20%"; (v) "I've got a flat tyre. Can you please help me?" and (vi) "Let's hurry up! We'll be late for the concert". These statements, over and above making implicit truth and truthfulness claims, thematize claims to the appropriateness of their regulatory content.

The regulatory "appropriateness" claims raised explicitly through the thematized content of these utterances may – as in example (iii) – appeal to moral, legal or other social norms which are presumed to be shared by, and legitimate for, a certain set of agents, thus determining practical commitments (obligations and prohibitions) for them; or – as in examples (iv) and (v) – they may constitute claims to the effect that the regulation proposed is acceptable to those affected – in these two cases, to speaker and addressee – on the basis not of any normative commitments of theirs but of the equal treatment their relevant egocentric interests, including, possibly, other-regarding motives; or, they may appeal to shared interests of the interlocutors, as in example (vi).

On the other hand, the regulatory claims raised by speech acts, mostly implicitly, about the "appropriateness" of their being performed in the given context, may constitute claims to normative commitments, as in example (i); or claims to intersubjective acceptability on the basis of egocentric goals, as in examples (ii) – (v), combined with claims to normative permissibility, insofar as these speech acts are performed within a normatively structured social context; or claims to common evaluative commitments on the basis of shared interests, perhaps also combined with claims to normative permissibility. Thus, the claims concerning the "appropriateness" of the proposed, or assumed, regulation of intersubjective relationships, which (claims) are inescapable presuppositions of every speech act, cannot be identified with Habermas's validity claim to normative rightness. Claims

to normative rightness, as we have seen, are contingent, with regard to the structure of speech acts, and would be contingent even if they proved to be present in every actual speech act, which may very well be the case, since speech acts tend to occur within normatively structured social contexts. Furthermore, claims to normative rightness, insofar as they are (contingently) present, are only one of three major types of regulatory claims raised in actual speech acts, alone or in conjunction with each other; the other two types being (a) claims to intersubjective acceptability on the basis of the communicatively necessary equal treatment of the relevant egocentric interests, and (b) claims to common evaluative commitments based on the coincidence of interests (either spontaneous or resulting from communicative negotiations).

What is universal-pragmatically inescapable, in other words, is that every speech act raises not a validity claim to normative rightness, but a claim to “appropriateness” in one or more of the three senses explained above; i.e., (a) a claim to normative rightness, (b) a claim to intersubjective communicative acceptability on the grounds of egocentric interests, and (c) a claim to common evaluative commitments based on shared interests. But valid norms, as argued earlier, are intersubjectively accepted, by those affected by their general observance, through processes that amount to large-scale communicative negotiations based on the equal treatment of interests; thus normative rightness is itself reducible to intersubjective, communicative, interest-based acceptability; while common evaluative commitments also signify interest-based intersubjective acceptability. It can therefore be concluded that speech acts inescapably raise, and offer to redeem by means of communicative discourse, validity claims to the *intersubjective acceptability* of their regulatory content. These claims imply the equal treatment of the interests involved, as these are (a major kind of) reasons relevant to the justification of the regulatory aspects of speech acts.

Such validity claims to intersubjective (communicative and interest-based) acceptability – a number of which might be simultaneously implicit in a single speech act – can take a direct form, appealing without mediation to shared or egocentric interests, or they can take the form of claims to normative rightness, thus appealing to interests in an indirect way, via (intersubjectively acceptable) moral or other norms.

Validity claims to intersubjective acceptability include, as explained in the discussion of naked imperatives, in section 3.2.2.2, above, a fundamental constituent claim to what we have called cross-subjective truth, that is, truth of beliefs about others' subjective states, particularly interests.

The term “rightness” can, perhaps, still be retained, and we can thus still speak of necessary validity claims to rightness, albeit on one condition: that it ceases to be used in the sense of *normative* rightness, and takes on the meaning of intersubjective (communicative, interest-based) acceptability, in the above sense. “Justice” may be yet another term that can be made to carry the meaning of intersubjective acceptability as defined here.

3.6 TAKING STOCK: THE HABERMASIAN AND THE REVISED CONCEPTUAL SCHEMES REGARDING PRACTICAL RATIONALITY

The preceding discussion has led to certain revisions of Habermas's understanding of action and practical rationality. Throughout his career Habermas has been operating with a two-fold distinction between *techne*, or purposive (instrumental and strategic) action/rationality, on the one hand, and *praxis*, or communicative action/rationality, on the other. The former prong of the distinction is understood to fuse together a means-to-ends, or "purposive" (in this narrower sense) or success-orientation of the subject and a causal-manipulative relationship between the subject and a (natural or human/social) object. The latter prong is understood to fuse an orientation of interacting subjects to determining values and norms, more simply, an ends-orientation, and a dialogical, or "communicative" relationship between them, in the stricter sense of communicative, referring to a mutually respectful relationship between agents oriented to reaching understanding.

This dichotomy has been replaced, in our discussion, by a four-fold distinction, on the basis of disconnecting questions concerning orientation, from those concerning relationship. Thus, we have differentiated between four types of action/rationality, as follows: (1) Means or success-oriented or purposive (in our narrower, orientation sense) *and* causal-manipulative, (2) oriented to values and norms *and* at the same time causal-manipulative, (3) means or success-oriented *and* dialogical, or communicative (in our narrower, relationship sense) and (4) oriented to values and norms *and* dialogical/communicative.

In addition to the above mentioned dichotomy, Habermas in his later work, i.e. in the work following his shift to the communicative paradigm, also proposes more differentiated models of practical rationality. Initially (in *Justification and*

Application) he develops a tripartite model of (a) moral, (including both moral justification and moral application), (b) ethical and (c) pragmatic forms of discourse. Later (in *Between Facts and Norms*) he puts forward a model of legal-political rationality that includes a fourth element, that of procedurally regulated, fair bargaining, in addition to the former three. Moral and ethical discourses are conceived by Habermas as processes of reflective communicative action, the former with an orientation to determining and applying moral norms, the latter with an orientation to clarifying values, and more generally conceptions of the good and personal and group identities. Pragmatic discourses, on the other hand, are conceived as purposive or success-oriented rationality processes involving causal-manipulative relationships, either between those participating in the discourse or between these and a third party, natural or human/social. These three basic forms of practical discourse are, according to Habermas, founded on the validity claims to moral rightness, truthfulness (in the sense of both sincerity and authenticity) and truth, respectively. These claims are associated with three respective world relations: those to an intersubjectively shared social world, to the inner, subjective world of the speaker and to an objective external world.

Fair bargaining is, in the last analysis, a strategic element of practical rationality, linked to the Discourse Principle and Universal Pragmatics only indirectly. Habermas represents this four-dimensional model, applicable in its entirety in political decision-making, in the diagram shown in Figure 5.

Habermas's model of "rational political will-formation"

(from Habermas, 1996: 168)

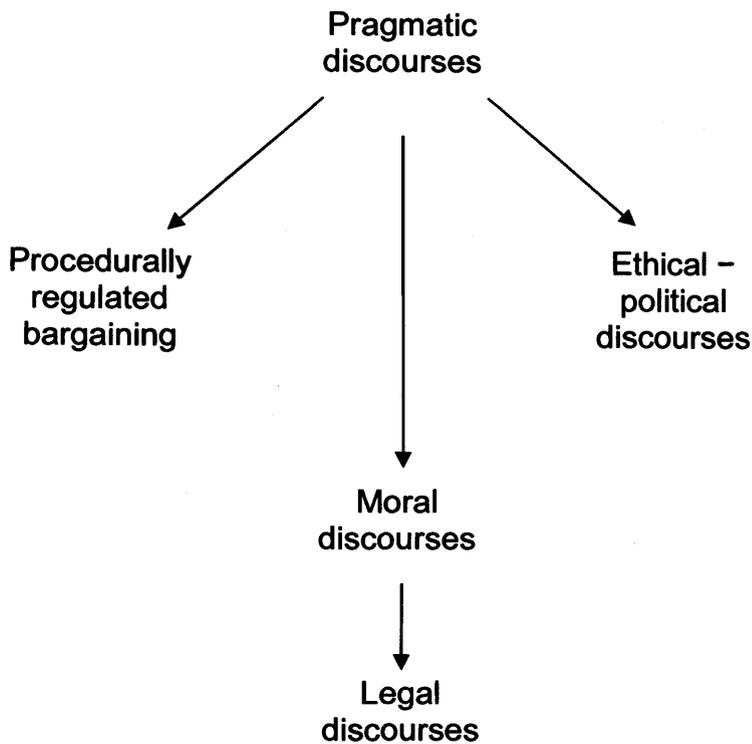


Figure 5

In the previous sections of this Part this model of practical rationality has, in effect, been revised in three main ways: (i) The category of success-oriented, yet communicative actions and discourses has been introduced. These, conceptualized also as communicative negotiations, are understood as processes in which interacting agents communicatively aim to use one another as means to their own ends including through composing collectively shared ends/interests/values out of diverse ones. (ii) Moral discourses have been relegated to a secondary, derivative status. Moral justification discourses, in the form of applications of the universalization principle, are considered to be special types of communicative negotiations (i.e. of success-oriented communicative discourses), moral application discourses being thus rendered dependent on the

latter. (iii) The Habermasian category of legal rationality, just as other non-moral normative forms of rationality, can, accordingly, be construed as a process that also relies on communicative negotiation: directly, by virtue of its significant elements of fair bargaining; and indirectly by virtue of its normative component. According to (ii) and (iii), therefore, moral and non-moral normative discourses are distinct forms of discourse only *prima facie*.

This revised model of practical rationality is founded on two of the Habermasian validity claims, namely those to truth and truthfulness, as well as on a new validity claim to “intersubjective acceptability”, an element of which is a claim to “cross-subjective truth”. The Habermasian validity claim to moral rightness, as well as any validity claims to non-moral rightness, are special forms of this, more inclusive, new validity claim. The validity claims to (objective) truth, truthfulness and intersubjective acceptability correspond, respectively, with world-relations to an objective, external world, to the subjective, inner world of the speaker, and to the intersubjective world of shared or communicatively reconcilable values, interests and concrete goals of the speaker and the addressees and others affected.

As will be shown below, there are important interdependencies between the different forms of discourse comprising this, revised model of practical rationality – and indeed this holds also insofar as these forms of practical rationality are components of the original Habermasian model. For a schematic representation of the revised model, see Figure 6.

Revised model of practical rationality

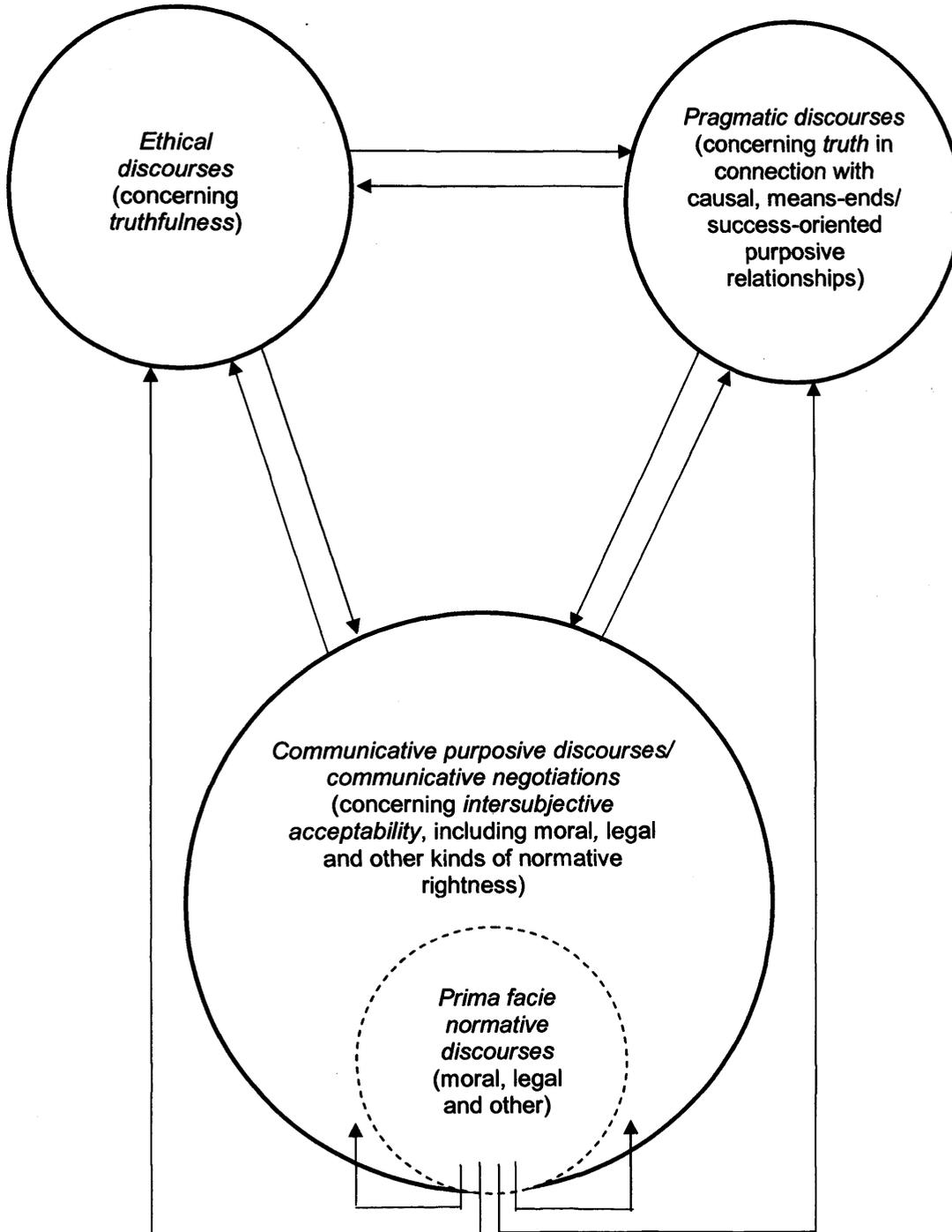


Figure 6

In the following section, 3.7, we shall discuss in more detail certain aspects of this revised model. Firstly, section 3.7.1 will consider the concepts of ethical, purposive (in both the causal-manipulative and the communicative sense) and *prima facie* normative discourses, discerning interdependencies between them. Secondly, section 3.7.2 will elaborate the concept of communicative negotiation, i.e., the specifically communicative form of purposive discourse to which the concept of *prima facie* normative discourse is, ultimately, reducible. The analysis will also bring to light the role played in communicative negotiation by the other three forms of practical rationality.

3.7 FURTHER ELABORATION OF THE REVISED MODEL OF PRACTICAL RATIONALITY: RECIPROCAL DEPENDENCES BETWEEN ETHICAL, PURPOSEIVE AND NORMATIVE DISCOURSES, AND THE CONCEPT OF COMMUNICATIVE NEGOTIATION

3.7.1 Ethical, Purposeive and *prima facie* Normative Discourses and relations of interdependence between them

This section is naturally closely related to section 2.4, which presents and critically examines Habermas's tri-partite model of practical rationality. The following discussion is an extension – and in some parts a summing up – of that in 2.4, now focusing attention mainly on issues of mutual dependence between the three forms of discourse.

3.7.1.1 Ethical Discourses

Both in the original Habermasian and in the revised model of practical rationality, ethical discourses deal with the clarification and determination of what is “good” for an individual or collective subject. When speaking of “the sphere of ethics”, i.e. that which ethical discourses are about, Habermas, in addition to the generic designation “good”, uses a variety of terms such as “values”, “ideals”, “conceptions of the good life”, “ego-ideals”, “needs”, “interests”, “(strong) preferences”, “inclinations”, “self-identity”, “intentions”, “ends” and “goals”. And if we specify further what comes under these terms, we can conclude that “the good”, for any particular subject, is likely to be a very complex and uneven ensemble of things.

In both the original and the revised version, the determination of the good for a certain subject rests, ultimately, on that subject's own *authentic* choices i.e. choices not involving any kind of self-deception or erroneous judgement. In other

words, validity claims as to what is good for a subject are, in the last analysis, redeemable only in terms of such choices. There is no guarantee that an ethical choice, even though expressed with sincerity, is *authentic* in the above sense. Thus, ethical discourses are concerned not simply with eliciting a subject's ethical choices, but with assisting the subject to make ethical choices that are free from error and self-deception. In that capacity ethical discourses deal also with additional validity claims for whose redemption third parties have a say equal to that of the ethical subjects themselves. Such validity claims, and the related ethical discourses, are concerned with three main criteria of authenticity:

a) *The internal coherence of the subject's conception of the good.*

What is positively evaluated – i.e. considered desirable – by a subject is normally an ensemble of diverse elements (states of affairs, objects, actions, properties) which can stand in various forms of relationship with each other. The main forms of such relationships are:

(i) Relationships between general and specific or abstract and concrete: a valued element can be a specific aspect of a more general value (e.g. freedom of speech in relation to liberty in general), or a concrete manifestation of an abstract value (e.g. my freedom to publish, today, my views on a particular current issue, in relation to freedom of speech, and, beyond that, to liberty).

(ii) Pragmatic relationships: some things are valued merely as means to other (valued) things. Money, for example, can be valued as means to such values as security, or social prestige; university education, as a route to a certain career. Some things can be valued both as means to other ends and as ends in themselves. Good health, for example can be valued for itself, and also as a means to other goods, such as longevity, or a sporting life.

(iii) Priority relationships: Some goods are valued more, or less, than others (i.e. preferred to others, or the opposite).

Validity claims explicitly or implicitly made with regard to such relationships can be tested in open discourses in which the subject and others have an equal say, with respect to such criteria as logical-conceptual consistency, evaluative consistency and factual truth. There has to be consistency between one's evaluations of something general and of its specific aspects. One cannot, consistently, value liberty and not freedom of speech, or do so to a disproportionate degree. Similarly, transitivity between preferences needs to be respected. If A is preferred to B, and B is preferred to C, then A must be preferred to C. Factual claims regarding relationships between means-values and end-values must also be discursively checked for truth; and where such means-ends links are not seen by the ethical subject concerned, they can be pointed out by others.

b) *Consistency between a subject's conception of the good and his or her behaviour.*

A subject's behaviour – which is in principle visible to all – can be tested for consistency with the subject's professed conception of the good in “objective” inquiries and discourses in which the subject does not have a privileged position *vis à vis* observers. Once aware of inconsistencies between their behaviour and their ethical choices, subjects should either reconfirm the ethical choices and change the behaviour accordingly, or revise the ethical choices, so that they are in line with the behaviour.

c) *Making ethical choices in full awareness of possible influences on them.*

Influences on a subject's ethical choices can be of three main kinds.

(i) Influences that occur as a result of a subject's specific life-history and socialization and which the subject can become aware of and acknowledge through ordinary hermeneutic processes. Such processes, for example, can make one understand – and possibly revise – her present religious affiliations and evaluative choices by enabling her to reflect on how these are linked to the religious family practices and school teaching of her childhood years; or lead one to review his present consumer habits through reflecting on the power of modern advertising techniques, and on his own early exposure to them.

(ii) Influences on a subject that are alleged by a third person to have occurred but which the subject does not acknowledge in the course of ordinary hermeneutic discourse. In such cases the subject refuses to accept either that an alleged biographical fact (e.g. a traumatic experience during infancy) has occurred, or that an (acknowledged) fact has had the alleged influence on his or her present subjective choices. In these cases the third-person allegations are either indeed false, or are true but the subject can come to accept them only through a process more radical than ordinary hermeneutics; such as a depth-hermeneutical, “therapeutic” process on the psychoanalytic model, combining hypothetical causal explanation by the “therapist” and emancipatory self-reflection by the subject/“patient”, as analysed by the early Habermas in *Knowledge and Human Interests* and discussed in Part One, above.

As we have seen, Habermas plays down this idea, together with the associated concept of ideology critique, due to its divergence from communicative discourse, and above all due to the authoritarian implications of the asymmetrical relation between therapist and patient. As a result, he deprives his communicative practical rationality model of valuable critical and self-critical potency (see section 2.4.2, above).

However, with the introduction of certain restrictions, depth-hermeneutic processes can, perhaps, be brought sufficiently close to fulfilling the conditions for communicative discourse as to be made admissible as possible aspects of ethical discourses, on a par with ordinary hermeneutics. Such restrictions are, (a) that only ethical subjects have the right to initiate depth-hermeneutic processes about their own choices (i.e. “therapeutic” processes in which they themselves are the “patients”); (b) that the “patient” has the right to terminate the “therapeutic” process at any time; (c) that ethical subjects have the last word concerning the alleged causal influences on their choices, and thus the privilege, in the last analysis, of determining their own authentic choices; and (d) that ethical discourse participants have the opportunity to play, symmetrically, both “patient” and “therapist” roles towards each other.

(iii) Influences exerted on a subject’s choices by his or her biological (genetic, hormonal etc) constitution. Such influences might have to do, for example, with one’s being left- or right-handed, with deep-seated personality traits such as extroversion or introversion, with one’s sexual orientation, and, with one’s desires, or felt needs, in general. Again, (theoretical) knowledge about influences of this kind can play a significant role in certain ethical discourses.

The above analysis makes it clear that ethical discourses cannot be divorced from theoretical and other non-ethical forms of discourse, including purposive ones, concerning means-ends relationships.

3.7.1.2 Purposive and Ethical Discourses

Let us now look at the structure of discourses concerned with finding and justifying the choice of means to given ends. Through an example, it will be shown that the selection of the “best” or “most efficient” means to a certain end is inseparable

from ethical choices over and above those involved in selecting the end in question.

A certain City Council has taken a policy decision to establish a new link across the river that flows in the outskirts of the city. The new link would be within a given stretch of the river and would be constructed with the aim of providing to a specified number of daily commuters easier access than now available to the city centre. The Council – having determined this goal – asks the relevant branch of the city administration to develop and assess alternative technical solutions to the problem, that is, alternative sets of means to the given goal, and select among them the one that would best/most efficiently realize the goal.

The following six types of solution seemed initially technically feasible: (a) a railway bridge, (b) a road bridge, (c) a rail and road bridge, (d) a railway tunnel, (e) a road tunnel and (f) a rail and road tunnel. In order to devise and comparatively assess specific solutions of each type, questions such as the following need to be asked and answered: (a) What are the advantages and disadvantages of building a tunnel rather than a bridge? E.g, which type is likely to be more costly? Which is likely to be safer? How would each type of solution affect the landscape and cityscape aesthetically? Which is likely to be more environmentally friendly, in terms of materials, energy, pollution? How would a bridge affect the economically important boat-traffic along the river? (b) What are the advantages and disadvantages of a road solution (bridge or tunnel) compared to a rail one, again in terms of criteria such as cost, safety, aesthetics, environmental impact, economic activity, etc.? How would the city decision-makers compare evaluatively the effect of the greater influx of cars into the city centre in the case of a bridge solution, to the inconvenience and cost to be suffered by commuters in case of a rail solution, according to which commuters would have to leave their motor vehicles in car

parks outside the city and make their way to the city centre by train? (c) Within each type of solution, how would different particular solutions, considered in all their specificity down to the smallest detail, compare in terms of the same, and perhaps more, evaluative criteria? For example, how would a bicycle lane on a road bridge be evaluated? Or, would a more costly but impressive suspension bridge – displaying architectural ingenuity, and adding prestige, and perhaps tourist income, to the city – be preferable to a more humble and functional, but less costly construction? (d) Would there be any differences between the various options with regard to jobs created (for realizing the option), thus with regard to easing the city's unemployment problem? (From a morally and legally permissive point of view – non-permissible within the communicative perspective – even more questions might arise; for example, which private companies are likely to benefit from each option? How close are such companies to the various political parties represented in the City Council? What – illegal – contributions have they made, or are likely to make to them?)

It is obvious that questions such as the above, concerning the development and comparative evaluative assessment of the various possible means to the end under consideration (with a view to selecting the best one), involve the clarification and assessment of goals, interests, values, other than those implicated in setting the initial goal. Means are thus value-laden through and through. Each alternative means-option, by virtue of its various technical characteristics, constitutes a complex value-ensemble (with evaluatively positive/desirable, but perhaps also negative/undesirable elements) which, when employed, includes – if it is an effective means at all – the initially given goal as *only one* of its elements.

The question “Which is the best means to the given end?” is thus not an entirely, even mainly, technical value-neutral question, reducible to one of theoretical

rationality. "Best means", in this context, means "the most (evaluatively) preferable among effective means", and not that which more closely approaches, in a factual/quantitative sense, a pre-given yardstick – the goal in question. Similarly, "most efficient", if it has a meaning other than "best", in the above general sense, means "best with regard to certain evaluative parameters", e.g. energy consumption, or (financial) cost, which are presupposed to evaluatively/ethically override other relevant parameters such as beauty, environmental impact, safety, etc.

What is evident in the above example applies to all issues concerning the choice of means to ends; from the broadest questions of social policy (e.g. designing and managing an "effective" national pension scheme), to the most trivial everyday instrumental choices (e.g. whether to take a lift to our office or walk up the stairs); it applies, not least, to all technical-economic decisions concerned with designing, mass-marketing and using technological objects such as computers, motor cars, the internet, genetically modified crops, etc. All means are thoroughly value-laden. All purposive, i.e. means-ends decisions are inseparable from ethical choices.

