NOTICE

AVIS

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

If pages are missing, contact the university which granted the degree.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

VALIDITY AND VALIDATION

IN

KANT'S AND HABERMAS' MORAL THEORIES
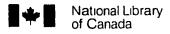

Chayah Vermes

A Thesis

in

The Department

of

Philosophy

Presented in Partial Fulfilment of the Requirement
For the Degree of Master of Arts at
Concordia University
Montreal, Quebec, Canada

January, 1994

ISBN   0-315-90929-3

Canadä

ABSTRACT

VALIDITY AND VALIDATION
IN KANT'S AND HABERMAS' MORAL THEORIES

Chayah Vermes

In this thesis, Jurgen Habermas' discourse ethics is examined in relation to Immanuel Kant's categorical imperative. Discourse ethics is considered by Habermas to be a procedural reformulation of the categorical imperative. The question is how to understand discourse ethics as such a reformulation given the fact that Habermas' worldview as well as his conceptions of consciousness and rationality differ from that of Kant.

In both cases we have a moral principle that is believed to be universal, formal and rationally justifiable. There are similarities and differences in Kant's and Habermas' conceptions of moral consciousness and their methods of validation of moral norms; these are examined, analyzed and compared.

The similarities are in moral concepts such as practical reason (a capacity for moral judgements in all human beings), autonomy, plurality of ends-in-themselves, a formal structure of human interrelations, moral reasoning. The differences are in the very conceptions of consciousness, rationality, morality and society.

Discourse ethics can be seen as a critical social scientific interpretation of the categorical imperative. Habermas rejects what he calls Kant's conception of monological rationality and morality in favour of a historically conscious communicative or dialogical rationality and morality. Kant's cognitivism, formalism and universalism, although retained, are given a different interpretation.

The value and significance of Habermas' discourse ethics, as compared to that of Kant, is that it is not only a moral principle guiding **public** discourse on the validity of moral norms (moral argumentation) but at the same time constitutes a normative standard for social criticism as well.

v

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

For Kant's works:

Fund.

Fundamental Principles of the
Metaphysics of Morals, trans.
Thomas K. Abbott. Indianapolis:
Bobbs-Merrill, 1949.

Second Critique

Critique of Practical Reason, trans.
Lewis White Beck. Indianapolis:
Bobbs-Merrill, 1956.

For Habermas' works:

CES

Communication and Evolution of Society,
trans. Thomas MacCarthy. Boston:
Beacon Press, 1979.

MCCA

Moral Consciousness and Communicative
Action, trans. C.Lenhardt and S.Weber
Nicholsen. Cambridge: Mass.: MIT,
1990.

TCA

Theory of Communicative Actions,
trans. Thomas McCarthy.
Boston: Beacon Press, 1984. 2v.

INTRODUCTION

The purpose of this thesis is to examine Jurgen Habermas discourse ethics in relation to Immanuel Kant's categorical imperative. Habermas has been influenced by Hegel (historical consciousness), Karl Marx (Critical theory), Max Weber (social theory), as well as by Ludwig Wittgenstein's philosophy of language, American pragmatism, hermeneutics, systems theory, and others, yet he considers his discourse ethics to be a procedural reformulation of the categorical imperative. This thesis will explore in more detail Kantian similarities as well as differences in Habermas' discourse ethics and its validating methods.

Although there are fundamental differences between Habermas' and Kant's worldviews, Habermas does share with Kant certain important concerns and conceptions which directly influence their respective moral theories. They are both concerned about the negative moral implications of certain philosophical theories of their own time. Kant was concerned about the negative moral implications of the dogmatism of traditional metaphysics and of moral fanaticism (uncritical acceptance of traditional moral norms); yet, he was equally concerned about the moral implications of empiricism and

ethical scepticism (Hume),[1] utilitarianism, as well as of the
moral implications of Newton's discovery that the natural
world is governed by morally neutral causal laws.

Habermas is, likewise, concerned about the negative
moral implications of positivism (which identifies rationality
solely with scientific knowledge, while ethics is relegated to
the status of emotivism and scepticism) and of philosophical
theories which are influenced by Nietzsche's will to power
(such as Michel Foucault and Jean-Francois Lyotard, among
others).[2] In particular, Habermas is concerned about the moral
implications of the positivistic approaches in the social
sciences (such as systems theories) and he is equally
concerned about the moral negative effects of what he calls
Kant's "methodological solipsism", or "philosophy of
consciousness", even while trying to retain Kant's moral
ideals. Richard Kroner describes Kant's worldview as wholly
determined by his moral outlook;[3] I believe Habermas'
worldview can be described in the same way.

What guides both Kant's and Habermas' thinking is
their concern for moral and individual autonomy. However,

---

[1] David Hume, <u>Treatise of Human Nature</u>, Selby-Bigge (ed.)
(London: Oxford, 1888), p.415 : "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".