This inseparability of purposive issues from ethical ones obtains not only when there is more than one effective option – each of them being a value-ensemble, as seen above – in which case we need to choose between them on evaluative/ethical grounds. It happens also when only one option is offered as a technically effective means to a given end. In such a case it might seem, on the face of it, that the facts of the matter, value-neutrally "force" that solitary option on us. Yet, since that means-option, just like any other, is a value-ensemble including positive elements, such as the realization of the end in question, but possibly also negative ones, i.e. costs/losses, the question necessarily arises whether, as a whole, that option is worth adopting; whether the relevant end justifies that

particular – one and only – means-option; or, more generally, whether the gains offered by that option – including the initial end – outweigh the losses (to be) suffered because of it. And that, surely, is an evaluative/ethical question – one that can be answered either positively or negatively. Thus, even a single means-option to a given end is not value-neutrally binding on us. In fact, the evaluative/ethical decision to give up an end altogether due to the unacceptability of the means available for its realization, is open not only when there is only one possible means-option but also when there are several. For the cost of even the most preferable among them may be judged to outweigh the value of the end to be achieved.

In our discussion on the structure of ethical discourses, above, it has been suggested that these discourses, with a view to eliciting authentic ethical choices, might involve questions of biological and socialization influences on one's choices, of consistency between one's choices and one's overt behaviour, and of internal coherence of one's set of choices with regard, in particular, to such parameters as relations of generality and specificity, preference-ordering and purposive (means-ends) relations between the different items chosen. Now, the discussion of purposive discourses shows that the development and selection of means to an end, far from being a value-neutral process, necessarily involves ethical discourses, possibly of the high complexity indicated above.

The internal structure of a subject's conception of the good – in particular the means-ends relations between elements of this conception – in conjunction with the necessary involvement of ethical discourses in purposive ones, has an additional implication for efforts to develop and select means to ends – an implication especially important, as we shall see, for negotiation processes. That is, the purposive search for a means need not terminate the moment it is decided

that no effective or acceptable means exists or can be devised to the given end. The search may continue in terms of questions such as the following: What higher goals or values was the initial end intended to fulfill? Can alternative ends be set which could satisfy those same higher goals or values at least to the same degree, and for which acceptable means can be found? And if yes, which of those alternative ends-means combinations is the most preferable? The City Council in our earlier example, for instance, might reject as unacceptable all the river crossing solutions proposed by its technical services; but then conclude – on the basis of further technical and political consideration of the problem – that the wider interests which were explicitly or implicitly intended to be served by the initial goal of constructing a river crossing could, instead, and even better, be served by moving a number of industries and offices from the central areas of the city to its outskirts, on the other side of the river; and that for this alternative goal acceptable means were possible.

The analysis so far has demonstrated mainly an one-way dependence of the (technical) development of appropriate means upon the choice of ends as well as on other evaluative decisions. Such a dependence exists within specifically purposive contexts – where the development of means is guided at every step and in every aspect by evaluative choices – as well as within specifically ethical contexts where elements not valued in themselves are determined according to their role as means to elements which are valued in themselves. Elements valued in themselves tend to be abstract values (health, beauty, security, convenience, economy, etc.) which can be realized as properties of concrete things (states of affairs, objects, actions), and, by extension, through purposive relations of the latter with other concrete things. Thus, the dependence of the purposive on the ethical is, by and large, also a dependence of concrete on abstract evaluative judgements.

However, further elucidation of the relations between the development of appropriate means on the one hand, and the adoption of goals and values, on the other, reveals also a reverse dependence. Every means to an end, every technical solution to a problem (e.g. a suspension bridge of a specific design and method of construction, in our example), embodies – as suggested above – a certain value-ensemble: the realization of a unique combination of specific degrees of different (positive and negative) values, including those corresponding to the end in question (e.g. the construction of a river crossing). The specific suspension bridge under consideration would thus realize, for example, X degree of convenience to the commuters, Y degree of aesthetic value, Z degree of environmental burden, etc. – a value-ensemble unique to that specific bridge; a different technical option (a tunnel, a different kind of bridge, or even the same bridge constructed at a different point along the river) would realize its own (different and also unique) value-ensemble.

Now, the evaluative assessment of a value-ensemble is dependent on the technically developed concrete means embodying it. The value ensemble embodied by the above mentioned suspension bridge would be cognitively unimaginable, and at any rate intuitively impossible to evaluate – either in absolute terms or in comparison to other value ensembles – in abstraction from and thus prior to technically developing the specific bridge-option itself. In fact, according to the explicit or implicit logic of purposive rationality processes, the value-ensemble embodied in a means-option is constructed and evaluatively assessed step by step – beginning with the relevant goal – in reciprocal relation to the process of designing, or, if ready-made, of technically assessing, the means-option in question; each stage in the process of designing or technically assessing a means – a process initiated by a given goal – throws up evaluative questions answering which adds new elements to the relevant value-ensemble and at the same time

guides the design, or assessment process to its next stage, which, in turn, raises new evaluative questions, and so on until the design or technical assessment of the means, on the one hand and the construction and evaluation of the value-ensemble embodied in it on the other, are completed simultaneously.

Thus, the clarification and adoption of values, interests, preferences, etc. – in short, of elements of a subject’s conception of the good – depends on what the subject perceives as possible means to their realization as much as the latter depends on the former; ethical discourses feed on purposive discourses as much as purposive discourses feed on ethical ones. This rational, reciprocal relationship of mutual dependence between what we value and what means we have and develop is, perhaps, the core of what the early Habermas – in important essays advocating the mediation between technical progress and the social lifeworld, and the cooperative interaction between science and politics, technical experts and interested parties – refers to as the “dialectic of potential and will”, and, even more aptly, the “dialectic of enlightened will and self-conscious potential” (Habermas, 1971: 61, 73).⁴

The above considerations direct our attention back to a – still important – major early Habermasian theme, referred to in Part One: the theme of technocracy, technocratic consciousness and technocratic ideology. Technocratic consciousness is blind to the evaluative/ethical aspects of purposive problem-solving, reducing the latter to a value-free technical process best left to those who have the know-how. Such an illusory, ideological consciousness tends to depoliticize citizens and undermine the role of politicians, leaving effective social power in the hands of technical experts – “technocrats” – whose “technical” deliberations and decisions conceal, usually also from the technocrats themselves, the evaluative choices necessarily implicit in them.

To be sure, decisions made by technical experts (engineers, doctors, accountants, etc.) embody values/evaluative choices that can legitimately be considered internal to the relevant technical disciplines and professions (e.g. precision; economy of energy, materials, time; minimum safety standards; professional ethics). Some are adopted directly from within the technical disciplines and professions, others are internalised by them in the form of legally binding norms originating from outside. And, they are legitimately considered internal insofar as there is social awareness and acceptance of their presence; in other words, insofar as they are, in a sense, “delegated” to the technical disciplines and professions by society. To that extent, society’s self-consciously standing back and allowing experts to make decisions involving such (“internal”) values autonomously does not betray technocratic consciousness. However, not all the values embodied in purposive decisions, concerning the development of means, are, or – as argued above – can be internal in this sense, that is, adopted in advance. Some evaluative issues are, and cannot but be, raised and settled in the course, and up to the very end, of purposive deliberations. Thus, the idea that all means-ends discourses are, at least from the moment the relevant end is determined, value-free, or value-neutral processes – exclusively the province of technical expertise – does constitute a technocratic illusion.

Technocratic consciousness, in the above sense, is by no means exclusive to advocates of technocratic society. It often has a strong hold on its opponents too. For example, it is perhaps exhibited nowhere as clearly as in Jacques Ellul’s, undoubtedly critical, *The Technological Society* (1964). In that work, a most thoroughgoing technological determinist interpretation of modern society is essentially founded on the argument that the “best” or “most efficient” means – the pursuit of which is taken to be the hallmark of modernity – are determined in an “automatic” way that leaves no room for human evaluations (*ibid.*: 79–80).

Technocratic thinking can also be said to underlie Max Weber's view that purposive rationality, though hard to do without due to its effectiveness, eventually leads to such evils of modern rationalization as "mechanized petrification", "dehumanization" and the entrapment of humanity in an "iron cage". As shown in Appendix III, entitled "Max Weber and the Discontents of Modern Rationalization", Weber is led to his pessimistic verdict on rationalization by conceiving purposive rationality not as a process defined by its pursuit – through a thoroughgoing dialectic of technical and evaluative considerations – of the maximum technically feasible realization of what human beings consider ultimately valuable, but as a formalistic procedure whose engagement with the substantively evaluative is severely truncated by being limited to taking for granted arbitrarily posited ends.

3.7.1.3 Ethical, Purposive and *prima facie* Normative Discourses

Norms are rules which regulate actions, i.e. specify – for certain sets of people and within certain contexts – which actions are obligatory, prohibited or permissible. It can be argued that communicative purposive discourses (i.e. communicative negotiations) concerned with determining norms, are reciprocally dependent on ethical and purposive discourses.

As explained in preceding sections of this Part, the communicative determination of *moral* norms is understood in Habermasian theory as taking place through applications of the Universalization Principle, processes which we have interpreted as universal communicative negotiations. Thus, moral discourses – those of justification but also the dependent discourses of application – are normative only *prima facie*. According to the Universalization Principle, a rule of action, proposed as universally applicable, is judged to be a valid moral norm if – in a (potential) universal human discourse where everyone's interests are equally taken into consideration – all can accept (and prefer to alternatives) the consequences

(including the side effects) which its universal observance can be anticipated to have on their interests. The choice of moral norms, therefore, depends on the one hand on the interests of all those involved – i.e. the interests of all human beings – and on the other hand on how possible applications of the norm being considered will relate, as means to ends, to those interests. Thus, discursive considerations about the determination of moral norms depend both on ethical considerations, concerning interests, and on purposive ones, concerning means-ends relations. Conversely, once a rule is recognised as a valid moral norm, it, in its turn, sets limits both on the ends/interests/ethical values that is morally permissible to pursue and on the (necessarily value-laden) means that it is morally permissible to employ. Thus, ethical and purposive discourses are, in turn, dependent on moral ones.

This reciprocal dependence between moral discourses on the one hand and ethical and purposive ones on the other exists, *mutatis mutandis*, also with regard to non-moral normative discourses, legal, organizational or other.

3.7.2 Communicative Negotiation

3.7.2.1 Introductory ideas

Communicative Negotiation, as defined earlier in this Part, is coordination of action between two or more agents which is at once communicative (i.e. oriented to reaching understanding) and purposive, or success-oriented (i.e. oriented to finding means to ends believed by the agents concerned to be egocentric, that is not genuinely regulated by shared norms or values). When discursively rational, communicative negotiation is communicatively fully rational, while it is also purposively rational to the maximum degree attainable by communicative means,

though this may often fall short of full purposive rationality, i.e. of the maximum purposive rationality attainable by any means, communicative or strategic.

The concept of communicative negotiation, which has been developed in order to fill an important gap in Habermas's conceptual framework and grounded in terms of a revised Universal Pragmatics, can be applied, in both a prescriptive and a descriptive sense, to a wide range of social interaction situations: from simple everyday cases of issuing and responding to (egocentric) imperatives, including requests, to economic exchanges, administrative interactions as well as complex processes of collective policy and law making.

As a field of academic study, negotiation analysis belongs, together with the study of individual decision making and Game Theory, to the broader discipline of Decision Theory. Scientific work in all the fields of Decision Theory is carried out within the paradigm of Rational Choice, that is from the perspective of (fully) purposive action and rationality. Game Theory and Negotiation Analysis – both areas of study of group decision making – are normally distinguished on the grounds that the former deals with separate though interacting decisions, whereas the latter is largely concerned with jointly taken decisions. However, the two have a great deal in common. Many decisions taken within a negotiation context are lone decisions, taken without knowing with any certainty what the other(s) will do. Conversely, lone game decisions are taken on the basis of reciprocal assumptions about each other's intentions, and lead to jointly produced results. For this reason, work done in Game Theory is often relevant to the study of negotiations, research on the Prisoner's Dilemma being a prominent example.

In a negotiation two or more parties, through communication (among themselves – and possibly their advisors – alone, or also with the participation of neutral, third parties such as facilitators, mediators, arbitrators, etc) try to reach an agreement

out of which each party would gain more than (or at the very least as much as) it would have gained in some alternative way. Such alternative ways might be (i) terminating the negotiation and leaving things as they were before, (ii) successfully negotiating with another party, or (iii) making some other (unilateral) efforts of its own. Thus, negotiating parties, aim to use each other – more precisely, each other's actions, services, consent – as means to the realization of some of their ends. To the extent that this can be achieved, the parties can be said to be interdependent with regard to realizing their relevant ends.

If any of the parties, at any point during a negotiation, comes to believe that its relevant ends cannot be satisfactorily realized in the context of that negotiation – i.e. that they cannot be realized to a greater, or at least equal, degree as they would in the most beneficial of the said party's alternatives – that party can, and rationally should, withdraw from the negotiation and pursue its best alternative; whereas if any party believed beforehand that an alternative approach better than the prospective negotiation was available to it, it could, and rationally it should, refuse to enter the negotiation in the first place. In some of the literature on negotiation, a party's best alternative is technically referred to by the term Best Alternative To a Negotiated Agreement, for short BATNA (see Raiffa, 2002: 110; Fisher et al, 1991: 101ff).

The above applies to both Rational Choice negotiations and communicative ones. The two differ, though, in that communicative negotiations, unlike Rational Choice ones, do not allow themselves strategic, i.e. causal-manipulative, behaviour such as deception or coercion, either in the actual negotiation itself or in the (hypothetical) negotiations or other actions considered as possible alternatives to it, especially in what they take to be their party's BATNA. Despite this evidently important difference, however, communicative negotiation can learn a great deal

from Rational Choice negotiation analysis. And this is so not only due to the considerable common ground between the two, but also because Rational Choice theorists have come to recognize that very often in negotiations it makes good Rational Choice sense to adopt approaches and methods akin to those we would characterize as communicative.

Below, we shall look – in a highly condensed fashion – into some elements of Rational Choice negotiation analysis which can be used profitably to clarify and elaborate our concept of communicative negotiation. These elements are of particular importance when they link up with the revised model of practical rationality explained in earlier parts of this study, especially with the concepts of ethical and purposive discourses and their interdependencies.

3.7.2.2 Elements of Rational Choice Negotiation Analysis

In the relevant scholarly literature, two aspects of negotiations are normally distinguished: the *integrative* and the *distributive* (Raiffa et al, 2002: 97, 191). The first aspect is concerned with developing, through the integrated efforts of the negotiating parties, potential agreement options which would leave the parties better off than they would be without agreement. The second aspect is about how the parties will share – distribute among themselves – the total benefit, or value, thus produced. These two aspects are often referred to by the terms “creating value” and “claiming value”, respectively (see *ibid.*, and Lax and Sebenius, 1986: 129ff).

In some cases, the total value to be created in the negotiation is considered by the negotiating parties to be fixed, or constant, regardless of how it is eventually shared among them. In these cases the parties concentrate their efforts on claiming and securing as big a share or the “fixed pie” – another metaphor used –

as they can. They are then said to be involved in a “zero-sum game”, (Raiffa et al, 2002: 97, 279) in the sense that the difference in the total value that results from different outcomes of the claiming and sharing game is zero; or, put differently, they are involved in a “constant-sum game”, in the sense that the total value secured (the sum of the values that each secures) in the game is constant regardless of its distribution among the parties: “what you gain the other side loses and vice versa” (*ibid.*). Often, however, the total value that can be created through negotiation is variable; more precisely, it is higher than what is initially visible to the parties. The “pie” is not fixed; it can be enlarged. The challenge for the negotiating parties in these cases is to envisage this potential and try to actualise it, by creating through a “variable-sum game” as much total value as possible before sharing it among themselves.

Let us consider two examples (both borrowed, with modifications and enrichments, from Fisher et al, 1991). The first is a well known simple story – though highly instructive in its simplicity – of two young sisters who were claiming the same orange (*ibid.*: 59). Each of the sisters, perceiving the problem as a fixed-pie, or zero-sum one, first insisted on getting the whole orange, and then claimed a disproportionately large portion of it for herself, using alternately, threats, promises, emotional appeals and even arguments such as “I was the first to ask for the orange” or “I have priority because I am older”, evidently based on principles presumed to be valid. But finally – neither of the two being able to overpower, outwit or rationally convince the other – they settled reluctantly on a fifty-fifty compromise solution, taking half of the orange each. On this understanding of the problem, no matter how the two sisters agreed to divide the orange between them – including the extreme option of one sister getting the whole orange and the other getting nothing – the total value created and distributed would remain constant: one whole orange.

But when the orange was cut into two and each girl got her own half, to their mutual astonishment, this happened: The first girl, as might be expected, peeled her half of the orange, ate the flesh and threw the peel away; the second girl, however, peeled her own half, threw the flesh away and used the peel as an ingredient of the cake she was baking. They then of course realized, belatedly, that both would have been better off by far had the first got all the flesh of the orange and the second all the peel, neither of them throwing anything away. In such a solution, the total value created in the negotiation and shared by the two sisters would have been twice as much as the one actually created: Each would have got twice as much of what she wanted and nothing of what she did not want. To achieve such a feat, the two young negotiators would have to approach their problem as a potentially variable-pie/variable-sum game. To do that – and this is of crucial importance for our argument – they would have to engage in two simultaneous and interrelated processes: on the one hand, a process of asking and telling each other the reasons (the underlying concerns, needs, desires – in general, the interests) because of which each so insistently claimed the whole orange for herself; and on the other hand, the process of trying to devise means by which those underlying *interests* of both (rather than their initial *positions*, i.e. their claim to the whole orange) could be satisfied as much as possible.

Of course, had the two sisters both wanted the flesh of the whole orange, or if they both wanted its peel, rather than different parts of the orange, the total value available for distribution would have been fixed, and no negotiation technique could yield extra value. On the other hand, in the event that the two sisters, in addition to wanting the flesh and the peel of the orange, respectively, also shared an interest – one inspired, for instance, by their botany teacher – in planting seeds and nurturing the sprouting plants, yet more value could be creatively conjured up, for both, this time out of the orange pips.

The second example is analogous to the first, but this time it is taken from the real world of international politics and diplomacy. It concerns the 1978 negotiations between Egypt and Israel on the status of the Sinai Peninsula which led to the peace treaty agreed at Camp David. A lengthy quote is here appropriate:

When Egypt and Israel sat down together in 1978 to negotiate a peace, their positions were incompatible. Israel insisted on keeping some of the Sinai. Egypt, on the other hand, insisted that every inch of the Sinai be returned to Egyptian sovereignty. Time and again, people drew maps showing possible boundary lines that would divide the Sinai between Egypt and Israel. Compromising in this way was wholly unacceptable to Egypt. To go back to the situation as it was in 1967 was equally unacceptable to Israel.

Looking to their interests instead of their positions made it possible to develop a solution. Israel's interest lay in security; they did not want Egyptian tanks poised on their border ready to roll across at any time. Egypt's interest lay in sovereignty; the Sinai had been part of Egypt since the time of the Pharaohs...

At Camp David, President Sadat of Egypt and Prime Minister Begin of Israel agreed to a plan that would return the Sinai to complete Egyptian sovereignty and, by demilitarizing large areas, would still assure Israel's security. The Egyptian flag would fly everywhere, but Egyptian tanks would be nowhere near Israel (Fisher et al: 42-3).

Here, again, as in the orange example, the negotiation began as a constant-sum game, on the assumption of a fixed total value. The perceived fixed total value – possession of the Sinai – was from the beginning claimed in its entirety by both parties. No compromise sharing of that total value between the two was acceptable to either. However, the total value was considerably enlarged when the parties moved to a variable-sum form of negotiation. On that approach, they engaged in the twin process of (a) exploring with each other the deeper interests (needs, etc.) behind their positions, and (b) creatively inventing solutions that would satisfy those deeper interests sufficiently. Through this process – in which, it must be said, a very significant role was played by skillful mediators – the parties were able to create sufficient additional value.

To be sure, in this case, unlike the earlier example, it can be justifiably argued that the eventual total value did not satisfy the relevant needs fully. Each of the two sisters got one hundred per cent of what she wanted – clearly, each was totally disinterested in the part of the orange that went to the other sister. The same, however, cannot safely be said about the Sinai case. Egypt would most likely prefer not to have any restrictions on the extent of militarization of – thus on its sovereignty over – the peninsula; while Israel would most probably feel more secure if the Sinai was under its own control, or at least if the whole of it was demilitarized, even if under Egyptian sovereignty. Thus, though what each side eventually got was far more than fifty per cent of the value initially claimed, (which would be what they could expect in an equitable compromise in the context of a constant-sum negotiation) it was still below one hundred per cent. Therefore, in addition to creating extra value, the two sides, unlike the two sisters in the orange example, also had to play a difficult claiming game over a part of the eventual total value. One might venture to say that one reason why they concluded that game successfully was that agreement between them meant yet more value, for both,

over and above the value accruing from the (partial) satisfaction of their interest in sovereignty and security, respectively: namely, the value resulting from the fulfillment of their *common* desire for peace between them; a factor not explicitly thematized in the streamlined description quoted above.

One might visualize, furthermore, (and here we are admittedly taking a flight from the real negotiation into the realm of the purely hypothetical) the creation of yet more (common) value, had those involved been even more creative, inventing ways of satisfying also other interests of the two countries. Such ways might be, pooling their (presumably, separately insufficient) resources, to explore and commonly exploit possible oil reserves in the Sinai, or to irrigate and make fertile certain parts of the peninsula.

But although it is the aim of all negotiations to create value, whether fixed or variable, not every negotiation does so. It is perfectly possible, and in practice often the case, that a negotiation does not produce any value at all. Such an outcome occurs when the negotiation does not lead to an agreement, leaving things as they were before. In some such cases this outcome is deliberately chosen by one or more of the parties because they have, or believe they can achieve, a better alternative than any feasible agreement achieved in the present negotiation would secure. In other such cases, however, the lack of agreement is not chosen, but results from incompetent negotiating by one or more of the parties, and deprives the parties of the opportunity to create more value for themselves than any of the alternatives available to them. Failure of the two sisters to agree on what to do with the orange, for example, would leave the needs of both totally unsatisfied in the absence of an alternative, whereas even a straightforward, equitable zero-sum compromise could have left both of them satisfied. And, if the “negotiation” led to a quarrel, the orange might even end up squashed on the floor. Similarly the (hypothetical) inability of Israel and Egypt to reach a negotiated

agreement on Sinai would have perpetuated the status quo and left unfulfilled the interests (in security and sovereignty, respectively) for the sake of which they had entered the negotiations in the first place.

But worse still, a badly handled negotiation may not only fail to produce any new value, it may even result in the destruction of existing value. A rowdy dispute over the orange, might not only deprive the two sisters of any part of the orange, it might, in addition, provoke their mother's punishment and/or cause lasting resentment between them. Similarly, a failure of the Egypt-Israel negotiations, accompanied by reciprocal public recriminations, could lead the two countries into another, mutually destructive war. Thus, in all negotiations value needs to be produced, or at least preserved, as well as claimed. And the extent to which in a negotiation the combination of value creation and value-claiming succeeds in satisfying the negotiating parties, either through an agreement, or through a deliberate non-agreement in view of a better alternative, depends on two factors: Firstly, it depends on the potential offered by the particular situation. The most important factor, in this respect, is whether we have a zero-sum situation or a potentially positive-sum one. Secondly, it depends on how skillfully the negotiation is carried out; in particular, on the ability of the negotiators to actualize the value-creating potential inherent in the situation.

With regard to the latter point – concerning effective negotiation approaches and methods – Rational Choice negotiation theorists identify a tension between creating value and claiming value, the techniques most suitable for successful value claiming being inappropriate for – indeed working against – successful value creation, and vice versa. So much so that certain theorists consider the effective handling of this tension, or “negotiator's dilemma”, as the heart of the science and art of negotiation (see Lax and Sebenius, 1986: 38ff; Raiffa et al, 2002: 95).

Value creation (in a broad sense, including preserving existing value, securing value visible and claimable ahead of the negotiation and generating extra value in the course of the negotiation) is best facilitated by a cooperative, mutually respectful and open approach, one of collaborative common-problem solving. In particular, the creation of extra value – based on the above mentioned twin processes of (a) exploring possible, situationally relevant, deeper interests underlying the parties' initial positions/claims and (b) inventing ways of satisfying the parties' deeper interests to an optimal degree – requires firstly that the parties, through mutual questioning and prompting, delve deep into themselves and honestly share information about their authentic subjective states and secondly that they pool their creative energies together to devise imaginative, situationally feasible solutions. Raiffa et al speak, in this respect, of a “Full, Open, Truthful Exchange” (FOTE) (2002: 83, 86–7). It is obvious that these requirements for value creation are very close to our own Habermasian communicative discourse.

The circumstances of value-claiming, however, call for a more differentiated approach. For there is a whole set of possible situations that differ with regard to the ways in which the respective total value available for claiming satisfies the requirements of the parties. Below, I shall provide a schematic presentation of this spectrum, taking, for the sake of clarity, the simplest case of a two-party negotiation between A and B, the argument being applicable, with suitable adjustments, to more complex negotiations.

In all the diagrams in Figures 7 to 9, the areas enclosed within the outer line represents the total value, V_T , which at a given phase in the negotiation is believed by agents A and B to be of interest to them and available for sharing between them. V_T is at the same time considered to be permissible according to norms shared by A and B in the context of the given situation. The area outside the outer

line, on the other hand, represents non-required, non-available, or non-normatively permissible value.

In these diagrams, areas designated V_A represent value which is required by agent A alone and areas V_B represent value required by B alone. Areas $V_{A \text{ and } B}$ represent value required by both A and B and capable of satisfying both requirements, whereas areas $V_{A \text{ or } B}$ represent value required by both agents but capable of satisfying only one or the other fully, or both of them only in complementary parts. Areas $V_{A'}$ and $V_{B'}$ represent values required by both agents but recognized, according to norms shared by them in the given situation, as rightfully belonging to (i.e. as normatively sanctioned entitlements of) A or B, respectively.

The diagram in Figure 7 represents a simple and rather extreme member of the set of possible value-claiming situations: the purely antagonistic, zero-sum situation, where all the value available, V_T , is of the type $V_{A \text{ or } B}$. This is the case of the two sisters claiming the whole orange, each for herself; or the case of Egypt and Israel each claiming, as they did initially, unrestricted sovereignty with full militarization rights over the Sinai peninsula.

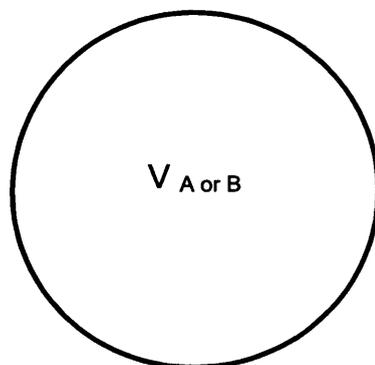
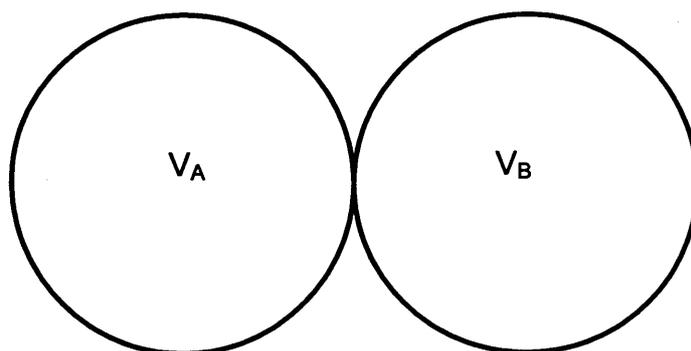
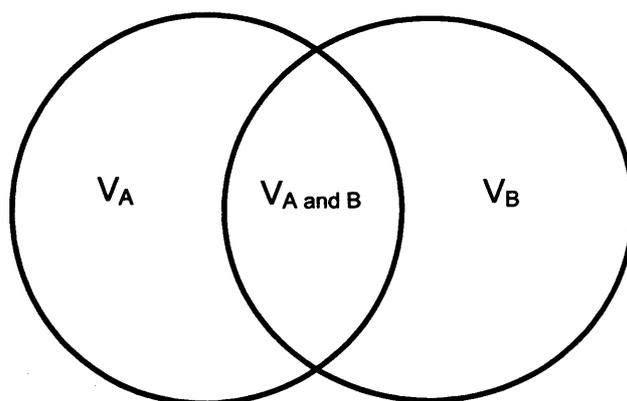


Figure 7

Another paradigmatic member of the set of possible value-claiming situations, and also extreme, though in the opposite way to the previous one, is the situation where the value available satisfies fully all the situationally relevant requirements of both parties. Here, two different versions can be discerned. In the first version (Figure 8a) there are no conflicting requirements of the parties, no zero-sum aspects to the negotiation. The requirements are simply different. Here V_T is equal to the sum of V_A and V_B . An example would be that of one of the two sisters claiming only the orange flesh and the other claiming only its peel. The second version (Figure 8b) combines requirements that are not conflicting but are different, in the above sense, with identical requirements all of which can be satisfied. Here the total available value V_T is the sum of V_A , V_B and $V_{A \text{ and } B}$.



(a)



(b)

Figure 8

A case in point would be the two sisters claiming, each for herself, the orange flesh and the peel, respectively, and at the same time both of them (together) wanting the orange seeds for their joint botany experiments.

The diagram in Figure 9 is a generic representation, where all the possible value-claiming situations, including the two described above, can be obtained by varying the magnitude of the different types of value involved, down to nil and up to the total available value V_T . In this generic diagram, where all the different types of value are present, the total value is the sum of V_A , V_B , $V_{A \text{ and } B}$, $V_{A \text{ or } B}$, $V_{A'}$ and $V_{B'}$.

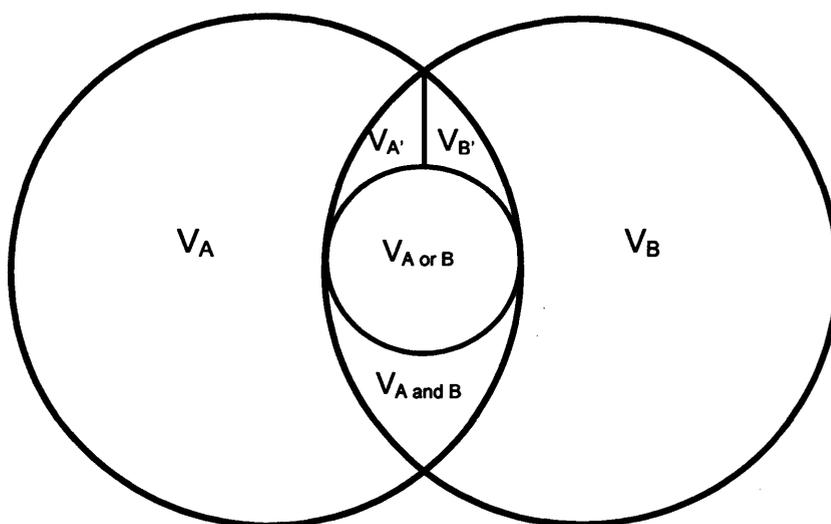


Figure 9

Such a maximally composite value-claiming situation, in other words, combines (i) conflicting requirements, (ii) requirements that are different but not conflicting, (iii) coinciding and jointly satisfiable requirements, as well as (iv) a conflict of requirements which is normatively resolved. Examples of $V_{A'}$ and $V_{B'}$ would be, (a) the whole orange, had the two sisters shared a norm such as "Priority goes to the one who asks first", and (b) the status of the whole or parts of the Sinai, had there been binding law for settling international disputes such as the one in question.

Evidently, in these situations, values V_A , V_B , $V_{A'}$ and $V_{B'}$ (Figures 8 and 9) are not contentious, and their claiming can, in principle, be pursued in a cooperative/communicative manner. Values $V_{A \text{ or } B}$ (Figures 7 and 9), on the other hand, are contentious, creating situations of zero-sum claiming: the more one party gets, the less will remain for the other. Here, the (contentious, zero-sum) claiming may have to be competitive/non-communicative, though not necessarily so, as we shall see.

Situations of genuinely conflicting requirements, thus the need for *some* zero-sum value claiming, can be expected to arise in the great majority of negotiations; for cases such as those represented in Figure 8 (a and b), where the situation is *entirely* non-contentious, are rare. And even those rare cases cannot be known until the final stages of a negotiation are reached. For these reasons, all negotiations need to be conducted on the assumption that in them some degree of contentious value claiming is likely to be called for.

Nonetheless, even contentious value-claiming need not require, from a purposively-rational point of view, competitive claiming methods. To begin with, even in cases where competitive claiming is not expected to impair value creating ability, it is often prudent to seek a cooperatively reached fair compromise rather than risk getting a smaller share as a result of a competitive process. This is true for negotiating parties that believe or suspect that they are not strong enough (i.e. that they do not possess sufficient capacity for deception and/or coercion to be able to secure for themselves a better deal than they would achieve by cooperative means).

But in many cases, the reason why it is purposive-rationally preferable, even for the party that feels stronger, to pursue contentious claiming cooperatively, rather than competitively, is the danger that competitive value claiming will have a

seriously adverse effect on value creation – that is, keep the total value eventually available for claiming and sharing to unacceptably low levels. If it were possible in a negotiation, first, through cooperative methods, to create maximum value and do all the non-contentious claiming, and only then use competitive methods to do the unavoidable contentious claiming, this danger might be considerably reduced; there might be much less tension between cooperative and competitive approaches. However, this is not possible; for the processes of value creation and value claiming (contentious and non-contentious) are necessarily intertwined. From the very beginning, while having to cultivate mutual trust and engage in an open, honest dialogue in order to create more value, the parties also need to be in some respects deceitful and even coercive in an effort to secure a bigger share of the value created. In this connection, Raiffa et al refer to “bluffing, threats, trickery, exaggeration, concealment, half-truths and outright lies” (2002: 83). Such means, however, can seriously damage trust and cooperation, and thus reduce the negotiators’ ability to create value, and even to successfully claim value that is of interest only to themselves, or value that could be enjoyed by the two parties jointly, or indeed value which could be shown to be rightfully theirs. Heavy handedness for the sake of better share of the eventual total value, may not only obstruct the creation of new value; it may also result in destroying existing value too, as pointed out in our preceding discussion of the two examples.

It is also noteworthy that excessively competitive methods, even when beneficial in the short run to those who employ them, (particularly in securing a better deal in the negotiation underway) can seriously harm their interests in the long run: they may, for example, impair their ability to engage in successful negotiations (with respect to both value-creating and value-claiming) or other exchanges with the same parties, not excluding those exchanges necessary for the implementation of the (beneficial) agreement reached in the negotiation in question. Furthermore,

excessively competitive methods may cause, for those who use them, loss of reputation and credibility among third parties (colleagues, potential customers or business associates, significant other countries, etc); which will damage future negotiations, or other relations, with them.

To be sure, the tension exists also in the opposite direction: Too much openness and honesty may in some cases diminish one's value-claiming ability. However, the value-creating potential of cooperation and the value-undermining potential of competitiveness are such that, more often than not, one can benefit more from getting a (cooperatively agreed) fair share of a larger (cooperatively enhanced) pie, than a disproportionately big (competitively achieved) share of a pie that has remained smaller due to interference from competitive claiming methods. This is particularly the case in situations where the balance of power – even though possibly in favour of one party, thus tempting that party to employ competitive value claiming methods – is not so one-sided that would enable the stronger party to secure a sufficiently disproportionate share of the eventual value to make up for the smaller size of the total value that would result from the interference of the competitive methods with value creation. It is also true in situations where much of the value at stake for a (currently) strong negotiating party lies in that party's future relations with its current negotiating counterparts and/or third onlookers, while its present comparative strength (possibly justifying a competitive approach now) cannot be assumed to last. Such circumstances are by no means rare in real social life. On the contrary, it can be said that they are part and parcel of much of it.

For these reasons it is very often preferable, from a purely purposively rational point of view, to resolve the above mentioned tension between cooperative value creation and competitive value claiming not through a delicate balancing act

between the two but by consistently employing throughout a negotiation – both for value creation and for value claiming – an exclusively cooperative approach such as Raiffa's "Full, Open, Truthful Exchange".

3.7.2.3 From Rational Choice Negotiation Analysis to Communicative

The similarity between what, within the Rational Choice paradigm, is called cooperative, or, collaborative, negotiation and what we have termed communicative negotiation is obvious. Thus, whatever support Rational Choice Theory provides for cooperative negotiation (which, of course, though considerable, does not amount to a justification for its universal use, as we have seen) can be said to apply also to communicative negotiation; which, in addition, is rationally fully grounded in terms of the communicative paradigm, as argued earlier in this Part. The support coming from the perspective of purely purposive rationality is important for communicative negotiation, as it can have a powerful influence on the will to employ in practice the communicative mode of negotiating. For this support – though admittedly partial – is accompanied by the strong motivational force of interests, which greatly enhances what Habermas calls "the weak motivational force of good reasons" (1996: 164), which comes with the (full) rational support provided to communicative negotiation from within the communicative paradigm. This strong motivational force of interests, in the context of a communicative negotiation is directly proportional to the value expected to be created (and equitably shared) in that context.

Below, we shall try to develop the concept of communicative negotiation a little further, through comparing and contrasting it with the Rational Choice conception of negotiation outlined above, and linking it with aspects of communicative practical rationality, discussed in the previous section. A major concern will be to

enhance as far as possible the value creating capacity – and thus the purposive rationality and the concomitant motivating power – of purely communicative negotiation.

According to the preceding analysis, negotiations are mutually purposive interactions into which two or more parties enter, seeking to use each other – strictly speaking, each other's actions or consent – as means to what at first sight seem to them to be egocentric goals of theirs, that is goals that are for them normatively or evaluatively permissible but not obligatory; goals that are, therefore, considered conflicting or at least at variance with each other. Negotiation situations typically arise – to remain with the simple, bilateral model – when one agent addresses to another what appears to at least one of them as an egocentric request (i.e. a request not regulated for them either normatively or evaluatively) which is normally accompanied, explicitly or implicitly with a reference to corresponding rewards or other addressee-relative reasons for fulfilling the request. A negotiation process gets properly underway when the addressee makes a counter-proposal which the former agent finds unsatisfactory but not a sufficient reason for not pursuing the particular interaction further.

Rational Choice negotiation, by definition, aims, overall, to be fully purposively rational, and is cooperative only insofar as this serves that aim. Communicative negotiation, by contrast, aims to be on the one hand, and primarily, fully communicatively rational, and on the other hand purposively rational to the maximum degree compatible with communicative rationality. This maximum possible degree of purposive rationality sometimes amounts to full purposive rationality, but sometimes falls short of it. Once engaged with each other in a negotiation over goals initially considered egocentric, in our sense of the term, communicative agents try to devise a solution – a negotiation “deal” – which they

can convince each other communicatively (i.e. by means of raising and redeeming validity claims) that it is for both, from a purposive-rational point of view, the best achievable by communicative means, either within the present negotiation or outside it, not excluding the possibility of a fully purposively rational outcome. Such a communicatively optimal solution, or deal, can be achieved according to our earlier analysis, by means of two analytically distinct processes. These are, in the terminology used earlier, (1) the communicative creation of maximum value to be available for sharing among the parties, and (2) the communicative claiming (and thus the equitable, or “fair” sharing) of that value. This section will conclude with some additional remarks on these two processes.

1) *The twin-process of communicative value creation.*

As we have seen, in the discussion of the *orange* and *Sinai* examples, value creation is achieved through a two-pronged communicative process: (a) the process of exploring (i) the interests (including – shared or divergent – values, principles, needs) underlying the parties’ initial positions/negotiation claims, (ii) other relevant interests of the parties (i.e. interests which, by being brought into the negotiation, could facilitate its successful resolution) and (iii) relevant shared norms; and (b) the process of inventing means (i.e. possible negotiation solutions or deals) by which to satisfy as much as possible the relevant underlying interests of the parties. The invention of such means involves the analysis of the negotiation “stake” (i.e. the contentious object which the negotiation is mainly about – the orange, the Sinai peninsula) into constituent elements which can satisfy the parties’ underlying interests either separately (e.g. the orange flesh, the orange peel) or together (e.g. the orange seeds). It also involves the mutual exploration and communicative influencing of those factual beliefs held by the parties which might have a role in the parties’ linking the means devised, and proposed, with

their interests (e.g. in linking the partial demilitarization of a Sinai that nevertheless is under Egyptian sovereignty, with Israel's fundamental interest in security). Such beliefs might include prejudices, false expectations, or fears, which can either be rectified or, if this is not possible, accepted and taken into account in the negotiation. Consider, for example, Egypt's and Israel's beliefs about each other and about the other's likely future attitudes and behaviour towards them.

The point of this twin-process is to translate the initial, apparently conflicting goals/positions/claims of the parties into requirements (that is, required elements of the invented means) that are as reconcilable as possible. In terms of Figures 7, 8 and 9, above, the aim of the twin value-creating process is to move from initial situations often approaching that in Figure 7 to situations as close as possible to those in Figures 8a and 8b, or, more generally to situations where (in terms of Figure 9) V_A , V_B , $V_{A \text{ and } B}$, $V_{A'}$ and $V_{B'}$ are maximized and $V_{A \text{ or } B}$ is kept to a minimum.

The appropriate *form* of this twin-process of value creation is described by Rational Choice negotiation analysts in terms such as Raiffa's "Full, Open, Truthful Exchange". This notion, however, can be greatly enriched, as well as gain in philosophical depth and justification, by drawing on the considerable body of Habermasian ideas on the conditions and characteristics of communicative discourse: ideas – summed up in such concepts as "ideal speech situation" and "unrestricted communication community" as well as in Habermas's theory of argumentation – which specify the meaning of a discourse in which the only influencing factor is the "force of the better argument".

With regard to the *substantive* aspect of the value creating process (that is, the mutual exploration and critical, including self-critical, examination of values, principles, interests, needs, desires, goals, preferences, attitudes, concerns,

beliefs, prejudices, fears, etc., on the one hand, and the creative invention of means that maximally satisfy the genuine requirements of the parties involved, on the other) we can, similarly, draw on the dimensions of practical rationality discussed earlier, in section 3.7.1; more precisely, on the ideas concerning ethical, purposive-rational and (*prima facie*) normative discourses and the necessary interdependence between them.

2) *The process of communicative value claiming.*

The various alternative solutions, or deals, proposed in a negotiation are likely to differ with regard to how much each satisfies the requirements of each party. In the best of cases the solutions invented manage to resolve all the normatively resolvable issues (concerning prohibitions, obligations and entitlements) and reconcile all the reconcilable remaining requirements. Such solutions, therefore, differ only in the extent to which they can satisfy each party with regard to normatively non-regulated and irreducibly non-reconcilable requirements.

The sharing among the parties of the total value available is effected through reaching agreement on one of the proposed alternative solutions, and thereby determining the degree of satisfaction of each party. In negotiations intended to be fully communicative, the tension identified within the Rational Choice paradigm between the twin-process of value creation (which needs to be cooperative/communicative), on the one hand, and the process of value claiming (which may need to be non-cooperative/non-communicatively), on the other, does not exist; for within the communicative paradigm, the option of using non-cooperative/non-communicative claiming methods for the sake of securing for one's own party, at the expense of the other(s), a better share of the value available, is not open.

Non-communicative methods – in our Habermasian terminology, “strategic” methods – of making others accept a less than fair deal, largely consist of forms of deception (including self-deception) and coercion. By contrast, communicative methods of reaching agreement on how to share the value available – including, of course, value of type $V_{A \text{ or } B}$, i.e. value with regard to which requirements are non-reconcilable, and where the main value-claiming problems therefore lie – are limited to methods based on raising and redeeming validity claims; and these exclude both deception and coercion. When Rational Choice negotiation analysts consider it more purposively rational to refrain from using non-cooperative/non-communicative methods of value claiming, they propose approaches such as that of “Full, Open, Truthful Exchange”, which, as we have seen, they propose also for value creation and which, as suggested above, can be greatly enhanced by being linked to the Habermasian communicative conceptual apparatus.

Concerning the exclusion of coercion from communicative value claiming, some further clarification is in order. In (non-communicative) negotiation – as opposed to open conflict – coercion is not exercised directly; it enters the process mainly through credible “threats” and “promises” of a certain kind. The use of threats and promises is indeed, for Habermas himself, the way in which power is exercised in negotiations, that is, how bargaining – even fair bargaining – is for him differentiated from communicative interaction (see section 3.3, above). Thus, one can make another party accept an otherwise unattractive deal by “promising” to offer something additional that would render the overall exchange more appealing; or by “threatening” to resort to one of three alternatives: (i) to withdraw from the negotiation and seek satisfaction of its relevant requirements independently of its present negotiating counterpart leaving the latter in an even more unattractive position, or (ii) to withdraw from the negotiation and in addition take away something the other party has, or (iii) to withdraw from the negotiation and use

methods other than negotiating to achieve with the (same) other party, an exchange which is even better for itself and worse for the other than the best negotiated deal under consideration.

However, the credible appeal to resources other than the negotiation at hand, or other than those expected to be used within the negotiation, in an effort to make the other party accept a deal less favourable to itself, need not be non-communicative, or strategic; i.e., it need not amount to coercion. Some forms of “threatening” – better called forms of warning, cautioning, notifying, informing – or promising, can be means to what can be called “communicative influencing”, rather than coercion or causal influencing more generally. To be sure, the (non-legitimate) promise made by A to B, to the effect that A will return to B – if only B agrees to a certain negotiated deal – an object which A had stolen from B, or taken from B by force or deception, clearly constitutes a non-communicative negotiation method; as is the offer of any bribes. Similarly, non-communicative are threats to use physical strength, underground connections, or military might to get what one wants and the other refuses to give within a negotiation context; and/or to punish the other for that refusal. But this is not so with (legitimate) promises to offer, in a negotiation context, what is rightfully ours (that is, what we have a right to dispose of at will); or with (legitimate) warnings –in case of non-acceptance of a certain proposal by the other – (i) that we shall break off the negotiation and conclude a (communicative) agreement with another negotiating partner who is prepared to accept the proposal in question, or (ii) that we would break off the negotiation and try to get what we believe we are entitled to by legal means (i.e. by appealing to a shared normative framework), or (iii) that we shall take away from the other party an established, though voluntarily granted favour. Such methods “of threatening” and promising, because they are legitimate, that is,

communicatively justifiable, are clearly admissible in a communicative value claiming process.

Thus, the notion of communicative negotiation – i.e. of (reciprocal) success oriented communicative action – and its rationality, can be conceptually clarified and elaborated, as well as methodologically enriched, to a considerable degree, by drawing on both Rational Choice negotiation analysis and Habermasian discourse theory, including such elements of the revised model of practical rationality as those concerning the interdependencies between ethical, purposive and normative discourses discussed in section 3.7.1, above. At the same time, however, the notions of “Full, Open, Truthful Exchange” and “cooperative”, or “collaborative” negotiation, which refer to an important, though not the only, approach to negotiation within the Rational Choice paradigm, could, I believe, significantly benefit from the essentially Habermasian, communicative concept of negotiation outlined in this study.⁵

3.8 TOWARDS A PRACTICABLE IDEA OF A COMMUNICATIVELY RATIONAL SOCIETY: PRELIMINARY CONSIDERATIONS ON THE SOCIAL APPLICATION OF THE REVISED CONCEPTION OF PRACTICAL RATIONALITY

3.8.1 The problem of the revised concept's practical application in modern society

As indicated in section 2.5, the later Habermas considers it unavoidable that, under conditions of modern complexity, the huge societal spheres of economic and state-administrative action will continue to be organized in the way he believes them to be organized at present, that is, as media-steered subsystems of strategic action. He is convinced that a modern society which is thoroughly integrated by means of action orientations – thus potentially integrated solely in terms of communicative, and communicatively rational, action – is not viable in practice.

It has been one of the main theses of this study that Habermas does not have adequate conceptual means for working out purely communicative models of society. Above all, he lacks a concept of social interaction and rationality that combines a communicative/dialogical process and a purposive/success/means-ends orientation, which would enable him to give at least a *prima facie* communicative account of the enormous volume and variety of social encounters where agents pursue through other agents goals that are not regulated either by shared norms or by shared values. And such encounters have a pervasive presence in all the fields of social activity, including those considered by Habermas to belong to the lifeworld, private and public. Interactions which are at once communicative and purposive are categorially ruled out by the early and late Habermasian dichotomy between communicative and purposive action and

rationality; a conceptual divide according to which the communicative is necessarily oriented to values and norms (and never to success concerning the use of means for the attainment of ends) while the purposive necessarily involves a causal relationship (and never a dialogical one). They are also excluded by his tripartite conception of practical rationality, consisting of moral, ethical and pragmatic discourses. Finally, they are left unaccounted for by the Universal-Pragmatic analysis of linguistic communication in terms of the validity claims to truth, truthfulness and normative rightness.

To be sure, the notion of fair bargaining, which, though essentially strategic in character, plays a central role in the later Habermas's idea of rational political, and in particular legal, decision making, comes sufficiently close to a discursive form of communicative purposive interaction. And as such, it could have been used by him to explore – and carry out well-grounded assessments of the practical viability of – possible forms of societal organization that were free from media-steered systemic social integration and almost free from strategic action itself. Nonetheless, he has chosen not to undertake such a task in any systematic way.

In an attempt to fill the gap identified in Habermas's conceptual framework regarding action and its rationality, and also address the corresponding shortcomings at the deeper level of the rational reconstruction of the universal pragmatics of language, I have developed, and provided an alternative rational-reconstructive justification for the concept of restrictedly purposive communicative action and a revised model of practical rationality that includes the discursive equivalent of that concept, namely communicative negotiation. In so doing, I have drawn on Habermas's own idea of *weak* communicative action. At the same time I have suggested – without, however, going into it in any detail – that the potent early Habermasian idea of depth-hermeneutical, emancipatory self-reflection, is

developed in a dialogical direction so as to be usable as an aspect of communicative negotiation and communicative practical discourse in general – a move that the later Habermas is distinctly reluctant to make.

In what follows, I shall put forward some preliminary, tentative ideas on how the revised model of practical rationality, in particular the concept of communicative negotiation, can be made in practice applicable to the complex, modern social reality: what practical forms might its employment take so as to render media-steered systems, and causal-manipulative, strategic action as such, unnecessary, or at least needed in as limited and as communicatively controlled areas of social action as possible?

To begin with, let us delimit the *in principle* scope of social applicability of the concept of communicative negotiation. According to our previous analysis, that scope is enormous, extending over all social interactions that involve the use of others as means for what are, immediately or in the last analysis, actor-relative reasons. There are three large categories of such interactions: a) Interactions pursuing through others goals that are – in the context of these interactions – egocentric. This category includes “lone” social actions that do not *directly* require others in order to realize their goals – and do not therefore *prima facie* constitute interactions – but do so *indirectly* by having to be permitted by norms already consented to by others; b) Interactions pursuing through others goals regulated by contextually shared norms; for such norms, as has been argued, ultimately depend on universal or other large-scale communicative negotiations; c) Interactions pursuing through other agents goals regulated by interests/values that are contextually shared (not spontaneously, but) as a result of communicative negotiation. The concept of communicative negotiation, therefore, is applicable

both to the interactions not covered by Habermas's conceptual framework at all, and to those covered by it only *prima facie*.

There are, in fact, only the following types of interactions that do not involve the use of others as means for actor-relative reasons, and to which, therefore, communicative negotiation is not applicable: interactions in which (a) an agent asks one or more other agents to act in a way that is in accordance or at odds with interests/values shared spontaneously, i.e. prior to any negotiation between them; (b) an agent exhorts one or more agents to act in a way that is in accordance or at odds with the latter's interests/values; and (c) agents help each other to identify/determine their own interests, values, self-identities – their conceptions of the ethically good, in general.

It goes without saying that communicative negotiation in the precise sense used here is not appropriate for non-interactive activities such as the self-reflective processes through which agents identify/determine their ethical preferences. The forms of practical rationality appropriate for these types of cases are, of course, varieties of ethical discourse.

Similarly, communicative negotiation is not appropriate for practical (as distinct from aesthetic or playful) relations of human beings to nature, or more precisely to objects or objective processes. Actions involved in such relations must be at once oriented to success and causal-manipulative; they are thus amenable to the instrumental (as distinct from the strategic) aspect of the pragmatic form of practical rationality. To be sure, these actions, and the corresponding rationality, can be and need to be internally differentiated, according to the possibilities of "interaction" offered by different types of objects and objectified processes, e.g. by inanimate objects, biological organisms of different levels of complexity and, especially, consciousness, and ecological systems. However, such internal

differentiation and respective development of the concept of instrumental action and rationality, though important and challenging, is beyond the scope of this study.

Having delimited the scope of the *in principle* applicability to social interaction of the concept of communicative negotiation, and thereby appreciated its truly enormous dimensions, we may return to the question of the concept's applicability and usefulness *in practice*. The problem arises as three major obstacles to the concept's direct application to every specific interaction that lies within the range of its applicability become apparent: (1) Its deficiencies with regard to the overall long-term interest-satisfaction of agents; (2) its social costliness; and (3) its motivational deficit.

1) As shown in section 3.5, the direct employment of the discursive method of communicative negotiation fails to secure the maximum communicatively attainable overall interest-satisfaction of agents with regard to a certain type of interests, those which we have termed "universalizable" or "generalizable" interests. These are interests whose maximum long-term (communicative) satisfaction for each agent requires that in the short term they must not be satisfied equitably, as would be demanded by a direct communicative negotiation; instead, interactions which involve such interests must always – within given contexts – be resolved in a standardized, norm-guided way that is fully in favour of the one who asserts them and at the expense of the other.

2) The discussion in section 3.7 makes abundantly clear how costly, in terms of social time and effort, is rational, value-maximizing, communicative (and often non-communicative) negotiation. For as we have seen, negotiation aiming to exhaust its value-creating potential for all parties is an extremely complex and demanding discursive undertaking, involving interpenetrating ethical and

pragmatic processes; namely, processes concerning the hermeneutic, including depth-hermeneutic, determination of the parties' relevant interests, on the one hand, and the creative development of means that optimally satisfy those interests, on the other. If communicative negotiation of this kind were to be employed with regard to every social interaction to which it is in principle applicable, there is little doubt that society would be led to paralysis.

3) For a negotiation to be conducted in a communicative, rather than a strategic, mode, all the parties involved must be sufficiently motivated in that way. Even if only one party decides to negotiate strategically, the negotiation will not be communicative. Very often, however, at least one of the parties is convinced that by negotiating strategically it would gain more than its equitable share, which would be allocated to it as a result of communicative negotiation. For that party, in other words, it would be fully purposively rational to negotiate strategically, and only restrictedly purposively rational – though fully rational from a communicative point of view – to negotiate communicatively. Thus, that party would be motivated by what Habermas calls the strong motive force of interests to negotiate strategically; while at the same time it would be motivated by what Habermas refers to as the weak motive force of (communicative) reason to negotiate communicatively. Evidently, the balance of motive forces would be in favour of strategic negotiation; hence, the motivational deficit with regard to employing communicative negotiation, in practice, in all the interaction situations where it is in principle applicable.

But of course, to be fully appreciated, the problem of applying communicative negotiation in practice must be seen in its true context, that of modern society. For certain fundamental features of modernity make the problem indeed formidable:

1) The great variety of interactive situations/coordination problems which are potentially open to negotiation

2) The complexity of each potential negotiation, owing to the modern forms of its two major, interrelated, component processes, the ethical and the purposive; owing, more precisely, to the enhanced reflexivity of the modern individual and collective subject, and the sophistication of modern technical/pragmatic issues.

3) The plurality of stakeholders in any modern interaction situation, and the complex ways in which each stakeholder is affected by different possible solutions to such a situation.

4) The increased pluralism of values and interests, in comparison to pre-modern societies, which in turn creates a greater demand for a negotiated resolution of interactive situations.

5) A sharply reduced stock of traditional, ready-made, unreflectively used solutions to interactive situations, thus yet more need for explicitly negotiated deals. This social reflexivity is in a relationship of continuous mutual reinforcement with process or social change.

3.8.2 Towards overcoming obstacles to the revised concept's practical application

Efforts to make the social application of the revised model of practical rationality – communicative negotiation in particular – more feasible can move in two directions:

1) Reducing the need for the *direct* employment of communicative negotiation and other complex forms of practical discourse by replacing them, to an optimal degree, with simpler, time- and effort-economical, *indirect* ways. This can be

achieved by using *devices that structure social actions, and interactions*, by “storing up” within themselves the outcomes of a manageable range of generalized communicative negotiations or other complex discourses, and then “releasing” these outcomes, in relatively simple, standardized ways, into large classes of specific (inter)action situations emerging over time, which would otherwise have to be tackled by means of the costly, complex methods.

2) Creating motivational, competence and other *conditions* that make possible (a) the development and proper functioning of the above mentioned social action-structuring devices, and (b) the direct application of the complex forms of practical discourse in the still extensive range of situations which remain outside the scope of the structuring devices, and thus need to be dealt with in terms of such direct application.

3.8.2.1 Social Structuring Devices

Modern societies meet many of their truly immense coordination and integration needs by structuring social interactions and decision making processes in various ways, among them the ones Habermas interprets as media-steered subsystems. Although existing structuring methods are anything but free from the strategic mode of action, in many cases they can, nonetheless, be drawn upon in order to develop, in theory as well as in fact, communicative devices that would facilitate the application in practice of the idea of communicative negotiation, and more generally of the revised conception of practical rationality. Some such devices are sketched below.

1) *Communicatively established normative frameworks*

In section 3.5 we have referred to the possibility of communicatively developing contextually binding sets, or frameworks, of norms, over and above the framework

of moral norms, which is binding in the context of universal human interactions. Examples of non-moral normative frameworks can be the laws of a state, the rules and regulations of an economic political, administrative, educational or other organization, the charter of an association of states or other organizations, etc. These frameworks are binding for agents insofar as the latter act in a respective, contextually determined capacity: moral norms are binding for agents in their capacity as (abstractly conceived) human beings; the laws of a state are binding for the citizens of that state; the rules of a school are binding (differentially) for the pupils, the teachers, other staff, and the school's governing bodies; the charter of an association of organizations or states is binding for the association's members, etc. For a normative framework to be communicatively determined it must be validated in terms of a respective generalization principle, applied with the participation, in principle, of all the stakeholders, that is, all those affected by the norms' general observance. In this way, their standardized general observance replaces – as a “condensed” equivalent of – a multitude of specific, situational, communicative negotiations.

As argued in section 3.5, such normative frameworks are possible only insofar as there exist, in a given context, what have been referred to as generalizable interests; that is, interests which, for each agent, are communicatively better satisfiable, in the long run, in terms of a generalized observance of a particular norm rather than by means of communicatively negotiating directly every specific relevant case. Context-relative sets of norms, however, can, if established and observed, go a good way towards overcoming two of the three major obstacles to applying communicative negotiation in practice: its lower levels of overall satisfaction of generalizable interests, and its forbidding costliness in terms of social time and effort.

Societies, or indeed the international community, can be furnished with comprehensive sets of communicatively established normative frameworks, structuring a sizable portion of interactions in each of the corresponding social contexts.

The frameworks belonging to such a society-wide or global set can be interlocking, in the sense that they stand to each other in relations of inclusion, exclusion or overlap.

2) Communicatively negotiated collective interests

A normative framework specifies what are, within its context of application, obligatory, prohibited or optional (i.e. permissible but not obligatory) actions/goals. Sometimes, some of the optional goals are regulated by a narrower normative framework, subsumed in the original one; as, for example, goals which are optional within the context of a certain state but become obligatory or prohibited in the context of an educational institution operating within that state. Otherwise, optional goals – insofar as they need to be (communicatively) realized by means of others, and if not spontaneously shared by those others – must be pursued in terms of communicative negotiations.

However, certain clusters of such optional goals can be found to follow from or be ruled out by relatively enduring interests or ethical choices (values, policies, projects) which, following communicative negotiation are collectively shared by a sub-group – or in some cases by the total group – of those for whom the said goals are optional. In those cases, the interactions through which those goals are pursued can be structured around those collective interests. Instead of being arduously coordinated by means of a multitude of individualized communicative negotiations, they can be, much more simply, regulated in terms of collective

interests, which are determined each by a generalized communicative negotiation among those concerned. The latter process is comparable to the large scale negotiations comprising the application of the universalization/generalization principle for determining norms which represent universalizable/generalizable (as distinct from collective) interests and serve to regulate – more generally and more enduringly in their case – certain other categories of goals and thus actions.

3) *Communicatively developed and pre-evaluated means-packages*

As argued in section 3.7.1.2, each proposed solution to a means-ends problem – i.e., each “means-option” – is value-laden through and through. It constitutes what we have called a “value-ensemble”. This is true of solutions to all purposive-rationality problems, including those involving purely causal-manipulative relations to objects/objectified processes, as well as those involving communicative interactions with other agents (thus communicative negotiations) in which causal-manipulative relationships, when significantly involved, constitute only one aspect of the means-ends problem.

As value-ensembles, solutions to purposive problems (i.e. means-options) must be acceptable to the agents concerned in the light of their relevant interests: the relevant interests of a single, individual or collective, agent, in cases of instrumental, subject-object problems, or the relevant interests of all the interacting agents concerned, in cases of communicative negotiations. Usually, as parts of means-options are included whole, ready-made objects or processes: tools, machines, computer software and other technological items; sociotechnical and other processes, such as manufacturing assembly lines and standardized customer service or back-office procedures; or suitably adjusted objects of nature, such as domesticated animals and plants, or other natural resources. Seen from this angle, even whole organizations, or institutions, with the normative

frameworks embodied in them, can be construed as parts of means to realizing (large scale collective and other) interests.

Such objects and processes constituting ready-made parts of means-options – let us refer to them as *means-packages*, for short – can normally be employed in diverse, optional ways in order to produce diverse sets of results. Nonetheless, they are only partly neutral with respect to the results their uses produce. For on the one hand, there are results none of their uses can produce; while on the other hand, and more importantly, there are results which are unavoidably produced by any of their (optional) uses. A routine office procedure, for example, cannot be used as a means to developing the innovative abilities or the initiative of staff. On the other hand, irrespective of whether such a procedure will be followed to deal with case A, B, or C (with regard to which the procedure is neutral), it will have certain more or less invariable effects; e.g. a certain expenditure of materials, certain telephone costs, and certain degrees of staff fatigue, boredom, etc. Similarly, a saloon car cannot fulfill the professional requirements of a plumber, but it could probably be used equally well (i.e. neutrally) for a family outing or as a robbery getaway vehicle. In the latter two cases, however, it would consume equivalent amounts of fuel and emit equivalent amounts of polluting gases into the atmosphere. Thus, means-packages are characterized by a combination of value-ladenness and value-neutrality. In a way that is interestingly analogous to the functioning of normative frameworks and collective interests, they necessitate or rule out certain results, while leaving others open.

Now, as we have seen in 3.7.1.2, in cases of communicative negotiations, as well as in other kinds of purposive discourses, means-options are developed and assessed in a deliberative process involving a complex interplay of values and interests on the one hand, and technical possibilities on the other; a process which

has been referred to – after the early Habermas – as a “dialectic of potential and will”. The social costliness of communicative negotiations or other discursive processes involving the creation and evaluative assessment of means-options could be reduced to a considerable degree, if ready-made means-packages that can be used as integral parts of large numbers of specific means-options, over substantial lengths of time, were communicatively evaluated and approved in advance, and then their value-profiles taken for granted and integrally incorporated in the relevant deliberations. An extensive stock of such communicatively developed and value-profiled objects and processes would greatly enhance the social efficiency of practical discourses.

If comprehensively set up and utilized, structuring devices of the three forms discussed above – (i) normative frameworks, (ii) collective interests, and (iii) pre-evaluated means-packages – would go a long way towards addressing difficulties with regard to the practical application of the revised model of practical rationality in conditions of modernity. More precisely, normative frameworks would overcome direct communicative negotiations’ ineffectiveness concerning the overall satisfaction of generalizable interests; whereas devices of all three types would contribute considerably to the social efficiency of communicative discourse.

To be sure, such devices, by virtue of their action-structuring nature, introduce at the same time a certain rigidity in social action. This rigidity, however, can be kept to levels which ensure that it is amply compensated for by the added social practicality. Moreover, it is unavoidable that society’s multiform discursive activities are constantly in a self-propelled, dynamic process of evolution; their subject-matter always being renewed and enriched; their outcomes being always revocable. This, surely, applies also to discourses about structuring devices; that is, discourses concerning the endorsement of a specific norm, a collective interest,

or a value-profile, as well as discourses that might concern the level of the social flexibility that should be allowed by the totality of society's structuring devices at any given time.

3.8.2.2 Enabling Conditions

The existence and proper functioning of social action-structuring devices – and therefore their above mentioned social effectiveness and efficiency benefits – continue, of course, to rely on communicative discourses, even though fewer and less varied, or of simpler forms. The establishment of norms and collective interests depend, as we have seen, on complex communicative negotiations, whereas the creation and evaluation of objects and processes in the form of ready-made means-packages require complexly interrelated ethical and pragmatic discourses. Both these kinds of communicative discourses, insofar as their practical applicability is concerned, would still face social costliness as well as motivational impediments. Motivational problems would arise also with regard to the functioning of the first two kinds of structuring devices; that is, with regard to the general observance of norms, and to conformity with communicatively negotiated collective interests.

In addition, the use of structuring devices, though it considerably narrows down the range of specific (inter)action situations that are amenable to direct communicative negotiation or other complex forms of discourse, it by no means does away with it entirely. Goals not governed either by norms or by values, including spontaneously shared or communicatively negotiated collective interests, still need to be (communicatively) pursued in either of two ways: through others, by employing complex, direct communicative negotiations; or through causal-manipulatively relating to objects or objectified processes, on the basis of specific pragmatic discourses involving the complex development of means-options, of

which pre-existing and pre-evaluated means-packages are only parts. All these complex forms of discourse, which continue to be needed for dealing with the social actions and interactions that would remain “unstructured” by structuring devices are, surely, still open to social costliness problems; whereas those of them that are of a communicative nature would face motivation problems as well.

For all the above reasons, the practical applicability, in modern society, of the revised model of practical rationality would depend on the existence of certain conditions; namely, conditions that would sufficiently reduce the social costliness of, and reinforce the motivation for, the discourses (communicative or pragmatic, as the case might be) which are needed on the one hand for the establishment and functioning of an optimal set of social structuring devices, and on the other hand for dealing with social situations left unstructured by such devices. Some conditions of this kind are tentatively outlined below:

1) *Institutionalized sanctions*

The often serious motivational problem of adherence to communicatively agreed normative frameworks can, at least in part, be dealt with – as it in fact is, with regard to existing social norms – with the introduction of positive and negative incentives or sanctions, i.e. rewards for adhering to norms and penalties for violating them. Sanctions are normally equipped with means of enforcement, based on institutionalized authority. Systems of sanctions, together with their means of enforcement can take stronger or milder, stricter or looser forms. These vary between different normative contexts (e.g. national state, international organization, economic enterprise, educational institution) as well as within each such context, according to the gravity and sharpness of the relevant norms.

Needless to say, the institutionalization of sanctions, like the establishment of normative frameworks, must be effected through communicative deliberation processes, with the participation of all the relevant groups of stakeholders.

2) *Shortcuts to decision making*

One way of reducing the time and effort expended on discourses, and thus increasing their social efficiency, is by introducing certain accelerated ways of concluding argumentation, or “shortcuts” to reaching discursive decisions. Three such shortcuts – already widely used – can be referred to:

a) Having decisions taken by *majorities* rather than on the basis of unanimous agreement, or consensus, as would be required, in principle, by the concept of unconstrained discourse

b) Limiting participation in certain discourses to *representatives* of the different groups of stakeholders rather than extending them to every individual agent affected.

c) Adopting the first option considered satisfactory, rather than searching for the best possible option. This so called “*satisficing*” approach to decision making is considered to be an aspect of what is sometimes referred to as “bounded rationality”.⁶

3) *Individual competences*

An important prerequisite for the social application of practical rationality is the possession by individual agents of well developed competences for communicative negotiations and other forms of discourse. These competences include self-reflective, hermeneutic and technical/pragmatic skills, a creative imagination, and, more generally, postconventional cognitive and moral abilities.

With regard to communicative negotiation, in particular, such advanced competences will not only enhance the efficiency of the process, and its effectiveness in relation to the optimum satisfaction of the interests of all the parties; by virtue of the latter, they will also mobilize the strong, interest-based motivational forces of the parties to negotiate in the communicative mode, rather than resort to strategic techniques in an effort to secure greater interest satisfaction for themselves.

4) *Equality*

(a) Deliberative equality among participants is of course an essential part of the definition of communicative discourse. It is a presupposition of a discourse's rationality and legitimacy. This issue, however, and the great difficulties in establishing true deliberative equality in practice, lie beyond the scope of this study.⁷ Here, suffice it to say that equal treatment in the context of negotiation and other discourses – in the sense of an equal opportunity to contribute to the discourse, to be taken seriously and to have an effect – provides participants with a sense of ownership, and boosts their motivation both to participate in the discourse itself and to abide by its resolutions.

(b) Another aspect of equal treatment which is important from the point of view of agent's motivation to act communicatively, is the fairness with which observance of norms is monitored and respective sanctions applied. Such fairness depends on the effective and unbiased functioning as well as the non-corrupt character of the state or other institutions invested with the relevant authority.

(c) A factor that strongly tempts participants to adopt a strategic approach in a negotiation is their belief that the balance of power (in the form of physical or military strength, economic resources, connections or alliances, information, persuasive ability, etc) is sufficiently in their favour to enable them to coerce or

deceive the other participants into accepting a deal which in a communicative negotiation they would reject as inequitable. Thus, the more equality of negotiating power there is in a society, or on the global stage, the more motivationally feasible becomes the employment of communicative negotiation.

5) *Socialization into social structuring devices*

According to our proposal regarding action-structuring devices, a sizable proportion of everyday social action situations could require agents to observe norms, respect commitments to collective interests or appropriately employ objects and processes as means to ends. In every situation of this kind, agents have to recognize that a certain structuring device is implicated, decide what the appropriate response to it is and act accordingly. It is obvious that carrying out these operations in an explicitly deliberative manner – though admittedly much simpler than considering every such case afresh, without the help of structuring devices – would be so inefficient as to bring society to a standstill. It would indeed be equivalent to a pianist playing a sonata – even though not having to compose it in the process – by focusing attention on every individual note and every necessary movement of the fingers. Evidently, these activities would be impossible if they did not rely heavily on agents' intuitive capacities – ones that need to be purposely developed through prior learning.

Thus, it is a prerequisite of (communicative) social life that agents are properly socialized into the world of the action-structuring devices which society considers well established at any given time. This will provide them with the intuitive ability necessary for dealing with situations involving those devices in a rather spontaneous, unmediated way. However, such a socialization must by no means amount to the inculcation of an unreflective, uncritically traditionalist attitude to the existing set of structuring devices. It has to remain compatible with the

development of agents' postconventional, critically-reflective stance towards existing social reality; of their readiness to reopen issues and revise previous decisions. For such a stance is a necessary aspect of practical rationality, albeit in a sense of the term not entailing constant explicit deliberation.

3.8.2.3 Lines of Further Research

The "structuring devices" and "enabling conditions" – which we have tentatively sketched above – constitute only a few generic elements of a practicable model of society based on a conception of practical rationality that crucially incorporates the notion of communicative negotiation. These elements, however, indicate the directions in which further research needs to move if such a model is to be adequately developed.

To begin with, each type of generic structuring device and enabling condition referred to requires considerable elaboration. At the same time, additional structuring devices, enabling conditions or other generic elements might be explored. But most importantly, these generically applicable elements need to be specifically applied to each particular area of social life in order to create communicative "sub-models" of these areas. Sub-models of this kind must be developed for all the fields of social activity which Habermas considers as parts of the lifeworld: family life, education, the political public sphere, etc; but they must also be developed for the two major sectors of society ceded by Habermas to media-steered subsystems of strategic action: state administration and the economy. With regard to the latter two, in particular, the challenge is to construct models which avoid systemic-strategic solutions, while at the same time not being communicative versions of historically failed, excessively structured and rigid forms of social organization; that is, models which efficiently coordinate and integrate the immense range of relevant social activity through an optimal

combination of communicatively developed, and adhered to, structures, on the one hand, and non-structured, direct communicative negotiation and other discourse, on the other. Such models might optimally combine, in the case of state administrative organizations, communicatively developed formal elements and post-bureaucratic, participative aspects, while in the case of the economy, communicative planning with “communicative market” elements.

The work of elaborating practically feasible models of communicatively rational social action – sectoral and societal – would have to be carried out in dialogue with descriptive as well as normative/prescriptive social theory: both the general theory of society and disciplines that specialize on particular sectors of social life, such as economic theory, organization theory, etc. Through such a dialogue, our model-construction will meet reality “halfway”, to use a favourite Habermasian expression. Existing social reality and its developmental trends will thereby be interpreted in the light of the communicative conception of social action – at a given stage of its development – in search of elements consistent with that conception, while that conception will, in return, be enriched by drawing on the elements thus identified. We shall, therefore, have a dynamic process of reciprocal reinforcement between – or a “dialectic” of – describing social reality on the one hand, and developing practically applicable communicative models of it, on the other.

It is of course conceivable that this dialectic does not yield a practically applicable model of a *fully* communicatively rational modern society; i.e. a model that is able to address with a sufficient degree of success the effectiveness, efficiency and motivational impediments to such a society. It may indicate that the practically feasible conception of modern society that comes nearest to the communicatively rational ideal will have to include pockets of strategic action – perhaps even of ones akin to what Habermas describes as media-steered subsystems, – for

example in parts of economic life. But now, owing to systematic explorations in the form of the above mentioned dialectic of interpretation and model-construction, it will be more likely that such unavoidable pockets of strategic action are kept to a minimum and are, communicatively, “anchored” and controlled as tightly as possible.

Furthermore, those unavoidable strategic pockets might, with time, become more and more restricted and communicatively controllable, as social reality itself, and its inherent potential, evolves. For the dialectic of interpretation and model-construction can be dynamic in an additional sense: it can involve the interpretation of a social reality which is itself *changing* through a similar reciprocal relationship to model-construction. This is so, because the communicatively rational models of social reality, over and above their interpretative/descriptive social-theoretic role, can also function as normative/prescriptive standards for a critical social theory intent on rationally legitimized social improvement, in the broad Frankfurt School tradition, including Habermas’s own work. Such a critical theory, in addition to developing its normative/prescriptive standards, is surely interested in exploring political and other practical means for promoting social change in the directions specified by those standards. And to that end, critical social theory needs to be in a continuous process of dialogue and mutual learning with kindred political and social movements. On this understanding of the vocation of critical social theory, one might, in fact, even entertain the hope that society could reach a stage when social interaction is, for all intents and purposes, communicatively rational; strategic action being retained only in areas in which it is deemed not unavoidable but desirable, as for example, in competitive games like football and chess.

CONCLUSIONS

1. The theme of practical rationality and its context

In the last few centuries, Western societies, and increasingly non-western ones too, have relied on the human capacity for reason as their dominant source of guidance for action in its different forms, individual and collective, private and public. Through a long process of European historical development, beginning for all intents and purposes in ancient Greece and, after the lengthy medieval interlude, culminating in the 18th century enlightenment movement, rationality has gradually come to discredit and overshadow alternative action-steering sources, primarily sacred texts, divine revelation and unreflective traditionalism. Thus rationality became the hallmark of modernity and the core of what was claimed to be modern humanity's potential for a continuous progress towards happiness, justice and freedom.

However, starting in the very age of enlightenment itself and to an increasing degree since, rationality, and modernity in general, have come under severe criticism from diverse quarters. Rationality is charged with lacking the capability to deliver what was promised, and for being in fact responsible for serious, specifically modern problems, such as the ecological crisis, loss of meaning, and novel, subtler, forms of domination. At the same time, the growing complexity of social conditions makes decisions ever more difficult and creates an ever greater demand for action-steering capacity. It looks, therefore, as if we are facing what can be called a "guidance-deficit" for human action.

On the basis of this diagnosis, and on the further assumption that pre- or non-modern action-steering principles cannot credibly be turned to for meeting today's challenges, this study has been intended as a contribution to the rich ongoing inquiries and debates that aim to explore rationality's potential, as well as its

limitations, for providing guidance to human action, in principle as well as in practice. This task is approached mainly through a critical examination and partial revision of Habermas's outstandingly important and fertile work on practical rationality, both in its early, 1960's phase and in its later communicative stages.

2. Habermas on practical rationality

The early Habermas rejects his Frankfurt School mentors' views encapsulated in their "Dialectic of Enlightenment" thesis. According to those views, modern reason is instrumental reason and cannot take any other form. As such, it turns everything it is applied to, as well as the human beings who use it, into objects of domination. Thus, modern, instrumentally rationalized society is a totally administered, reified society. Habermas maintains that the phenomena of reification and domination in modern society – which are for him far less widespread than Horkheimer, Adorno and Marcuse believe – are not due to inherent properties of instrumental rationality but rather due to its being used beyond its proper domain, essentially that of the practical interchange between human beings and nature. The illegitimate extension of the application of instrumental rationality into the field of purely social interaction – including the development of social science on the methodological model of the natural sciences – is a symptom of technocratic consciousness, a way of thinking that confuses ethical-moral issues with technical ones.

Moreover, Habermas insists that in addition to instrumental rationality, which is appropriate to instrumental action, or *techne*, a second form of practical rationality is possible, appropriate to human interaction, or *praxis*, and referred to in this study as "praxial". Praxial rationality is concerned with determining values and norms that must guide the ethically and morally imbued domain of *praxis*. In his early work Habermas develops a conception of praxial rationality whose most important component is a form of "depth-hermeneutic" self-reflection, on the model

of Marx's critique of ideology and Freud's therapeutic method of psychoanalysis. Depth-hermeneutics combined causal-explanatory analysis and ordinary hermeneutic methods of self-understanding. This type of self-reflection could enable individual and collective subjects to emancipate themselves from self-deception and unconsciously self-imposed bondage, and to discover their authentic interests, values and identities. It could not, though, provide objective foundations for values or norms.

Habermas grew dissatisfied with this first version of non-instrumental, praxial rationality. In particular, he considered untenable the image of society as a kind of undifferentiated macro-subject capable reflecting upon itself, an image that was implicit in that first version. In addition, he took careful notice of critics' fears that the power-asymmetry inherent in the therapist/patient model on which the relationship between emancipator and emancipated-to-be was based, could have sinister political implications; especially, given what he acknowledged as being another shortcoming of his idea of social self-reflection: an insufficient distance between critical theory and political action. Thus, he eventually abandoned the concept of critical-emancipatory self-reflection, together with the philosophy of consciousness, which he held responsible for that concept's major shortcomings. In its place, he developed an alternative version of praxial rationality, worked out within an entirely new, intersubjectivist, communicative paradigm.

The communicative paradigm is based on the rational reconstruction of universal and inescapable features implicit in all linguistic communication. Habermas understands linguistic communication as action primarily oriented to understanding, or as communicative action. The rational-reconstructive method is said to establish normative commitments that are universally binding on all language users, thus on all human beings. According to Habermas, implicit in every speech act are three rationally redeemable validity claims: the claims to truth

about an objective, external world, to truthfulness concerning the speaker's own subjective world, and to rightness with regard to legitimate – ultimately moral – norms.

Implicit in each type of validity claim are the conditions for its justification. These conditions constitute rules of a corresponding form of rationality. Some of these rules are common to all the three forms of rationality implicit in the three types of validity claim. The three-fold rationality that Habermas reconstructs from communicative action is referred to as communicative rationality. Communicative rationality is not a process of deliberation carried out merely in the heads of separate subjects. It is a process of (potentially universal) intersubjective argumentation, or discourse, oriented to (potentially universal) agreement, or consensus.

Thus, in Habermas's view, the rational reconstruction of linguistic communication establishes three forms of rationality, or discourse: (1) Theoretical rationality, concerning factual truth. When employed in questions regarding the use of means in order to causally bring about an end, theoretical rationality becomes the foundation of "pragmatic" rationality. Pragmatic rationality itself is subdivided into instrumental and strategic rationality and corresponds, respectively, to instrumental and strategic action. The former type of action is about the causal manipulation of objects (that do not have purposes of their own) by purposive subjects. The latter concerns the causally effected manipulation of subjects (who are themselves purposive) by other purposive subjects. Such causal manipulation involves the use of deception and/or coercion. (2) Ethical rationality. It concerns the clarification of an individual or collective agent's authentic and sincerely expressed subjective states, including values (or conceptions of the good) and self-identity (who one is and who one wants to be). (3) Moral rationality. It

concerns the application of Habermas's version of the universalization principle to test the validity of universally binding moral norms.

The conception of practical communicative rationality that emerges from this scheme includes (1) the old instrumental type (internally differentiated into instrumental in a narrower sense and strategic), now termed "pragmatic", and (2) a new two-fold version of praxial rationality consisting of an ethical and a moral dimension. Ethical discourses cannot establish rationally valid identities or values. Their function – just like that of the early, critical-emancipatory self-reflection – is to clarify the subject's authentic values and identity, though now by using only ordinary hermeneutic methods, since Habermas has given up the depth-hermeneutic methodology which was available to the early version. Moral discourses, however, are intended by Habermas to be strongly normative, in the sense of providing rational justification for universal moral norms.

Habermas employs the new conceptual framework, principally the distinction between communicative and strategic action and rationality, to develop a new critical theory of modern society. He begins by distinguishing between lifeworld and system. The lifeworld consists of the domains of (private or public) social activity mainly concerned with the production of meaning and identity: family life, education, the political and legislative sphere. Social action in these domains ought to be predominantly communicative, and coordinated in terms of agreements based on the action orientations of agents. Domains of social life concerned primarily with the reproduction of material life – mainly the economy and state administration – cannot, according to Habermas, under conditions of modern complexity realistically rely on communicative action and communicative modes of action coordination. The domains of economy and state administration must, instead, be organized as systems of strategic action steered by the media of

money and power, respectively, which is said to be the way they are in fact organized in modern societies.

The problem with modern societies, according to Habermas, is that the media-steered systems of the market economy and state administration are not as tightly controlled by the communicative powers of the lifeworld as they ought to, and tend to turn back upon the lifeworld in a “colonizing” manner. They encroach on its different domains and increasingly turn them into fields of media-steered strategic activity, in their own image. Thus, critical social theory, and critical political practice, should concentrate on securing the communicative character of the lifeworld, by reversing and stemming its colonization by the systems, and keeping the systems themselves institutionally well anchored to the lifeworld and firmly under its communicative control.

3. Critical and constructive contribution

The present study largely remains within the broad framework of Habermas's communicative paradigm, taking it to form a sufficiently sound and fertile basis from which to explore the central issues concerning practical rationality. It is thus accepted that the application of Habermas's rational-reconstructive method on linguistic communication can, via the identification of implicit, universal validity claims, yield normatively binding conclusions regarding the nature of rationality, theoretical and practical. It is also accepted that the kind of rationality thus specified is a process of potentially universal, communicative discourse, oriented to reaching agreement, or consensus, and carried out on the basis of the equal treatment of everyone's relevant contributions, and the exclusion of all coercion and deception, and generally of all force other than that of the better argument.

But at the same time, this study identifies, internal to this general framework, certain flaws in Habermas's work relating to practical rationality, and proposes

suitable remedies. The weaknesses mainly concern the typology of action and rationality, the classification of validity claims, and the implications of those for moral and social theory.

To begin with, an important conceptual gap is located in Habermas's typology of action and its rationality. The dualism of *techne* and *praxis*, or of strategic and communicative action/rationality, which is present in both his early and later work, categorially excludes action and rationality which concern the pursuit by means of other agents, in a communicative, agreement-oriented way, of agent-relative, or "egocentric", goals – i.e. goals not governed by shared norms or values. This is due to the assumptions that means-using (or purposive, or success-oriented) action is *ipso facto* causal-manipulative, and that communicative action is *ipso facto* oriented solely to shared norms or values. In view of these assumptions, the pursuit of agents' egocentric goals through using other agents as means, can only be conceptualized by Habermas in strategic terms.

A corresponding lacuna is found at the level of validity claims. None of the three types of validity claim seems to cater for requests, or imperatives, not governed by shared norms or values, or more generally, for bargaining-related or other speech acts whose rational redemption would require an appeal to the compatibility between the egocentric interests of different agents. Thus, such speech acts cannot be part of linguistic communication in its Habermasian interpretation as primarily communicative, oriented to understanding. They can only be primarily strategically oriented utterances, in contradiction to that interpretation.

The significance of these conceptual gaps cannot be overestimated; for a very large proportion of social interactions, in all fields of social life, including the ones belonging to what Habermas describes as the lifeworld, are unavoidably of the kind that these gaps conceptually ban from the realm of potentially communicative

action, thus condemning them to be strategic, i.e. necessarily based on forms of coercion and/or deception.

In some writings that followed his initial comprehensive presentation of the communicative paradigm, Habermas puts forward certain ideas which move in the direction of addressing the above deficiency. These ideas are mainly encapsulated in the notions of “fair bargaining” and “weak communicative action”. Both these notions remain, in different ways, insufficient to the task. Fair bargaining is capable of covering the whole range of interactions kept out of the communicative realm by the above mentioned lacuna; however, though linked to moral, thus communicative, principles, – hence “fair” – it does so indirectly, continuing in the last analysis to rely on the balance of forces, thus being strategic. “Weak” communicative action, on the other hand, while being interpretable as genuinely communicative (despite Habermas unjustifiably treating it as communicative action of an inferior kind) is capable of covering only some of the interactions in question: those interactions which all agents involved would be fully justified, with respect to their egocentric interests alone, to pursue without the use of coercion or deception. Moreover, the interactions treated previously as non-communicative and now as weakly communicative are still not accounted for by Habermas’s system of validity claims, which remains unchanged.

In order to render potentially communicative the entire range of interactions which have been excluded from that category by the conceptual gap in question, this study has developed, and justified in terms of a revised system of validity claims, the concept of *communicative negotiation*.

A communicative negotiation is a social interaction in the form of a bargaining process which is carried out by purely communicative means – i.e., without the use of any coercion or deception – and which aims at a resolution, or deal, which

is acceptable to all parties as satisfying their relevant egocentric interests, not necessarily fully, but to the greatest degree achievable by those, communicative, means. It is, in other words, a restrictedly purposive communicative interaction, with its own distinctive form of rationality. The concept has been elaborated – partly by drawing on Rational Choice negotiation theory – with a view to explaining how to enhance as much as possible the communicatively achieved interest-satisfaction of the negotiating parties. Such a communicative maximization of interest-satisfaction, over and above its intrinsic value, could render the strategic pursuit of those interests unnecessary, or at any rate less attractive, and thus minimize agents' temptation to resort to it.

In the light of the concept of communicative negotiation, Habermas's understanding of morality has also been reconsidered. In particular, applications of Habermas's universalization principle to test moral norms have been interpreted as multiple, large scale` communicative negotiations. Thus, like any other negotiations – indeed like any other specific social interactions – they are not necessitated by the rationally reconstructible structure of linguistic communication. Also contingent with regard to language – and depending on the existence of universalizable interests – are moral norms, in the sense of norms validated by the universalization principle, as distinct from similar, (akin to moral), rules of argumentation reconstructively found to be inherent in linguistic communication.

In the same way as in the case of morality, communicative negotiation is understood to underlie also communicative ways of establishing non-moral norms, valid for contexts in which agents participate in capacities other than the abstractly conceived human, which applies to the moral context. Such non-moral norms are, for example, the laws of a state, which are binding on agents in their capacity as that state's citizens, or the rules and regulations of an organization which are binding for that organization's members. The justifiability of non-moral norms in

any given non-moral context is contingent upon the existence of interests that are generalizable within that context.

On the basis of the concept of communicative negotiation and the subordination to it of the concept of moral and non-moral normative rightness discourses and moral and non-moral norms, Habermas's conception of practical rationality as a whole has been revised. The tripartite model of pragmatic, ethical and moral employments of practical rationality, later supplemented with fair bargaining, has been replaced by a model consisting of pragmatic, ethical and communicative negotiation discourses, the latter including processes of negotiated constitution of collective values and collective subjects, and also subsuming moral and non-moral discourses. The revised model acknowledges the mutual dependence and interpenetration between the different types of discourse, most notably the complex interrelationship between values and the means to their realization, an interrelationship referred to, after the early Habermas, as a "dialectic of potential and will". It is also suggested that depth-hermeneutics, which has been largely abandoned by the later Habermas, is reintroduced as part of ethical discourses, provided that it has been developed in a communicative direction and freed from power-asymmetry.

In addition to the revised model of practical rationality, the study has developed, as that model's rationally reconstructed linguistic underpinning, a correspondingly revised system of validity claims. More specifically, the validity claim to (ultimately moral) normative rightness has been replaced by the validity claim to *intersubjective acceptability*. This new type of validity claim caters both for speech acts directly oriented to the communicative pursuit of egocentric goals through other agents, and for speech acts that are indirectly so oriented, via appealing to moral or non-moral norms or to negotiated collective values, which depend on the former kind.

Habermas has defended his much criticized “surrendering” of two hugely important domains of social activity, namely the economy and state administration, to media-steered systems of strategic action, by arguing that coordinating and integrating these domains in purely communicative ways would be, under conditions of modern complexity, impossible in practice. However, given firstly the fact that much of the interaction in these domains concerns the pursuit of egocentric success through others, and secondly the absence from Habermas’s model of action and practical rationality of means by which to conceptualize such interactions in communicative terms, a communicative economy and state administration would be for him impossible not just in practice but also in theory. Furthermore, it would also be in principle impossible to have a fully communicative lifeworld. For a great deal of lifeworld activity, from politics to family life, also heavily involves, inescapably, success-oriented interactions.

The revised conception of action and practical rationality proposed here, by contrast, makes a fully communicative society in principle possible. However, the impediments to the practical application of this conception in conditions of modernity are admittedly great. There are two main kinds of such impediments: firstly the high social cost, in terms of time and effort, of inefficient, communicative discourses, communicative negotiations in particular; secondly the weak motivational force of purposively restrained communicative action and rationality, as opposed to the strong motivational force of strategic forms of action and rationality, which are oriented to the maximization of egocentric-interest-satisfaction.

The study explores, in a preliminary and tentative manner, ways in which the practical applicability of the revised conception of action and rationality, in conditions of modernity, can be facilitated. Proposals, in this respect, move in two directions. On the one hand they concern the creation of “social structuring

devices”, which are means of reducing the need for the most complex and inevitably less efficient forms of communicative discourse by replacing them with simpler, more efficient ones. Such devices include (i) communicatively established normative frameworks, (ii) communicatively negotiated collective interests and (iii) communicatively developed and pre-evaluated means-packages. On the other hand, proposals concern (communicatively created) “enabling conditions”. These are conditions for making the creation and functioning of social structuring devices as well as the undertaking of the remaining complex discourses (those which cannot be replaced by structuring devices) as motivating and as socially efficient as possible. Suggested enabling conditions are, (i) institutionalized positive and negative incentives regarding the observance of agreed norms; (ii) shortcuts to communicative decision making; (iii) developed competences regarding communicative negotiation and other complex forms of discourse; (iv) as much equality as possible among participants in communicative negotiations; and (v) the effective socialization of individuals into the world of the agreed social structuring devices.

4. Some directions for further research

Further theoretical work is no doubt required in order to elaborate the revised conception of communicative action and practical rationality. Particularly challenging, in this respect, would be efforts in the direction of developing a communicative form of depth-hermeneutical, critical-emancipatory self-reflection. Very important would also be further analysis of communicative negotiation, especially in cross-fertilizing dialogue with Rational Choice work in this field. Of particular mutual interest would be further exploration of the nature and role in communicative negotiations of ethical and pragmatic discourses and their interpenetration.

But beyond work on different dimensions of the pure concept of practical rationality, much additional research is of course needed also in order to develop adequate practically applicable communicative models of modern society and its different specific domains, especially those of the economy and state administration. Such model-construction cannot, surely, rule out the possibility that, at any given time, certain pockets of strategic action – even of media-steered strategic action – would be deemed practically inescapable, though now being more amenable to communicative control and kept to the minimum necessary. But even such a need for strategic action does not have to be static. There can always be efforts to reduce it further. The process of designing practically applicable, communicative social models can develop, not just through theoretical analysis but by continuously seeking and drawing ideas from a rich and constantly changing modern social reality, which, in spite of being still extensively strategic, already incorporates considerable communicatively oriented elements, as Habermas would be the first to stress. After all, the process of communicative model-construction can be – as it must – in a two-fold relationship with existing social reality: that of interpreting and describing reality from a communicative angle, and that of evaluating reality and helping it transform in rationally justifiable ways.

APPENDIX I

HUME'S SUBJECTIVE-INSTRUMENTAL CONCEPTION OF PRACTICAL RATIONALITY

The classical presentation of the Subjective-Instrumental conception of practical rationality is found in David Hume's *A Treatise of Human Nature* (1978), first published in 1739, and *Enquiries Concerning Human Understanding and Concerning the Principles of Morals* (1975), first published in 1748–1751.

In his deeply sceptical epistemological and ontological speculations, Hume demolishes fundamental categories such as those of *self* (1978: 252), *substance* (*ibid.*: 635) and *causality* (*ibid.*: 91–2), destroying thereby the foundations of the very instrument of this demolition, reason itself:

the *understanding*, [which Hume – unlike Kant – does not distinguish from reason] when it acts alone, and according to its most general principles, entirely subverts itself, and leaves not the lowest degree of evidence in any proposition, either in philosophy or common life (*ibid.*: 267–8).

On that account, one could justifiably count Hume as an important anti-Enlightenment figure. However, when Hume comes to philosophical issues that are more directly related to the conduct of everyday life – issues such as those concerning morality, politics, religion, or the explanation of natural phenomena – he suspends his "cold and strained" speculations which lead to "philosophical melancholy" (1978: 269), and operates, by and large, with common sense assumptions about the self, the external world and causality. Indeed, in these practice-oriented discourses Hume, as Ayer aptly points out (1980: 75), conceives of the world in entirely deterministic terms. Hume knows that human beings must

make such common sense assumptions if they are to carry out their lives as "real" selves in a "real" world. These, after all, are the assumptions he himself makes in his everyday life, not least when he merrily dines, plays back-gammon and converses with his friends (1978: 269). Radical scepticism cannot be sustained; for if it were, "[a]ll discourse, all action would immediately cease" (1975: 160). For good or ill, such is the "whimsical condition of mankind, who must act and reason and believe" (*ibid.*: 160).

In addition to these pragmatic considerations, Hume provides another, perhaps more philosophically respectable, reason for suspending his destructive metaphysical radicalism: The consistent sceptic cannot follow the path of rational doubt in a non-sceptical manner.

A true sceptic will be diffident of his philosophical doubts,
as well as of his philosophical convictions (1978: 273).

The impossibility of consistently doubting the possibility of reason is thus affirmed. But the precise nature of Hume's conception of reason's action-guiding role must be seen in the light of a negotiation between a self-constrained but nonetheless live scepticism on the one hand, and the imperatives of human existence on the other.

The outcome of this negotiation is encapsulated in the famous dictum:

Reason is and ought only to be the slave of the passions,
and can never pretend to any other office than to serve
and obey them (1978: 415).

The idea that reason is the slave of passions has two main aspects. On the one hand it denies that reason can determine the ultimate ends of life, assigning that

master-role to the passions (1). On the other hand it affirms that reason does have the capacity to be of service to the passions in their drive to attain those ends, by providing guidance concerning appropriate means to their realization (2).

1) For Hume reason has only two kinds of objects: "Relations of ideas" and "matters of fact" (1975: 25, 287). The former, the prime example of which is mathematical relations, constitute the province of formal, *a priori* reason. They are "discoverable by the mere operation of thought, without dependence on what is anywhere existent in the universe" (*ibid.*: 25). The latter are based on the evidence of the senses. Reason is able to make matters of fact its objects, and thus go beyond the evidence of the senses, only on the basis of the "relation of *Cause and Effect*" – the principle that every event has a cause (*ibid.*: 26). In contrast to the objective-substantive conception of practical rationality, Hume is convinced that in neither of these two areas of competence (its only ones) does reason have the capacity to make judgements about ultimate ends and values – about "right and wrong", "good and evil", "vice and virtue", "beauty and deformity":

The ultimate ends of human action can never, in any case,
be accounted for by *reason* (1975: 293).

Ultimate ends and values (and Hume is interested above all in questions of morality) neither consist in nor can be rationally derived from either formal relations or matters of fact (1978: 468–9. See also 1975: 285–94)); for ends and values are "entirely different" from formal relations and facts.

A fundamental difference between them, according to Hume, is that ends and values are inherently linked to motivation, whereas formal relations and matters of fact are not. Ends and values have an internal relation with volition, inclination to

action being an essential aspect of adhering to them. Speaking specifically of morality, Hume makes the point abundantly clear:

Morals excite passions, and produce or prevent actions...

If morality had naturally no influence on human passions and actions, 'twere in vain to take such pains to inculcate it (1978: 457).

On the other hand, knowledge of facts and relations of ideas – which is the sole business of reason – is dispassionate, not inherently linked with motivation:

Reason being cool and disengaged, is no motive to action (1975: 294).

Abstract or demonstrative reasoning... never influences any of our actions (1978: 414).

Reason is perfectly inert...wholly inactive (ibid.: 458).

Reason of itself is utterly impotent (ibid.: 457)

Another important difference, according to Hume, between relations of ideas and matters of fact on the one hand and ultimate ends and values on the other, is that the former are susceptible of truth and falsehood, whereas the latter are not. This makes reason – which "is the discovery of truth and falsehood" – unqualified to make pronouncements about ultimate ends and values (1978: 458).

Ultimate ends and values are, for Hume, exclusively the province of affective faculties – "passions", "sentiments", "feelings", "emotions" – which alone have the capacity to motivate.

[T]he ultimate ends of human actions... recommend themselves entirely to the sentiments and affections of

mankind, without any dependence on the intellectual faculties (1975: 293).

A true child of the Age of Enlightenment, Hume considers sentiments concerning moral and aesthetic matters to be uniform throughout humankind, due to "the structure of human nature". Such a universalism, however, is not a necessary component of the subjective-instrumental conception of practical rationality; it is in fact rejected by most other adherents to this conception.

Reason, for Hume, does nonetheless have an important role in guiding action. Though not in itself sufficient for determining ultimate ends, it can be of great service in specifying the most effective means to those ends. Through ordinary experience and, even more, through systematic scientific inquiry, "reason instructs us in the several tendencies of action" (1975: 286). It shows us, in other words, the consequences of different actual or possible courses of action – "it directs our judgement concerning causes and effects" (1978: 414). In this way, we can know the extent to which a particular course of action can bring about a particular state which is independently identified as desirable by sentiment. And it is precisely this sentiment which provides – indirectly, so to speak – the necessary motivation for undertaking whatever course of action reason specifies as an effective means to a desired end: "the impulse arises not from reason, but is only directed by it" (*ibid.*: 414).

Hume, in typically Enlightenment fashion, is thoroughly optimistic about the effects of systematically employing reason as an instrument of sentiment. He derived this positive attitude, firstly from his belief in the goodness of human nature – not least the benign character of our "moral sense" – and secondly from his high regard for science.

This optimism is shared by many who have since espoused the Humean conception of practical rationality, even if they do not agree with Hume's diagnosis of the universality and goodness of human sentiments. Most notable among the optimists are the advocates of economic liberalism, beginning with Hume's good friend Adam Smith. Such thinkers are confident that the mechanism of the market will ensure that the combined effect of the multitude of diverse economic actions of instrumentally rational individuals – however antagonistic their pursuits – is on the whole socially beneficial. Most 20th century non-cognitivist moral philosophers are also relatively sanguine.

There are, however, also important anti-Enlightenment adherents to the subjective-instrumental conception of practical rationality: thinkers who make a darkly pessimistic diagnosis of what they see as a modern world dominated precisely by the (Humean) instrumentally rational pursuit of non-rationally chosen ends.

APPENDIX II

KANT'S TRANSCENDENTAL CONCEPTION OF PRACTICAL RATIONALITY

This, transcendental conception of practical rationality has received its original and purest formulation in the works of Immanuel Kant – above all the *Groundwork of The Metaphysics of Morals* (1948), first published in 1785, and the *Critique of Practical Reason* (1996), first published in 1788 – and has since been greatly influential.

Prominent among English-speaking moral and political philosophers whose work incorporates strong Kantian elements are John Rawls and R.M.Hare. More importantly, from the point of view of this study, Habermas's own discourse theory of morality is a broadly Kantian one, albeit situated within an intersubjective, "communicative" philosophical paradigm rather than in Kant's subject-centred paradigm of the "philosophy of consciousness".

Here, I shall outline the main aspects of Kant's own conception of practical rationality, his own version of the view that reason can provide guidance both with regard to means and with regard to the ultimate ends or, more precisely, norms of action.

For Kant, action and the *will* are inherently linked to *reason*; the key element in their relationship being what he variously refers to as *laws*, *principles*, or *maxims*.

Everything in nature works in accordance with laws. Only a rational being has the power to act *in accordance with his idea* of laws – that is, in accordance with principles – and only so has he a *will*. Since reason is required in order

to derive actions from laws, the will is nothing but practical reason (1948: 76).

Thus, events in nature, including animal behaviour, happen in accordance with “laws” in the sense of cause-and-effect regularities. This is how involuntary human bodily movements also are experienced by human beings. Those modes of behaviour of human or other agents, however, which we call *actions* are not understood as events that (causally) happen to those agents; they are viewed as being *brought about by the agents themselves* in accordance with the agents’ understanding of a practical law, or principle, that is, as instances of general ideas of what one is doing. It is by virtue of this ability to bring about acts which come under principles that agents are, for Kant, said to have a will.

This subsumption of individual actions under general principles is, for Kant, the work of reason in its practical capacity, just as the subsumption of objects under concepts in sense experience is the work of reason in its theoretical capacity. In this way, the will is identified with practical reason. Thus, all actions – that is, all *willed* behaviours – are, by definition, rational in the minimal sense of being derived from principles.

However, actions can, according to Kant, be rational also in another, stronger sense. To appreciate this, we must first look into Kant’s distinctions between *subjective* principles, or maxims, *objective* principles, or practical laws, and *imperatives*.

A *maxim* is a subjective principle of action and must be distinguished from an *objective principle* – namely a practical law. The former contains a practical rule determined by reason in accordance with the condition of

the subject (his ignorance or again his inclination): it is thus a principle on which the subject *acts*. A law, on the other hand, is an objective principle valid for every rational being; and it is a principle on which he *ought to act* – that is, an imperative (1948: 84. See also, 1948: 66 and 1996: 31-2).

Thus, the principle on which an agent actually acts in a particular situation is that agent's subjective principle, or maxim, for that particular action. Principles of action, on the other hand, which are valid for all rational beings regardless of whether or not they are actually acted upon, are objective principles or practical laws.¹

A purely rational being, a being whose actions are motivated only by reason, would always – and effortlessly, so to speak – act in accordance with objective principles. In any action of such a being, in other words, the operative subjective principle would always be none other than the objective principle applicable to the situation. Human beings, however, are not purely rational in this sense. Their actions are subject, over and above objective principles, to two additional sources of influence, not present in the case of purely rational beings: these sources of influence are – to use the terms employed by Kant in the above-quoted passage – *inclination* and *ignorance*.

Inclinations (Kant refers also to desires, impulses, etc, as having a close affinity to them) are psychological states which have motivating power over human action and which often come into conflict with objective principles by motivating agents towards courses of action that are at variance with the ones specified by them.

Because of the possibility of such a conflict, in the case of non-purely-rational beings such as humans, objective principles take on the status of *commands of reason* or *imperatives*: they acquire the character of statements that specify *duties*, that prescribe what *ought* to be done.

To be sure, imperatives, or commands of reason, are not always obeyed, despite being recognized as valid, and thus as binding for human and other rational beings. In cases of inner conflict between imperatives and inclinations human beings can, and often do, act on maxims which accord with the latter.

What is important for Kant, however, is that commands of reason, or imperatives, do have motivating power. We always have the option to follow the imperative rather than the competing inclination. What we cannot do is exercise choice as to the nature, strength or object of our inclinations. These simply happen to us. As MacIntyre (in his *A Short History of Ethics*) puts it,

Inclination belongs to our determined physical and psychological nature; we cannot in Kant's view choose our inclinations. What we can do is to choose between our inclination and our duty (1966: 192-3).

Human actions can follow maxims that are at variance with objective principles – imperatives or commands of reason – not only because of the force of opposing *inclinations* but also due to *ignorance* of the appropriate imperative; whether this ignorance is known to the agent at the time of the action, or not, and whether it amounts to complete lack of knowledge about the appropriate imperative or is mere uncertainty or ambivalence about it. In any case, Kant is confident that human beings do have ample scope for reducing it. How they can reduce this ignorance – the way in which they can get to know the right imperative of reason

applicable to any given situation – varies with the *type* of imperative the situation calls for.

Objective practical principles, according to Kant, are of two kinds, *conditioned* and *unconditioned*, corresponding to two types of imperatives, *hypothetical* and *categorical*.

All *imperatives* command either *hypothetically* or *categorically*. Hypothetical imperatives declare a possible action to be *practically necessary* as a means to the attainment of something else that one wills (or that one may will). A categorical imperative would be one which represented an action as *objectively necessary* in itself apart from its relation to a further end (1948: 78).

That among reason's practical imperatives there are some which are binding on the will unconditionally, or "categorically", is the central, and by far the most important, claim of Kant's theory of practical rationality. Before looking into this claim more closely however, let us consider briefly what he has to say about hypothetical imperatives, especially his further subdivision of them into two distinct kinds; for apart from its undeniable intrinsic interest, this distinction proves to be also of considerable consequence for Habermas's tri-partite conception of practical rationality, more specifically for his distinction between "pragmatic" and "ethical" discourse.

Kant identifies two different kinds of hypothetical imperatives: "*rules of skill*" and "*counsels of prudence*" (1948: 80). *Rules of skill* specify means to well-defined ends which an agent might seek, or as a matter of fact does seek. These ends can be entirely arbitrary. In so far as hypothetical imperatives are concerned, "there is

absolutely no question about the rationality or goodness of the end, but only about what must be done to attain it" (*ibid.*: 79).

Counsels of prudence, on the other hand, propose various means to a special kind of end, namely "happiness" or "well-being". Happiness differs from other ends in that it is not arbitrary. As an end happiness is "given" – it is actually pursued by every human or other rational being though the content of happiness – the totality of the elements that constitute it – differs from one such being to another. However, what warrants the separate classification of the imperatives that concern means to this end is another characteristic of happiness, one that is more important with regard to the logic of imperatives: this is the inherent indeterminacy of happiness – the impossibility for any human or other finite being to specify at any given moment what his or her happiness might consist in, happiness being a property that applies to one's whole existence, past, present and future.

Unfortunately...happiness is so indeterminate a concept that although every man wants to attain happiness, he can never say definitely and in unison with himself what it really is that he wants and wills. There is required for the idea of happiness an absolute whole, a maximum of well-being in my present and in every future state. Now, it is impossible for the most intelligent, and at the same time most powerful, but nevertheless finite, being to form here a determinate concept of what he really wills (*ibid.*: 81).

Interestingly, Kant assigns the task of forming the idea of happiness for any human being primarily to imagination rather than to reason ("happiness is an ideal, not of reason, but of imagination", *ibid.*: 82). Reason, on the other hand – here in the form of practical wisdom, or "prudence" – is restricted to the role of providing

not strict commands, but only recommendations, or “counsels”, concerning how to attain happiness; hence, “counsels of prudence”.

Despite the considerable differences between rules of skill and counsels of prudence, however, they both, as hypothetical imperatives, have the same logical structure: “Do A if you will B (under circumstance C)”. Each such imperative is premised on a theoretical judgement of the form “Action A (under circumstance C) leads to B”.

Kant raises the question as to how a hypothetical imperative of either kind – even though it might be overpowered by a competing inclination – can be necessary or binding for a rational will; and he finds the answer to it relatively unproblematic: the necessity involved is analytic – one regarding the logical self-consistency of the will.

Who wills the end, wills (so far as reason has decisive influence on his actions) also the means which are indispensably necessary and in his power. So far as willing is concerned, this proposition is analytic (1948: 80-1).

Thus, all hypothetical imperatives presuppose the higher level (analytic) principle “To will the end is to will the means”, and derive their legitimacy and binding force for the rational will from the combination of that principle with the prior adoption of an end by the will and the relevant theoretical proposition(s) linking that end with the means prescribed by the hypothetical imperative in question.

So far Kant has in essence remained within the Humean-instrumentalist conception of practical rationality, despite having formulated his views primarily in the logical language of imperatives of reason and bindingness on a rational will, rather than in Hume’s predominantly psychological language of passions,

emotions, etc. The decisive break with instrumentalism – and also with the naturalism of the objectivist-substantive conception – comes with the idea that there are also “categorical imperatives” of reason.

In contrast to hypothetical imperatives, which are conditional upon the prior willing of an end, and thus have the form “Do (or do not do) A, if you will B”, categorical imperatives are not conditioned by anything beyond the actions prescribed by them. Their form is simply “Do (or do not do) A”. They specify actions that ought, or ought not to be performed – duties that ought to be fulfilled – for their own sake, not for the sake of something else. Categorical imperatives are the imperatives of morality. They specify what is morally impermissible to do or impermissible not to do (i.e., obligatory). Categorical imperatives (with small “c” and small “i”) can be rationally derived, Kant maintains, from a single principle, the Categorical Imperative (with capital “C” and “I”), which is itself in no need of any further rational justification. Kant famously offered several formulations of the Categorical Imperative. The most representative of them – for it best captures their common core idea, that of *Universalizability* – is the first formulation:

Act only on that maxim through which you can at the same time will that it should become a universal law (1948: 84).

The Categorical Imperative can be seen as a procedure for testing maxims (and the actions that follow from them); as a criterion for distinguishing morally impermissible from morally permissible actions. As such, it is an entirely *formal* principle. It is independent of the substance of any possible maxims that might be tested by it, of any interests, needs, desires, etc, that might be expressed in the maxim or with reference to the consequences that would result from applying the maxim. The Categorical Imperative cannot itself produce any maxims of action, it

can only test maxims that are already formulated and which do involve substantive issues. Körner rightly compares the Categorical Imperative with the laws of logic:

Just as the formal principles of syllogistic reasoning divide all syllogisms clearly into two classes, the valid and the invalid, so, Kant believes, the formal principle of morality divides all maxims, and consequently all actions based on them, into those which are moral and those which are not (1955: 135).

According to Kant's above-quoted criterion of universalisability, an action is morally permissible only when it is based on a maxim which a human, or other rational being "can will" that it becomes a universal law, i.e., that it is applicable to him or herself as well as to all other rational beings at all times; and it is morally impermissible if one "cannot will" its maxim to become universal law in this sense. Kant clarifies the key phrase "can/cannot will" as follows:

We must *be able to will* that a maxim of our action should become a universal law – this is the general canon for all moral judgement of action. Some actions are so constituted that their maxim cannot even be *conceived* as a universal law of nature without contradiction, let alone be *willed* as what *ought* to become one. In the case of others we do not find this inner impossibility, but it is still impossible to *will* that their maxim should be raised to the universality of a law of nature, because such a will would contradict itself (1948: 86-7).

Thus, a maxim fails to pass the test of universalizability – that is, cannot be willed by a rational being to become a universal law – in one of two circumstances. Firstly it fails the test when becoming a law is conceptually self-contradictory, therefore impossible to realize. Kant illustrates this point by reference to the practice of promising. A maxim such as “You may break a promise if it is in your interest to do so” is non-universalizable in this sense. For if everyone were entitled to break their promises whenever it were in their interest to do so, promising would become pointless and cease to be practised; consequently, breaking promises would become impossible, not only under the circumstances the maxim presumably allows, but under any circumstances whatsoever.

Secondly, a maxim fails the test when its becoming universal law, though not conceptually self-contradictory and thus impossible to realize, as in the previous case, is nonetheless a state of affairs which a rational being cannot will without self-contradiction. A maxim such as “Do not help others promote their well-being” is, for Kant, non-universalizable, in this second sense. For no rational being would want to give up all prospect of ever receiving help from others.

In the *Critique of Pure Reason*, Kant argues that certain theoretical categories, such as the category of causality (that every event has a cause), necessarily underlie our experience of events in the world. In a similar way, in his moral writings Kant tries to demonstrate that the Categorical Imperative is implicit in the everyday moral consciousness of humankind.

Despite the similarity, however, there is a fundamental difference between the two cases: whereas the “transcendental” argument linking the theoretical categories with our ordinary experience of events in the world is believed by Kant to establish the validity of those categories, in the practical sphere Kant insists that the

analogous argument leading from ordinary moral judgements to the Categorical Imperative does not amount to a justification of the latter.

For with regard to the former case, ordinary experience is “given” and must be taken for granted – it cannot be chosen or altered; thus, the theoretical categories are valid as long as they can “explain” ordinary experience as it presents itself. In the practical sphere, Kant recognizes that ordinary moral consciousness is often misled by inclinations to make judgements that are at odds with the prescriptions of the Categorical Imperative; but this is not why the Categorical Imperative cannot be justified in terms of ordinary moral consciousness. After all – Kant maintains – a systematic philosophical clarification of the Categorical Imperative in its different versions – such as the one he himself undertakes – would significantly enhance the ability of our moral consciousness to match its judgements with the Categorical Imperative, achieving a an appreciable measure of harmony between the two.

The point however is different. The Categorical Imperative – unlike the theoretical categories in their relation to ordinary experience – is supposed not to *explain* but to *establish* ordinary moral judgements (the words in italics are Kant’s own). It is supposed to determine what such judgements ought to be, to provide, in other words, justification for them. It cannot, therefore, at the same time derive its own validity from them, no matter how well the two match each other. If the whole edifice of morality is not to remain groundless, relying on a logic which circularly tries to justify morality’s two poles – ordinary moral judgements and the Categorical Imperative – in terms of each other, if, in Kant’s own vivid metaphors, morality is to avoid being “merely a chimerical Idea without truth” or “a mere phantom of the brain” (1948: 106), the Categorical Imperative must derive its validity independently of ordinary moral judgements, only thus becoming competent to confer validity upon them.

Kant's arguments aiming to establish the validity of the Categorical Imperative are extremely complex, and admit of alternative interpretations, especially if one considers both of the major works in which they are put forward, the *Groundwork* and the *Critique of Practical Reason*. A thorough, sympathetic but not uncritical, discussion of them is presented in H.J. Paton's *The Categorical Imperative* (1948). The essentials of Kant's argument, along the lines of Paton's interpretation, can be put as follows (at the risk of oversimplification):

The Categorical Imperative is a principle at once *practical*, *a priori*, and *synthetic*. It is a *practical* principle in the sense that it is not simply cognitively binding, but is binding on the will to initiate action of a certain kind. It is *a priori* in that it is not conditioned by anything empirical, indeed by anything outside of itself, such as human needs, inclinations, preferences or ends, or even the will of God. Finally, it is *synthetic*, as opposed to analytic, in that its negation is not self-contradictory, or, its truth does not result from the meaning of its terms.

The Categorical Imperative, therefore, must be rationally justifiable as a principle that is binding on the will, and this must be so not by reference to anything external to it and not by mere clarification of the meaning of its terms. Kant maintains that the Categorical Imperative is indeed justifiable in this way. His point is that practical reason, by reflecting on itself, recognizes the Categorical Imperative as its own principle of operation; we might say, as a kind of principle of "practical self-consistency" or "practical non-contradiction" in the sense of self-consistency and non-contradiction with regard to actions willed rather than to propositions held. Put another way, The Categorical Imperative is recognized by the rational will – which for Kant is identical with practical reason – as its own law. The will, therefore, is justified in abiding by the Categorical Imperative, for in doing so it is obeying a self-imposed law; it is, in other words, exercising autonomy.

APPENDIX III

MAX WEBER AND THE DISCONTENTS OF MODERN RATIONALIZATION

Weber considers certain forms of "rationalization", that is, the application of certain forms of rationality, to be the most pervasive feature of modern civilization.¹ The term "rationality", and its several cognates, are used by Weber in a great and often bewildering variety of ways (see Brubaker, 1984: 2). It is, however, possible to extract from his writings a sufficiently systematic and consistent set of concepts that will enable us to grasp what he has to say about processes of rationalization in various domains of modern life, and how he links rationality with what he diagnoses as the main negative consequences of these rationalization processes – namely, the loss of meaning and the loss of freedom.

Types of Action and Rationality

In his well known four-fold typology of action Weber distinguishes two forms of non-rational and two forms of rational action. The non-rational action types are *affectual* action, which is action "determined by the actor's specific affects and feeling states", (Weber, 1968: 25) and *traditional* action, that is action "determined by ingrained habituation" (*ibid.*: 25). Both affectual and traditional action lack the property that, for Weber, constitutes the defining characteristic of rational action: the deliberate control of action by conscious ideas (see Swindler, 1973), often expressed, alternatively, as control of action "by the intellect". The two types of rational action identified by Weber are "value-rational" (*wertrational*) and purposively rational (*zweckrational*, often translated rather misleadingly as instrumentally rational) action. They are differentiated from each other according to the way in which they are controlled by conscious ideas.

Value-Rational and Purposively Rational Action

Value-rational action is defined as action which is

determined by a conscious belief in the value for its own sake of some ethical, aesthetic, religious or other form of behaviour, independently of its prospects of success (Weber, 1968: 25).

Action of this type is characterized by "its clearly self-conscious formulation of the ultimate values governing the action", the latter taking the form of "unconditional demands" on "persons who, regardless of possible cost to themselves, act to put into practice their convictions", (p. 25) examples of which are those that "seem to them to be required by duty, honour, the pursuit of beauty, a religious call, personal loyalty, or the importance of some cause" (ibid.: 25).

Thus, a particular value-rational action does not aim at a "result ulterior to it", but has intrinsic value itself, as it constitutes an example – an instance, a particular realization – of the general idea of the ultimate value by which it is governed. Furthermore, the intrinsic value of any particular value-rational action is, at least implicitly, considered as overriding with respect to the action's consequences; and it is so considered *a priori*, that is, independently of any calculation of such consequences

Purposively rational action, on the other hand, is defined as action which is

determined by expectations as to the behaviour of objects in the environment and of other human beings; these expectations are used as "conditions" or "means" for the attainment of the actor's own rationally pursued and calculated ends (ibid.: 24).

In these calculations

the end, the means, and the secondary results are all rationally taken into account and weighed. This involves rational consideration of alternative means to the end, of the relations of the end to the secondary consequences, and finally of the relative importance of different possible ends (*ibid.*: 26).

It is widely agreed among commentators – and Weber's own remarks on the matter seem to justify this view – that for Weber purposively rational action is more rational than value-rational action. Indeed, according to the above definition purposively rational action is more comprehensively controlled by conscious ideas – absolute as well as comparative ideas regarding ends, means, and consequences – than value-rational action, which is controlled only by ideas concerning ultimate values. Nevertheless, this interpretation needs to be qualified. As already mentioned, it can be said that in cases of value-rational action, alternative ends, alternative means and the relevant consequences are also taken into consideration, in a particular sense: they are *a priori* judged to be collectively overridden – even if they are not known – by the particular realization of the ultimate value which governs the value-rational action in question. The degree of rationality possessed by value-rational action seems to depend on how consciously this judgement is made in any particular case.

An important question arising with regard to Weber's conception of purposively rational action concerns the rationality of the choice between alternative and conflicting ends. Weber identifies two ways in which such a choice can be made.

Firstly,

the actor may... take them as given subjective wants and arrange them in a scale of consciously assessed relative urgency. He may then orient his action to this scale in such a way that they are satisfied as far as possible in order of urgency, as formulated in the principle of "marginal utility" (*ibid.*: 26).

Here, choices regarding ends are rational to the extent that such choices are governed by conscious assessments of relative urgency, but no further, for these assessments are themselves non-rational, expressing "subjective wants" that are taken as "given" rather than being brought under the conscious control of any further ideas.

The second way in which choices regarding ends might be made, according to Weber, is "in terms of a rational orientation to a system of values" (*ibid.*: 26). Thus, in this case – unlike the previous one – the ordering of priorities is itself carried out in a rational way, by being brought under the conscious control of ideas regarding values. It is worth noting, however, that Weber describes such choices as being "determined in a value-rational manner" (*ibid.*: 26), which, according to his own definition of "value-rational", means that choices are determined by treating the relevant ends as possessing intrinsic value, that is as relating in a non-mediated way to some absolute and overriding values (or unconditional commands). This way of understanding the "rational orientation to a system of values" raises, of course, the question how a choice between conflicting ends is possible if more than one of them is related to values in such a value-rational manner.

There is, however, a more important point that needs to be made with regard to this interpretation of the relation between ends and values. By construing the relation in terms of value-rationality, Weber seems to disregard another likely way in which – in terms of his own conceptual schema – ends might be brought under the conscious control of ultimate value; namely, by means of purposive rationality. According to this interpretation, ends would be treated as means to other ("ulterior", i.e. external) ends, which might themselves be means to other ends, and so on, until the chains of means and ends reach some ultimate value, or values, where they terminate.

This apparent capacity of (Weberian) purposive rationality to follow means–ends chains right through to ultimate values is of considerable significance, and will be revisited in later sections. For now, we shall turn to Weber's views on ultimate values themselves and on the capacity of rationality with regard to their adoption.

Rationality and Ultimate Values

Weberian rational action, whether of the value-rational or the purposively rational type, evidently does not involve the rational choice of ultimate rational values, or any other rational value judgement. The rationality of action lies only in the ways in which the action is related to ideas, often including ideas of ultimate values. It involves (a) understanding – and making understandable, i.e. clarifying – ideas, such as concepts of values, (b) making logical inferences, and (c) making judgements of a factual nature, such as those concerning relations of cause and effect; but it does not involve making value judgements.

However, Weber goes beyond merely saying that for an action to be rational it is not necessary that the values involved in it are chosen rationally. Under the strong influence of Nietzsche, and in a manner reminiscent also of Hume, he rejects the

very possibility that value judgements can, under any circumstances, be made in a rational, or "scientific", way, affirming the inescapability of "irrational" ethical and other value choices and commitments, and the fundamentally irrational springs of human motivation.

For Weber, as Giddens puts it, "[s]tatements of fact, and judgements of value are separated by an absolute logical gulf" (Giddens, 1972: 42) In his renowned public lecture "Science as a Vocation", he points out that

it is one thing to state facts, to determine mathematical or logical relations or the internal structure of cultural values, while it is another thing to answer questions of the *value* of culture and its individual contents and the question of how one should act in the cultural community and its political associations (Gerth and Wright Mills [henceforth G & M], 1970: 146).

and speaks of 'the impossibility of "scientifically" pleading for practical and interested stands – except in discussing the means for a firmly given and presupposed end' (G & M, 1970: 147). To quote Giddens, once more, 'in Weber's thought, the notion of "value" becomes synonymous with (irrational) conviction' (Giddens, 1972: 57).

Closely associated with these views is Weber's firm belief in the existence of a plurality of fundamental values, value spheres, or attitudes to life, which are, and will always be, in constant and unresolvable conflict – a "polytheism" of unceasingly struggling "gods and demons"; and his consequent belief in the need to choose between them non-rationally.

[T]he ultimately possible attitudes toward life are irreconcilable, and hence their struggle can never be brought to a final conclusion. Thus it is necessary to make a decisive choice (G & M, 1970: 152).

Theoretical Rationality

In addition to the forms of rationality that pertain to action – which might be referred to as forms of "practical" rationality – we also find in Weber's work the notion of a "theoretical" or "intellectual" rationality, a rationality that is appropriate to "the intellectual comprehension of reality" (G & M, 1970: 293). It involves

the kind of rationalization the systematic thinker performs on the image of the world: an increasing theoretical mastery of reality by means of increasingly precise and abstract concepts (G & M, 1970: 293).

It can be said that theoretical rationality is characterized by the deliberate, methodical, logical linking of beliefs about reality to other consciously held relevant ideas and beliefs, in the same way as practical rationality is characterized by the establishment of such a relationship between action and ideas. Thus, theoretical rationality tends to create consistent and coherent networks, or systems, of concepts and propositions about different domains of reality and experience including not only aspects of the natural world but also areas of social and cultural life. Weber, indeed, discusses processes of theoretical rationalization, or intellectualization, with regard to such domains as the natural and social sciences, religions, ethics, law, and music. From a motivational point of view, whereas practical rationality is based on the desire to realize certain valued ends,

theoretical rationality, as Kalberg points out, is anchored in the "cognitive need of intellectuals" (Kalberg, 1994: 135).

Apart from its role in the "intellectual comprehension of reality", theoretical rationality can also serve practical purposes, by being incorporated in processes of practical rationality. A major example is the integration of science and technology. "Natural science gives us an answer to the question of what we must do if we wish to master life technically" (G & M, 1970: 144). And more generally, "science contributes to the technology of controlling life by calculating external objects as well as man's activities" (G & M, 1970: 150). Scientific knowledge of the natural world can contribute to the rationalization of technical action by rendering technical means more calculable, predictable and controllable, while social scientific knowledge can have an analogous function in the field of economic activity as well as in other areas of social action.

A close union of theoretical and practical forms of rationality can also exist in legal and administrative systems (whose importance for Weber's work cannot be overestimated). For such systems, when rationalized, constitute internally consistent and coherent networks of conceptually clear and precise propositions, and at the same time – these propositions being rules of conduct – they are also devices for the regulation of action in a calculable, predictable, controllable manner. As we shall see, this form of regulation of action lies at the heart of his theory of bureaucracy.

Formal and Substantive Rationality

Another conceptual distinction that figures prominently in Weber's studies of rationalization processes in the West is that between *formal* and *substantive* rationality. In a nutshell, an activity is formally rational to the extent that it involves

calculability, and substantively rational to the extent that it is oriented to the attainment of ultimate values.

Two of the main areas in relation to which Weber draws this conceptual distinction are (a) economic action and (b) law.

(a) On the one hand, '[t]he term "formal rationality of economic action" will be used to designate the extent of quantitative calculation or accounting which is technically possible and which is actually applied' (Weber, 1968: 85). And, in somewhat more concrete terms,

A system of economic activity will be called "formally" rational according to the degree in which the provision for needs, which is essential to every rational economy, is capable of being expressed in numerical, calculable terms, and is so expressed (ibid.).

On the other hand,

[t]he "substantive rationality"... is the degree to which the provisioning of given groups of persons...with goods is shaped by economically oriented social action under some criterion... of ultimate values...(ibid.).

Such ultimate values might be "ethical, political, utilitarian, hedonistic, feudal... egalitarian..."(ibid.).

(b) In the area of law, formal rationality is understood as the kind of calculability of decisions and actions that is present when "only unambiguous general characteristics of the facts of the case are taken into account" (ibid.: 656-7), which is possible to the extent that we have "the collection and rationalization by logical

means of all the several rules recognized as legally valid into an internally consistent complex of abstract legal propositions" (*ibid.*: 657). On the other hand, legal activity is substantively rational to the degree that decisions and actions are influenced by fundamental norms such as "ethical imperatives, utilitarian and other experimental rules, and political maxims" (*ibid.*).

It must be emphasized that in formally rational action, calculability goes hand in hand with predictability and controllability, thus with efficacy; what is calculable is not just any imagined outcomes (including ends, means and possible secondary effects) of action, but its real outcomes – the outcomes that actually obtain, or would obtain if the action were undertaken. What formal rationality is not concerned with is whether what is calculable, predictable, controllable, is also desired or desirable. These points are made abundantly clear in what Weber says about "technical rationality", a term which can be understood as more or less equivalent to "formal rationality" but reserved by Weber for use in connection with certain domains of activity. Thus,

[f]or purposes of the definition of technical rationality, it is wholly indifferent whether the product of a technical process is in any sense useful... we can conceive of a rational technique for achieving ends which no one desires. It would, for instance, be possible, as a kind of technical amusement, to apply all the most modern methods to the production of atmospheric air. And no one could take the slightest exception to the purely technical rationality of the action (*ibid.*: 67).

In terms of the above definitions, there is an evident affinity, on the one hand between formal and purposive rationality (as well as a close association between

both and theoretical rationality), and on the other hand between substantive and value-rationality.

Let us, however, look a little more closely into the conceptual relationships between these forms of rationality.

Relationships between Types of Rationality

Theoretical and Formal Rationality

The calculability of the results of an action, which would be the hall-mark of the action's formal rationality, evidently presupposes the existence of a relevant field of theoretical rationality, that is, the existence of a consistent and coherent system of concepts and propositions about the domain of reality within which the action in question is to take place.

However, the mere presence of theoretical rationality does not guarantee its practical use, i.e. its utilization for rendering relevant actions formally rational. It can be said, therefore, that theoretical rationality is a necessary but not sufficient condition for formal (practical) rationality.

Formal and Purposive Rationality

An analogous relationship can be discerned between formal and purposive rationality. The calculability of the results of following a certain course of action – i.e., that action's formal rationality – is a prerequisite for making means–ends calculations with regard to that action, i.e. for the action's purposive rationality. But for the exercise of such purposive rationality, there must also be – in addition to the possibility of the requisite formal rationality – an intended end which is among the calculable results of the formally rational action. Let it be made clear, though,

that the degree of formal rationality – i.e. calculability – necessary for the rationality of any purposive action, is never greater than that required by the degree of precision with which the action's ends are specified. The same, of course, applies in the case of theoretical rationality.

Theoretical, Formal and Value-Rationality

The value-rationality of a particular action consists in the logical subsumption of that action under a concept of an ultimate value, that is, its linking with the ultimate value in question without the mediation of causal (i.e. means–ends) relationships. The action would be formally rational if the logical relationship between the particular action and the relevant concept were calculable irrespective of whether the agent held the concept as an (ultimate) value. Similarly, the clarity and coherence of the concepts involved would signify the presence of theoretical rationality, regardless of whether those concepts were used for any practical purposes.

Thus, (a certain kind of) theoretical rationality is a necessary but not sufficient condition for (a certain kind of) formal rationality, and the latter is a necessary but not sufficient condition for value-rationality.

Substantive, Value- and Purposive Rationality

On the basis of Weber's definitions, value-rationality always (by conceptual necessity) amounts to substantive rationality, as it involves the linking of action to some ultimate value, as substantive rationality requires. The latter, however, need not always be of the value-rational variety, that is, involving the linking of action to an ultimate value not in a causally-mediated, but in a logically subsumptive manner.

Substantive rationality may, conceptually speaking, be of a purposive type, in the sense that it involves relating an action to an ultimate value via a chain of means and ends. In that case, of course, substantive rationality would also have to be formal in ways and to the extent necessitated by the respective purposive rationality.

One of the most crucial questions arising out of Weber's work concerns the degree to which – in the above terms – substantive rationality of the purposive, but also of the value-rational form can be practically realized in modern society. In other words, when the costs of formalization are taken into consideration, is it feasible for modern society to attain the formal rationality necessary for achieving as fully developed a substantive rationality as socially needed; or is it necessary to compromise, either

(a) by opting for a sufficient formalization of purposive action with respect to specific ends but without sufficiently linking those ends to ultimate values, ending up with what can be called a sufficiently formalized but substantively truncated purposive rationality; or,

(b) by pursuing a less than adequately formalized, less than fully developed, substantive rationality, where an effort is made to optimize the relationship between formal and substantive rationality in favour of maximizing the realization of ultimate values?

With the above conceptual analysis in mind, let us now turn to Weber's diagnosis of modernity in terms of the important idea of rationalization.

Western Rationalization

The idea that widespread and growing rationalization is the dominant characteristic of western or occidental civilization – and the main feature distinguishing this civilization from all others in history – is justifiably considered to be one of the central tenets defining Weber's sociological understanding of modernity.

Three preliminary qualifications, however, must be made in order to protect this immensely important Weberian idea from likely misinterpretations.

Firstly, western rationalization does not consist in the undifferentiated spread of rationality in all its forms – as these have been identified by Weber – but in the dominance of some of its forms at the expense of others.

The precise character of the forms of rationality that, according to Weber, prevail in modernity will be discussed later. For the moment, suffice it to say that they are akin to the purposive, formal and theoretical forms, as opposed to the value- and substantive forms, which are becoming less and less influential.

Secondly, Weber is not postulating a determination or teleology inherent in rationality. Rationalization (in the specific forms in which it has occurred) has not been historically inevitable, but depended on the existence of particular sets of conditions, which happened to arise at certain times and not at others, and in certain regions of the world and not in others. Such conditions include economic interests, power relations, religious and cultural movements, etc. Indeed, much of Weber's work is concerned with identifying, and explaining the emergence of, conditions that were conducive to the development of modern civilization, and ultimately with explaining why the phenomena of modernity appeared in the West

and not in the context of other civilizations such as those of China or India. Most elaborate, in this respect, is his thesis developed in *The Protestant Ethic and the Spirit of Capitalism* (Weber, 1958), that the emergence of capitalism, and consequently of modernity, in the West was significantly facilitated by certain religious beliefs, and associated attitudes and practices with regard to work and consumption, that were adopted by certain protestant sects. While rejecting the historical inevitability of rationalization, however, Weber insists that once the forms of rationality characteristic of modernity do take root, they become – due to what Weber calls their technical superiority – virtually irresistible. Their application therefore acquires a strong tendency towards irreversibility as well as towards intensification and further expansion.

Thirdly, modern rationalization is not understood by Weber as a blanket phenomenon covering uniformly and simultaneously all aspects of social life; but rather as a set of processes which take place in different domains of social activity (economic, political, scientific, legal, artistic, etc.), and which – while exhibiting strong family resemblances – might differ in several respects, such as the specific mix of rationality types involved, the intensity of application of each rationality type, the stage of development of the rationalization process, the rhythm of its development, etc. The specific character of a process of rationalization is shaped by the particular set of conditions in which it occurs.

Keeping these qualifications in mind, we shall now turn to Weber's account of rationalization and its consequences in three domains of modern social activity: (1) science, (2) the capitalist market, and (3) organizations.

Science

Modern science, according to Weber, is the result of a process of intellectualization – that is, of the growth of what we have earlier referred to as theoretical rationality. This process has gone on in the West for millennia but has sharply intensified in modern times.

For Weber, science, in addition to providing a "theoretical mastery of reality", is practically valuable mainly in three ways (G & M, 1970: 150-2): Firstly, it is a powerful technical instrument, making possible the technological control of natural as well as socio-cultural objects and thus the adoption of effective means to chosen ends. Secondly, science serves as a good exercise for the mind; it provides "methods of thinking, the tools and the training for thought". Thirdly, it enables us to attain self-clarification with respect to our evaluative standpoints.

What science cannot do, however, – and what it has, by and large, given up all ambition of doing – is to provide justification for evaluative standpoints of any sort. In this context, Weber approvingly quotes Tolstoi's claim that science 'gives no answer to our question, the only question important for us: "What shall we do and how shall we live?" ' (G & M, 1970: 143).

More specifically, the very value of science's own theoretical or practical attainments – the very value of science itself – cannot be scientifically proved; it must simply be presupposed:

Science... presupposes that what is yielded by scientific work is important in the sense that it is "worth being known"... this presupposition cannot be proved by scientific means. It can only be *interpreted* with reference to... our ultimate position towards life (*ibid.*).

Similarly science

leaves quite aside, or assumes for its purposes, whether we should and do wish to master life technically and whether it ultimately makes sense to do so (*ibid.*: 144).

Furthermore, science cannot discover any meaning in the objects it studies, nor can it demonstrate that those objects have any value. For science, as opposed to religious or other modes of comprehending reality,

there are no mysterious incalculable forces that come into play... one can, in principle, master all things by calculation. This means that the world is disenchanted (*ibid.*: 139).

This disenchanted world – which includes the artificial, man-made world of technology just as much as nature and society – is one about which it cannot be scientifically proved that it is 'worth while, that it has any "meaning", or that it makes sense to live in...' (*ibid.*: 144)

Disenchantment and meaninglessness are, thus, necessary concomitants of modern science's rational mode of relating to the world. And since this mode has displaced the earlier, religious one from its position of dominance, disenchantment and meaninglessness have become pervasive characteristics of our experience – the price our civilization has to pay for the theoretical and technological mastery over reality.

Thus, for Weber, science's unsurpassed theoretical rationality, and – through it – its great and indispensable contribution to the formal and purposive rationality of technological action, go hand in hand with definite substantive irrationalities,

associated, above all, with the undermining of such a paramount ultimate value as the meaningfulness of life and the world.

Market-Oriented Economic Activity

Capitalist economic activity, oriented to a competitive market is, for Weber, another major field of increasing rationalization.

Weber refers to "consociation through exchange in the *market*" as "the archetype of all rational social action" (Weber, 1968: 635). Market-oriented action is (formally) rational in that it is based on the strict, quantitative calculation of all the factors involved in the action, that is the factors relevant to the pursuit of the agent's interests (*ibid.*: 85-7).

For a capitalist economic unit, such relevant factors that need to be subjected to calculation might include the prices at which raw materials can be purchased and products sold, as well as technological and social processes of production, administration and management.

The highest degree of quantitative calculability is achieved, in Weber's view, when every relevant factor can be accounted for in monetary terms:

From a purely technical point of view, money is the most "perfect" means of economic calculation. That is, it is formally the most rational means of orienting economic activity (*ibid.*: 86).

In modern market-oriented economic activity, the calculability of all the relevant factors in terms of money goes hand in hand with the market's orientation to these factors solely as commodities. What is common to money and commodities is their "abstract" nature : money is an abstraction from all particular qualitative and

quantitative criteria for evaluating things, whereas the commodification of things constitutes the exclusive focussing on their market-exchangeability, in abstraction from all their specific qualities; on their "exchange-value" as opposed to their "use-value", as Marx would have put it.

The calculability of commodities in terms of money, which characterizes modern economic action and makes it formally rational, facilitates the purposively rational pursuit of specific market-oriented goals on the part of economic agents. Indeed, any failure by an economic agent to be rational in this way can be very costly, incurring even the ultimate penalty of economic extinction (*ibid.*: 1186). The substantive rationality of such action, however, is another matter altogether. To begin with, monetary calculability is not inherently linked, positively or negatively, to any form of social distribution of goods, or to the satisfaction of real wants, or to any such socio-economic values, though it may coincide as a matter of fact with the satisfaction of some such value. Thus,

Formal and substantive rationality, no matter by what standard the latter is measured, are always in principle separate things, no matter that in many (and under certain very artificial assumptions even in all) cases they may coincide empirically... [I]f the standard used is that of the provision of a certain minimum of subsistence for the maximum size of population, the experience of the last few decades would seem to show that formal and substantive rationality coincide to a relatively high degree (1968: 108-9).

However, this (alleged) independence of substantive from formal rationality seems to exist, for Weber, only if substantive rationality is defined in a very restrictive way, merely in terms of socio-economic values. For Weber considers the

monetary calculability of market-oriented economic action – thus, its formal rationality – to be inherently at odds with a whole range of other substantive values:

Money is the most abstract and "impersonal" element that exists in human life. The more the world of the modern capitalist economy follows its own immanent laws, the less accessible it is to any imaginable relationship with a religious ethic of brotherliness (G & M, 1970: 331).

And,

The market community as such is the most impersonal relationship of practical life into which humans can enter with one another... The reason for the impersonality of the market is its matter-of-factness, its orientation to the commodity and only to that. Where the market is allowed to follow its own autonomous tendencies, its participants do not look toward the persons of each other but only toward the commodity; there are no obligations of brotherliness or reverence, and none of those spontaneous human relations that are sustained by personal unions... Market behaviour is influenced by rational, purposeful pursuit of interests... Such absolute depersonalization is contrary to all the elementary forms of human relationship... The "free" market is an abomination to every system of fraternal ethics (Weber, 1968: 636-7).

In these, as in many other similar passages, Weber, with considerable eloquence and hardly restrained passion, depicts what he sees as the heavy, albeit

necessary costs of the process of rationalization in the field of modern, market-oriented economic action; costs that can be summarized by the phrase "depersonalization of human relations".

Thus, the kind of rationalization that, according to Weber, is taking place in the realm of market-oriented activity can be described as (a) highly formal, (b) purposive in a restricted way – namely, only with regard to a narrow range of ends, and neglectful of, and often in direct conflict with ultimate values – and (c) insufficiently substantive, not only in a purposive way but also in value-rational one.

Bureaucratic Organizations

The third field of occidental rationalization to be discussed – the most important one for Weber – is that of modern organizational life.

Weber analyzes organizational rationality in terms of the crucial concept (the "ideal type") of "bureaucracy", maintaining that modern organizations are becoming increasingly bureaucratic, and that bureaucratic organizations are coming to dominate more and more areas of modern social life.

Among modern organizations that, according to Weber, are rationalized through bureaucratization are such varied institutions as economic enterprises of all kinds, public administration organizations, political parties, voluntary associations, armies, hospitals, trade unions, churches, schools, universities and scientific research centres.

Bureaucracy, for Weber, is "*the* means of transforming social action into rationally organized action" (1968: 987). It constitutes a form of organization where action is regulated by a legally legitimated system of general, impersonal, precise, calculable rules (*ibid.*: 220-1, 956-8).

Using the term "rules" in its most abstract sense to include laws as well as all types of regulations (state and private, administrative and technical), we can say that bureaucratic rules first of all define – in an impersonal way – all the role-positions in an organization. These role-positions – which in the most advanced, "monocratic" (*ibid.*: 223), form of bureaucracy are occupied by individuals, as opposed to collectivities – are arranged in a pyramid-like, multi-layered hierarchical structure. A bureaucratic organizational structure has two distinct but closely associated aspects, which can even be thought of as two different but superimposed pyramids, which might be referred to as the "task-pyramid" and the "authority-pyramid".

The "task-pyramid" concerns the allocation of technical tasks that need to be carried out at every role-position. This allocation is based on the hierarchical division and subdivision of the total work of the organization into smaller and smaller parts down to quite narrow, specialized task-areas. These narrowest, most specialized task-areas are assigned to the role-positions at the bottom of the pyramid, whereas to hierarchically higher positions are assigned correspondingly more encompassing task-areas, with the (technical) responsibility for determining or clarifying, offering technical advice on, supervising, coordinating and fitting together the work of those occupying the positions hierarchically below them. Thus, the individual holding the top-position in the pyramid has this kind of responsibility with regard to the entire work of the organization.

The "authority pyramid" is a hierarchical structure of command and control. It concerns the allocation of rights to issue instructions of certain kinds and duties to obey instructions of certain kinds to "authority positions" which coincide with the "task-positions" that belong to the "task-pyramid", as explained above. The higher we go in the pyramid – thus, the more encompassing a position's task-area – the

greater the authority belonging to the position, with those at the bottom of the pyramid having only duties to obey instructions coming from above, and the one at the top having only rights to issue instructions to those below (being, though, accountable to bodies outside the bureaucratic pyramid, such as government ministers, shareholders, etc.).

In addition to the rules defining this double-aspect bureaucratic organizational structure, there are, for every role-position in the structure (in both its task and authority aspects), further administrative and technical regulations setting out precise, calculable methods and procedures for carrying out the normal tasks of the position and for tackling special situations that might arise within its area of jurisdiction.

Members of such bureaucratic organizations – that is, individuals who occupy the various role-positions in them – are appointed or promoted to their respective position on the grounds that they possess the specialized knowledge necessary for successfully handling the task and authority responsibilities of that position. Members must have the legal, administrative and technical expertise necessary for interpreting and following the whole body of rules applying to their specific position, as well as for working out within the framework of the relevant rules effective ways of fulfilling the tasks belonging to that position. So central does Weber consider the place of expert knowledge in bureaucracy that he goes as far as to maintain that "[b]ureaucratic administration means fundamentally domination through knowledge" (1968: 225).

Organizational rules, for Weber, are obeyed because they are considered legitimate. Without such an attribution of legitimacy to its rules, an organization would simply not function. Now, it is one of the central properties of Weberian bureaucracy that it possesses a particular form of legitimacy, namely "legality"; or,

alternatively, Weberian bureaucratic organizations function on the basis of a particular form of authority (*herrschaft*, translated also as "domination", or "rulership"), namely, "legal", or "legal-rational" authority. In Weber's words, "the purest type of legal authority is that which employs a bureaucratic administrative staff" (*ibid.*: 220).

Weber distinguishes legal authority from two other pure types of authority, "traditional" and "charismatic". "Traditional" is the type of authority that "[rests] on an established belief in the sanctity of immemorial traditions and the legitimacy of those exercising authority under them" (1968: 215), whereas "charismatic" is authority "resting on devotion to the exceptional sanctity, heroism or exemplary character of an individual person, and on the normative patterns or order revealed or ordained by him" (*ibid.*: 225). Legal authority on the other hand is based neither on traditional nor on charismatic, but on "rational grounds – resting on a belief in the legality of enacted rules and the right of those elevated to authority under such rules to issue commands" (*ibid.*: 215).

Thus, in bureaucratic organizations obedience is owed to rules in so far as they have been *enacted* in a legally proper manner, and to individuals' commands in so far as the individuals concerned derive authority to issue commands of that kind from legally enacted rules. The legal enactment of its system of rules is, therefore, the ultimate foundation of every bureaucratic organizational order. It must be added, however, that Weber leaves wide open both the formal requirements for legal enactments and their substantive grounds. Thus, "any given legal norm may be established by agreement or by imposition, on grounds of expediency or value-rationally or both" (1968: 217). Therefore, a bureaucratic order may be established and kept in motion by an individual entrepreneur, a group of shareholders, a charismatic religious leader, a dictator, a democratic political leader, or a

parliament; and it may be created in order to promote private economic interests, national interests, spiritual or social values, etc.

Whatever the mode and ultimate goals of its enactment, a system of bureaucratic rules regulates the conduct of those subject to its authority in a strict manner, rendering their decisions and actions – thus the results of a bureaucratic organization's operations – calculable, predictable and controllable.

Weber firmly believes that bureaucracy has unsurpassable technical advantages over other forms of organization, which, to his mind, explains what he sees as "the irresistibly expanding bureaucratization of all public and private relations of authority" (*ibid.*: 1002) and justifies such sweeping and daring statements on his part as "[t]he future belongs to bureaucratization" (*ibid.*: 1401).

Telling, in respect of this positive aspect of Weber's estimation of bureaucracy, is the following passage:

The decisive reason for the advance of bureaucratic organization has always been its purely *technical* superiority over any other form of organization. The fully developed bureaucratic apparatus compares with other organizations exactly as does the machine with the non-mechanical mode of production. Precision, speed, unambiguity, knowledge of the files, continuity, discretion, unity, strict subordination, reduction of friction and of material and personal costs – these are raised to the optimum point in the strictly bureaucratic administration....
(1968: 973)

As a result of its technical superiority, "bureaucracy is among those social structures which are the hardest to destroy" (*ibid.*: 987).

There is, however, also a deeply negative and pessimistic side to Weber's evaluation of bureaucracy. Its machine-like character, which he so often points out with unconcealed admiration, is at the same time a source of grave concern for him. Bureaucracy, while being an extremely efficient and powerful social machine, is also a frightful force of dehumanization and enslavement.

Bureaucracy develops the more perfectly, the more it is "dehumanized"; the more completely it succeeds in eliminating from official business, love, hatred, and all purely personal, irrational, and emotional elements which escape calculation (*ibid.*: 975).

As bureaucracy operates inexorably 'according to *calculable rules* and "without regard for persons" ' (*ibid.*),

the professional bureaucrat is chained to his activity in his entire economic and ideological existence. In the great majority of cases he is only a small cog in a ceaselessly moving mechanism which prescribes to him an essentially fixed route of march (*ibid.*: 988).

Thus, through its dehumanizing, yet irresistible effects, bureaucracy is, according to Weber, increasingly becoming a "shell of bondage" (1968: 1402) or an "iron cage" (1958: 181), for modern humanity and a force leading society to a state of "mechanized petrification" (1958: 182). (For a collection of Weber's "melancholy metaphors" about our modern predicament see Kontos, 1994: 233).

Bureaucratization is thus for Weber a major – in fact the single most important – strand of rationalization in Western society; and its far reaching consequences – both for good and evil – are products of the particular forms of rationality embodied in bureaucracy. On the basis of the preceding account, how can these forms of rationality be characterized, in Weberian terms?

In the first place, bureaucracy possesses a high degree of formal rationality. Weber stresses, time and again, the calculability, as well as the predictability and controllability, of bureaucratic action, which is achieved through obedience to rules and instructions and reliance on specialized knowledge. Calculability is made possible by bureaucracy for three categories of stakeholders:

- (a) its extra-bureaucratic creators and/or superiors, those who set its goals, design and enact its system of rules, and/or overview its operation;
- (b) its members, who carry out the day to day work of a bureaucratic organization along narrowly determined paths and on the basis of a reliable anticipation of the actions of all others involved;
- (c) its clients, whose interaction with the organization takes place within well defined limits and on the basis of a specified range of expectations.

Over and above formal rationality, bureaucracy is, according to Weber, purposively rational with regard to the system of ends embodied in the tasks carried out by its members and expressing the goals of its external superiors and its clients. It is precisely its purposive rationality – its ability not only to calculate the results of action but to bring about intended results through the employment of appropriate means – that makes it "technically superior", and hence "irresistible".

Where, however, bureaucracy is severely lacking is in substantive rationality. To begin with, there is no guarantee that the ends pursued in the bureaucratic context

are rationally, or indeed in any other way, linked with ultimate values, though its "technical superiority" and "irresistibility" may testify to a certain degree of substantive rationality in respect of those ends.

But the real problems with regard to substantive rationality lie not so much with the ends as with the means and side effects of bureaucratic action. They concern the substantive irrationalities produced by bureaucratic formalization, those captured by terms such as "dehumanization", "iron cage", "mechanized petrification" and all the other "melancholy metaphors".

Thus, using terminology suggested earlier in this section, we can say that Weberian bureaucracy exhibits a formally well developed but substantively severely truncated purposive rationality.

Concluding Remarks

The above account of Weber's understanding of rationalization in connection with three major domains of modern social life – those of science, market-oriented action, and bureaucratic organizations – has shown that, for Weber, modern society is dominated by a particular kind of rationality : a formally extremely advanced but substantively highly deficient purposive rationality; one that involves the systematic use of sophisticated quantitative and logical calculation for the achievement of well defined, specific ends by the most efficient means, but which is, at the same time, seriously inadequate with regard to relating those ends and means to the realization of ultimate human values.

Weber is deeply ambivalent about this rationality, recognizing both great virtues and great vices in it, and staring unflinchingly at both these aspects at once, refusing to succumb to a facile choice between them. He is, however, also certain that this is the only rationality that can be had.

Weber flatly rejected any suggestion that problems of modernity such as the ones he had identified could be overcome by means of more rationality. As is well known, he vehemently repudiated socialist claims that a "rationally" planned economy would cater for fundamental human needs left unsatisfied by capitalism. He insisted that, on the contrary, "socialism would... require a still higher degree of formal bureaucratization than capitalism" (1968: 225), thus broadening and strengthening the "shell of bondage".

At the same time, he clearly conceives the (partial) remedies to the evils of rationalization which he proposes at times as non-rational, or even irrational. First and foremost among these is the creative force of charisma (see Mommsen, 1974: 82), which paradigmatically can be displayed at the political level by the great leader and at the economic level by the innovative entrepreneur.

Thus, Weber raised, and answered affirmatively, two questions: Firstly, whether western society is dominated by a highly formalized but substantively deficient purposive rationality; and secondly, whether that rationality is the only conceivable form of rationality.

There have been many serious challenges to Weber with regard to these questions, both on conceptual and on empirical grounds. Here it may suffice to make two interrelated points, while remaining within a Weberian conceptual framework: Firstly, it seems that it may be possible to have a substantively more satisfactory purposive (and in some areas also value-oriented) rationality than Weber allows. This could be achieved in large part by reducing formalism – the culprit, according to Weber, for the evils of rationalization – to levels that ensure the optimum realization of ultimate values. Secondly, certain developments since Weber's lifetime seem to support not only the possibility of a substantively more adequate (and less formalistic) rationality but also its empirical realization. Among

the most important of these developments is the considerable scientific and practical interest in less bureaucratic private, and more recently also public, organizational life.

To that we may add a definite movement in the direction of developing holistic (e.g. ecological) forms of science and technology, and even – despite the counter-evidence provided primarily by Soviet-type socialism, but also in large measure by the (considerably bureaucratized) welfare state – the emergence of economic forms where political, social, environmental and other non-market considerations based on ultimate values play a significant role.

NOTES

GENERAL INTRODUCTION

- 1) Despite this view, Kant was to become an enthusiast for the French Revolution, at least in its early stages.
- 2) Horkheimer and Adorno, 1972.
- 3) Horkheimer, 1974.
- 4) Marcuse, 1964.
- 5) Adorno, 1973.

PART ONE

- 1) This usage is of course narrower than the one I have adopted throughout the thesis whereby “practical rationality” concerns all action-related questions including both the “practical and the “technical” in the above narrower senses. This broad, inclusive sense of “practical” is employed also by the later Habermas, as for example in the important essay “On the Pragmatic, the Ethical and the Moral Employments of Practical Reason” (in Habermas, 1993), which we shall discuss in Part Two. In this study I shall use the term “praxial”, to correspond to *praxis*, as distinct from “practical”, corresponding to practice in the broad sense that covers both *praxis* and *techne*.
- 2) In Aristotle’s words “the rational faculty exercised in doing is quite distinct from that which is exercised in making” (1955: 175).

- 3) I first came across this passage in Georgia Warnke's discussion of Gadamer on the same issue (Warnke, 1987: 92-4).
- 4) See Habermas's essay "Labour and Interaction: Remarks on Hegel's Jena Philosophy of Mind", 1974: 142-169.
- 5) (a) In Habermas's own words, "Hegel is able to resolve [sublate] scientifically based social philosophy into a dialectical theory of society, and thereby select and develop the categories in such a way that this theory at every step is guided and permeated by the self-consciousness of its own relationship to *praxis*". (1974: 81)

(b) Referring, again, to Hegelian dialectics, Habermas speaks of "the dangerous potential of a theory which still understood its own critical relationship to *praxis*" (1974: 131).

(c) In discussing Marxist theory as critique and its relations to capitalist contradictions - in this case sufficiently deep to have acquired the character of crisis - Habermas maintains: "...critique becomes conscious of its own peculiar involvement in the object of its criticism. The objective complex of relations, which critique, although also encompassed within it, reflects as a totality and which, just by means of this, it wants to drive to the conclusion of the crisis, is obdurate. Consequently all efforts are equally condemned to remain without consequences, if they do not go beyond critique and intervene in the crisis, employing the means of the crisis itself, namely, practically... [O]nly by the success of this *praxis* can the critique itself become valid" (Habermas, 1973: 214).
- 6) This is analogous to (but drawn by Marx without knowledge of) the distinction between labour and interaction which Hegel works out in his early Jena

philosophy but later gives up (see "Labour and Interaction: Remarks on Hegel's Jena Philosophy of Mind", in Habermas, 1974: 142–169).

- 7) In this regard, Habermas also draws, though to a lesser degree, on another main exemplification of the method, namely Marx's critique of commodity fetishism and of the ideology of equal exchange (see Habermas, 1974: Chapter 6).
- 8) "In technical control over nature we get nature to work for us through our knowledge of causal connections. Analytic insight, however, affects the causality of the unconscious as such. Psychoanalytic therapy is not based, like somatic medicine, which is 'caused' in the narrower sense, on making use of known causal connections. Rather, it owes its efficacy to overcoming causal connections themselves" (1972: 271).
- 9) Habermas expresses his appreciation for McCarthy's contribution to the analysis of these concepts (Habermas, 1979: 224, n.31).
- 10) See Taylor, 1947. For a critical analysis of 20th century work processes Honneth refers us to Braverman (1974).
- 11) Giddens also refers approvingly to, and to some extent recapitulates, McCarthy's above-mentioned discussion of ambiguities in the concept of labour and interaction.
- 12) For an interesting attempt to reconcile the two positions, see Ricoeur, 1990. See also Ricoeur, 1970. For further discussions of the Habermas–Gadamer debate, and issues dealt with here in particular, see Mandelson, 1979, Warnke, 1987: Chapter 4, and Holub, 1991: Chapter 3.

PART TWO

- 1) It should be remembered that Habermas's new paradigm was being developed in a philosophical climate characterized by concerted onslaughts on the philosophy of the subject and philosophy of consciousness in both analytic and continental contexts: critiques of this kind came, most prominently, from the followers of Heidegger and the later Wittgenstein as well as from neo-pragmatism and poststructuralism.
- 2) These core ideas of Habermas's Universal Pragmatics are clearly prefigured in earlier writings such as the "Appendix" and "Postscript" to *Knowledge and Human Interests* (1972: 314, 363); *Legitimation Crisis* (1976: 107–8); and the "Introduction" to *Theory and Practice* (1974: 25). The following is a particularly telling passage:

What raises us out of nature is the only thing whose nature we can know: *language*. Through its structure, autonomy and responsibility are posited for us. Our first sentence expresses unequivocally the intention of universal and unconstrained consensus (1972: 314).
- 3) For a critical perspective on this point, see Schnädelbach, 1991; for Habermas's response see Habermas, 1991: 228ff.
- 4) For a list of rules of rational argumentation, which Habermas puts forward in the context of his discussion of moral discourse, though they are applicable to discourse in general, see Habermas, 1990: 87–9.
- 5) A notable exception, though one focusing on a particular aspect of the issue – that of human genetic engineering – is his *The Future of Human Nature* (Habermas, 2003a).

- 6) For a critical discussion of Habermas's relevant arguments in *Knowledge and Human Interests*, see Part One, above.
- 7) See Christian F. Rostboll, 2008.
- 8) For a brief discussion by Habermas of the possible *moral* status of our natural environment, and more particularly of possible duties towards animals, within a deontological perspective which distinguishes sharply between the ethical and the more narrowly moral, see Habermas, 1993: 105–111.
- 9) For Habermas, a substantive theory of justice such as Rawls's is permissible only "as a contribution to a discourse among citizens" (1990: 94). Rawls is able to propose it as a view possessing universal validity only after he has essentially rendered his perspective –wrongly, for Habermas – monological, through the device of the veil of ignorance, which eliminates all differences between potential participants in moral discourse and makes the position from which they would judge substantive norms identical (see Habermas, 1990: 66).
- 10) More precisely, Habermas distinguishes three forms of abstraction involved in the perspective of morality;

In fact, the moral point of view requires a threefold abstraction: (1) abstraction from the motives required of those involved, (2) abstraction from the particular situation, and (3) abstraction from existing institutions and forms of life (1993: 118).

- 11) For Günther's presentation of the principle of appropriateness, see Günther, 1993, especially, pp 221–245; for Habermas's discussion of the principle, see, Habermas, 1993: 35–39, and Habermas, 1996: 217–19.
- 12) In analyzing the rules of argumentation Habermas draws on the work of Robert Alexy. See Alexy (1990) in Benhabib and Dallmayr (1990).
- 13) For Habermas's fuller definitions of these terms, see his 1987: 138.
- 14) "Both aspects of society, which are initially introduced merely as different perspectives adopted in observing the same phenomena, also acquire essentialist connotations for modern societies and open up a view of differently structured domains of social reality itself" (Habermas, 1991: 255).
- 15) Habermas points out that systems theory – unlike communicative theory which possesses the evaluative standards of communicative action and communicative rationality – is not equipped with the conceptual tools that would enable it to comprehend any form of erosion of the lifeworld "as cost" (Habermas, 1987: 186).
- 16) See, for example, Blau, P.M. (1963).

PART THREE

- 1) These considerations show at the same time that when Habermas, in the theory of Universal Pragmatics, maintains that every speech act implies three distinct and communicatively redeemable validity claims – the claims to truth, truthfulness and rightness – what he can consistently mean by rightness is ultimately *moral* rightness. For a claim to rightness with respect to shared non-moral norms is neither entirely distinct (for it involves the other two claims, together with the moral one), nor necessarily redeemable in a purely

communicative way (for the sharing of the norms in question may well depend on strategic bargaining processes).

- 2) Examples of concepts akin to those of success-oriented communicative action and communicative negotiation (fully or restrictedly purposive-rational) are (a) Rawls's "overlapping consensus" (Rawls, 1993, 1996: 133ff); (b) Bohman's "moral compromise" (Bohman, 1996: 89ff); and (c) Aristotle's idea of imperfect friendship, which refers to a relationship into which people enter not because they "love one another for their personal qualities" but "only so far as they are useful to one another" (Aristotle, 1955: 231).
- 3) The complexity and costliness, in terms of time and effort, of negotiation processes will become clearer in the next section, where the process of communicative negotiation will be further clarified.
- 4) In this respect, Habermas refers to John Dewey's similar "pragmatistic model" of decision making, warning, though, against the model's direct application to decision making processes in modern mass democracies, because of the risk of ideologically "short-circuiting the connection between technical expertise and a public that can be influenced by manipulation" (1971: 69–70).
- 5) Joseph Heath's *Communicative Action and Rational Choice* (2001) represents an important effort to bring the Habermasian and Rational Choice views of practical rationality into dialogue with each other. Heath, unlike the present study, rejects a Habermasian rational-reconstructive justification of communicative rationality. He does, however, in line with this study, draw considerable support for cooperative modes of interaction and decision making from within Rational Choice Theory. Studies within the latter paradigm are shown to demonstrate that from a purely instrumental/purposive point of view, a cooperative approach is very often more advantageous than a

strategic-manipulative one – which is found to be beset by problems and antinomies.

- 6) For classic accounts of the “satisficing” method of reaching decisions, and of “bounded rationality”, see Simon, 1997: 88–9, 118–22. See also, Simon, 1957.
- 7) For a very interesting discussion of “deliberative inequalities” see Bohman, 1996: chapter 3.

APPENDIX II

- 1) Kant seems to be using the term “practical law” in three different senses, according to context: (a) in the sense of “principles” in general, in contradistinction to inclinations, impulses, etc; (b) in the sense of “objective principles”, in contradistinction to subjective principles; and (c) in the sense of “unconditional (objective) principles” in contradistinction to conditional ones.
- 2) In doing so, Kant exhibits a profound faith in the natural goodness of the ordinary person, one which many commentators trace to the influence of Rousseau as well as to Kant’s pietist family background.

APPENDIX III

- 1) For a comprehensive presentation of Weber’s life and work, see Bendix, 1962.

BIBLIOGRAPHY

- Adorno, Theodor W. (1973). *Negative Dialectics*. New York: The Seabury Press.
- Adorno, Theodor, et al (1976), *The Positivist Dispute in German Sociology*.
London: Heinemann.
- Alexy, Robert (1990). A Theory of Practical Discourse. See Benhabib and Dallmayr
(1990), 151–190.
- Apel, Karl-Otto (1987). The Problem of Philosophical Foundations in Light of a
Transcendental Pragmatics of Language. See Baynes, Bohman and
McCarthy (Eds) (1987), 250–295.
- Arendt, Hannah (1958). *The Human Condition*. Chicago and London: The
University of Chicago Press.
- Aristotle (1955). *The Nichomachean Ethics*. London: Penguin.
- Axelrod, Robert (1984). *The Evolution of Cooperation*. New York: Basic Books.
- Ayer, A.J. (1980). *Hume*. Oxford: Oxford University Press.
- Baynes, K., Bohman, J. and McCarthy, T. (Eds). *After Philosophy: End or
Transformation?*. Cambridge, MA, and London : The MIT Press.
- Bendix, Reinhard (1962). *Max Weber : An Intellectual Portrait*. London : Methuen.
- Benhabib, S. and Cornell, D. (Eds) (1987). *Feminism as Critique*. Minneapolis :
University of Minnesota Press.

Benhabib, Seyla (1987). *The Generalised and the Concrete Other: The Kohlberg–Gilligan Controversy and Moral Theory*. See Benhabib and Cornell (Eds) (1987), 77–96.

Benhabib, Seyla (1992). *Autonomy, Modernity and Community: Communitarianism and Critical Social Theory in Dialogue*. See Honneth, McCarthy, Offe and Wellmer (Eds) (1992), 39–59.

Benhabib, Seyla and Dallmayr, Fred (Eds) (1990). *The Communicative Ethics Controversy*. Cambridge, MA and London, England : The MIT Press.

Berger, Johannes (1991). *The Linguistification of the Sacred and the Delinguistification of the Economy*. See Honneth and Joas (Eds) (1991): 165–180.

Berlin, Isaiah (1979). *Against the Current: Essays in the History of Ideas*. London: The Hogarth Press.

Blau, P. M. (1963). *The Dynamics of Bureaucracy*. Chicago: The University of Chicago Press.

Bohman, James (1996). *Public Deliberation: Pluralism, Complexity, and Democracy*. Cambridge, MA, and London, England: The MIT Press.

Bohman, James (2000). *Distorted Communication: Formal Pragmatics as a Critical Theory*. See Hahn, (Ed) (2000), 3–20.

Braverman, H. (1974). *Labour and Monopoly Capital: The Degradation of Work in the Twentieth Century*. New York: Monthly Review Press.

Brubaker, Rogers (1984). *The Limits of Rationality*. London : Allen & Unwin.

Burns, Tom and Stalker, G. M. (1961). *The Management of Innovation*. Oxford: Oxford University Press.

Calhoun, Craig (ed) (1992). *Habermas and the Public Sphere*. Cambridge, MA and London: The MIT Press.

Cook, Maeve (1999). Habermas, feminism and the question of autonomy. See Dews, P. (Ed) (1999), 178–210.

Crozier, Michel (1964). *The Bureaucratic Phenomenon*. Chicago: The University of Chicago Press.

Dews, Peter (1993). "Agreeing what is Right: Review of *Faktizität und Geltung* by Jürgen Habermas", in *London Review of Books*, Vol. 15: 25–26.

Dews, Peter (Ed) (1999). *Habermas: A Critical Reader*. Oxford : Blackwell.

Ellul, Jacques (1964). *The Technological Society*. New York: Vintage Books.

Fisher, Roger, et al; (1991). *Getting to Yes: Negotiating An Agreement Without Giving In*. London: Century Business Books.

Fishkin, J. S. and Laslett, P. (Eds) (2003). *Debating Deliberative Democracy*. Oxford : Backwell.

Gadamer, Hans-Georg (1967). On the Scope and Function of Hermeneutical Reflection. See Gadamer (1976).

Gadamer, Hans-Georg (1975). *Truth and Method*. New York: Seabury Press.

Gadamer, Hans-Georg (1976). *Philosophical Hermeneutics*. University of California Press : Berkeley.

Gadamer, Hans-Georg (1981). *Reason in the Age of Science*. Cambridge, MA: The MIT Press

Gerth, H. H. and Wright Mills, C. (eds.) (1970). *From Max Weber : Essays in Sociology*. London : Routledge & Kegan Paul.

Giddens, Anthony (1972). *Politics and Sociology in the Thought of Max Weber*. London : Macmillan.

Gilligan, Carol (1982). *In a Different Voice : Psychological Theory and Women's Development*. Cambridge, MA : Harvard University Press.

Gouldner, A. W. (1954). *Patterns of Industrial Bureaucracy*. New York: The Free Press.

Günter, Klaus (1993). *The Sense of Appropriateness: Application Discourses in Morality and Law*. Albany: State University of New York.

Habermas, Jürgen (1970). "On Systematically Distorted Communication", in *Inquiry*, vol. 13, 1970: 205–18.

Habermas, Jürgen (1971). *Toward a Rational Society*. London: Heinemann.

Habermas, Jürgen (1972), *Knowledge and Human Interests* (Second Edition), Heineman : London.

Habermas, Jürgen (1974), *Theory and Practice*, Heineman: London.

Habermas, Jürgen (1976). *Legitimation Crisis*. London : Heinemann.

Habermas, Jürgen (1979). *Communication and the Evolution of Society*. London: Heinemann.

Habermas, Jürgen (1982). A Reply to my Critics. See Thomson and Held (Eds) (1982), 219–283.

Habermas, Jürgen (1984). *The Theory of Communicative Action (Volume One): Reason and the Rationalization of Society*. London: Heinemann.

Habermas, Jürgen (1987). *The Theory of Communicative Action (Volume Two): The Critique of Functionalist Reason*. Cambridge, UK : Polity, Press.

Habermas, Jürgen (1988), *On the Logic of the Social Sciences*. Cambridge, MA: The MIT Press.

Habermas, Jürgen (1989). *The Structural Transformation of the Public Sphere*. Cambridge, UK: Polity Press.

Habermas, Jürgen (1990). *Moral Consciousness and Communicative Action*. Cambridge, UK : Polity Press.

Habermas, Jürgen (1991). A Reply. See Honneth and Joas (Eds) (1991), 214–64.

Habermas, Jürgen (1992). *Postmetaphysical Thinking*. Cambridge, UK: Polity Press.

Habermas, Jürgen (1993). *Justification and Application*. Cambridge, UK: Polity Press.

Habermas, Jürgen (1996). *Between Facts and Norms*. Cambridge, UK: Polity Press.

Habermas, Jürgen (1998a). *On the Pragmatics of Communication*. Cambridge, UK: Polity Press.

Habermas, Jürgen (1998b). *The Inclusion of the Other*. Cambridge, MA: The MIT Press.

Habermas, Jürgen (2001). *On the Pragmatics of Social Interaction : Preliminary Studies on the Theory of Communicative Action*. Cambridge, UK: Polity Press.

Habermas, Jürgen (2003a). *The Future of Human Nature*. Cambridge, UK: Polity Press.

Habermas, Jürgen (2003b). *Truth and Justification*. Cambridge, MA: The MIT Press.

Hahn, Lewis Edwin (Ed) (2000). *Perspectives on Habermas*. Chicago and La Salle, Illinois: Open Court.

Heath, Joseph (2001). *Communicative Action and Rational Choice*. Cambridge, MA and London, UK: The MIT Press.

Heidegger, Martin (1993). The Question Concerning Technology. See Krell (Ed) (1993).

Heller, Agnes (1982). Habermas and Marxism. See Thompson and Held (Eds) (1982), 21–41.

Hobbes, Thomas (1962). *Leviathan*. UK: Fontana.

Holub, Robert (1991). *Jürgen Habermas: Critic in the Public Sphere*. London and New York: Routledge

Honneth, A., McCarthy, T., Offe, C. and Wellmer, A. (Eds) (1992). *Cultural-Political Interventions in the Unfinished Project of Enlightenment*. Cambridge, MA, and London: The MIT Press.

Honneth, Axel (1982), "Work and Instrumental Action", in *New German Critique*, No 25, pp31-54.

Honneth, Axel and Joas, Hans (Eds) (1991). *Communicative Action : Essays on Jürgen Habermas's "The Theory of Communicative Action"*. Cambridge, UK: Polity Press.

Horkheimer Max and Adorno, Theodor W. (1972). *Dialectic of Enlightenment*. New York: The Seabury Press.

Horkheimer, Max (1972). *Critical Theory: Selected Essays*. New York: Herder and Herder.

Horkheimer, Max (1974). *Eclipse of Reason*. New York: The Seabury Press.

Horowitz, A. and Malley T. (eds.) (1994). *The Barbarism of Reason : Max Weber and the Twilight of Enlightenment*. Toronto: University of Toronto Press.

Hume, David (1975). *Enquires Concerning Human Understanding and Concerning the Principles of Morals*. (Edited by L.A. Selby-Bigge and P.H. Nidditch). Oxford: Clarendon Press.

Hume, David (1978). *A Treatise of Human Nature* (Edited by L.A. Selby-Bigge and P.H. Nidditch). Oxford: Clarendon Press.

Joas, Hans (1991). The unhappy Marriage of Hermeneutics and Functionalism. See Honneth and Joas (Eds) (1991) : 97-118.

- Kalberg, Stephen (1994). *Max Weber's Comparative–Historical Sociology*. Chicago: The University of Chicago Press.
- Kant, Immanuel (1948). *Groundwork of the Metaphysics of Morals* (Translated and analysed by H.J. Paton). London: Hutchinson.
- Kant, Immanuel (1983). "An Answer to the Question: What is Enlightenment", in Kant, Immanuel, *Perpetual Peace and Other Essays*. Indianapolis: Hackett Publishing Company.
- Kant, Immanuel (1996). *Critique of Practical Reason* (Translated by T.F. Abbott). New York: Prometheus Books.
- Kontos, Alkis (1994). The World Disenchanted, and the Return of Gods and Demons. See Horowitz, A. and Maley, T. (eds.) : 223 – 247.
- Körner, S. (1955). *Kant*. London: Penguin.
- Krell, David Farrell (Ed) (1993). *Martin Heidegger : Basic Writings*, Second Edition. New York: Harper Collins.
- Lax, David, A. and Sebenius, James, K. (1986). *The Manager As Negotiator: Bargaining for Cooperation and Competitive Gain*. New York: The Free Press, Macmillan.
- Locke, John (1988). *Two Treatises of Government* (Edited by Peter Laslett). Cambridge: Cambridge University Press.
- Lukács, Georg (1971). *History and Class Consciousness*. London: Merlin.
- Lytard, Jean–Francois (1984). *The Postmodern Condition: A Report on Knowledge*. Minneapolis : University of Minnesota Press.

- MacIntyre, Alasdair (1996). *A Short History of Ethics*. London: Routledge and Kegan Paul.
- Marcuse, Herbert (1964). *One-Dimensional Man*. Boston: Beacon Press.
- McCarthy, Thomas (1978). *The Critical Theory of Jürgen Habermas*. Cambridge, MA and London, UK: The MIT Press
- McCarthy, Thomas (1982). Rationality and Relativism : Habermas's "Overcoming" of Hermeneutics. See Thomson and Held (Eds) (1982), 55–78.
- McCarthy, Thomas (1991). Complexity and Democracy: or the Seductions of Systems Theory. See Honneth and Joas (Eds) (1991): 119–139.
- Mendelson, Jack (1979), "The Habermas–Gadamer Debate" in *New German Critique*, 1979, no. 18, pp. 44–73.
- Merton, R. K. (1957). "Bureaucratic Structure and Personality", in *Social Theory and Social Structure*, pp 195–206. New York: The Free Press of Glencoe.
- Mommsen, Wolfgang J. (1974). *The Age of Bureaucracy : Perspectives on the Political Sociology of Max Weber*. London : Harper and Row.
- Ormiston, G.L. and Schrift, A.D. (Eds) (1990), *The Hermeneutic Tradition: From Ast to Ricoeur*. State University of New York Press: Albany.
- Ottmann, Henning (1982). Cognitive Interests and Self-Reflection. See Thomson and Held (Eds) (1982): 79-97.
- Paton, H.J. (1948). *The Categorical Imperative: A Study in Kant's Moral Philosophy*. Philadelphia: University of Pennsylvania Press.

- Raiffa, Howard, et al (2002). *Negotiation Analysis: The Science and Art of Collaborative Decision Making*. Cambridge, MA, and London: Harvard University Press.
- Rawls, John (1993, 1996). *Political Liberalism*. New York: Columbia University Press.
- Ricoeur, Paul (1970). *Freud and Philosophy: An Essay on Interpretation*. New Haven and London: Yale University Press.
- Ricoeur, Paul (1990). Hermeneutic and the Critique of Ideology. See Ormiston and Schrift (Eds) (1990): 298:334.
- Rostboll, Christian F. (2008). Emancipation or Accommodation: Habermasian vs Rawlsian deliberative democracy, in *Philosophy and Social Criticism* vol. 34, (2000) no. 7, pp 707–736.
- Schnädelbach, Herbert (1991). The Transformation of Critical Theory. See Honneth and Joas (Eds) (1991), 7–22.
- Simon, Herbert (1957). *Models of Man: Social and Rational*. New York: John Wiley.
- Simon, Herbert (1997). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations* (Fourth Edition). New York: The Free Press.
- Swindler, Ann (1973). “The Concept of Rationality in the Work of Max Weber”, in *Sociological Inquiry* 43 : 35–42.
- Taylor, Charles (1991). Language and Society. See Honneth and Joas (Eds) (1991), 23–35.

Taylor, Frederick (1947). *Scientific Management*, London: Harper and Row.

Thomson, J. B. and Held, D. (Eds) (1982). *Habermas: Critical Debates*. London and Basingstoke : Macmillan.

Warnke, Georgia (1987). *Gadamer: Hermeneutics, Tradition and Reason*. Cambridge, UK: Polity Press.

Weber, Max (1958). *The Protestant Ethic and the Spirit of Capitalism*. New York: Charles Scribner's Sons.

Weber, Max (1968). *Economy and Society : An Outline of Interpretive Sociology*, 2 vols., edited by Günther Roth and Claus Wittich. London: University of California Press.

Wellmer, Albrecht (1971). *Critical Theory of Society*. New York: Continuum, The Seabury Press.

Young, Iris Marion (2003). *Activist Challenges to Deliberative Democracy*. See Fishkin and Laslett (Eds) (2003), 102–120.