[2] Jurgen Habermas, <u>The Philosophical Discourse of Modernity:
Twelve Lectures</u>, trans. Frederick Lawrence (Cambridge, U.K.: Polity
Press, 1987). chaps. iv-x.

[3] Richard Kroner, <u>Kant's Weltanschauung</u>, trans. John E.Smith
(Chicago : The University of Chicago Press, 1956).

their conceptions with regard to these ideals differ. Kant's concern was for rational and moral freedom or autonomy of the individual consciousness from traditional thinking, while Habermas' concern is for the freedom of the individual from coercion, be it political (ideology, forced legitimacy, forced consensus), social, or psychological. It is a concern for the freedom and autonomy of individuals in their interpersonal relations and communication, which Habermas considers to be the basic unit of society.

Both conceive of morality as universalistic and are in search of a formal principle which will serve as a basis for morality, and they both believe that the principle can be justified in a rational manner, although their conceptions of rationality differ. Kant thinks in terms of a transcendental justification of the supreme moral principle, while Habermas regards all knowledge including morality as hypothetical.

Kant was influenced by Jean-Jacques Rousseau, who was the first to see the necessary connection between freedom, equality and the law. 'Autonomy' means in Greek "self-legislation" which is to be distinguished from 'autocracy' which is "self-rule" (that refers to a relation of power). Rousseau understood autonomy in political terms (a self-governed state).[4] Kant was the first to extend it to moral individual autonomy, i.e. self-legislation of the moral law

---

[4] Jean-Jacques Rousseau, The Social Contract, Bk.II, chap. vi.

(or "republicanism in moral philosophy").[5] In both cases the idea is rooted in the Enlightenment's individualistic conception of what a human being is or should be.

The Enlightenment's conception begins with Descartes' "solitary thinker", obtains its positive form in Kant's morally autonomous subject, and is finally transformed into the Romantic conception of individual self-expression. The overall guiding idea in this conceptual progress is the abstract individual who is envisioned as the bearer of given interests, wants, needs, etc., which are independent and prior to society. This abstract individual is also conceived of as having inalienable rights for self-determination and self-realization. Society is conceived of as an aggregation of autonomous individuals who come to live together for the purposes of security with the means of a social contract.[6]

This individual is also envisioned as an independent center of consciousness that is rationally creative and by itself can gain rational insight into moral or theoretical judgements valid for all rational beings. This is based on the assumption that rationality is substantive, a-historical and identical for every individual. This conception is now being referred to as "methodological solipsism" or "philosophy of consciousness" or "the monological subject" by Habermas and is

---

[5] Rudiger Bittner, What Reason Demands, trans. Theodore Talbot (Cambridge : Cambridge University Press,1989), p.75.

[6] Steven Lukes, Individualism: Key Concepts in the Social Sciences (Oxford: Blackwell, 1973), pp.43-78.

rejected by him.

In addition to this individualistic conception, Kant conceives of nature as a harmonious system of purposes. All things in nature serve some purpose which is external to them, except the human being who as a rational being is believed to be the subject of its own ends; in other words, the human being is an independent end. Being an independent end endows the human being with an **intrinsic** or absolute value, while all other things in nature have only a **relative** value.[7]

Kant conceives of reality as having two **aspects**: the phenomenal and the noumenal (intelligible). He does not say that there are two worlds, only that it has two aspects for us, so that certain things can be understood either from one aspect or the other. The phenomenal world is that which is given to our senses and thus we can have theoretical knowledge of it, but as moral beings we do seem to belong to the noumenal world, the intelligible world, of which we can have no theoretical knowledge but Ideas of reason that do seem to guide us in our thoughts and actions (practical reason). As Kant understands it, the noumenal world is the ground for the phenomenal world so that things such as the moral phenomenon cannot be explained by our limited reason because they

---

[7] Immanuel Kant, <u>Fundamental Principles of the Metaphysic of Morals</u>, trans. Thomas K. Abbott (New York : Bobbs-Merrill Co. Library of Liberal Arts, 1949), p.45. Later references to pages in this book will appear in the text in parentheses (<u>Fund</u>.).

originate in the noumenal self (which he refers to as our "proper self") (Fund. 75).

Habermas' moral theory, on the other hand, is part of his critical social theory, which includes a theory of society, a theory of rationality as well as a moral theory. It is a social philosophy with a **practical** (or normative) intent. Critical social theory

> is critical both of contemporary social
> sciences and of the social reality they are
> supposed to grasp...
> Critical social theory does not relate to
> established lines of research as a competitor;
> starting from its concept of the rise of modern
> societies, it attempts to explain the specific
> limitations and the relative rights of those
> approaches.[8]

It is important to note that for Habermas, rationality has less to do with knowledge (the positivistic understanding of it) than with how, as speaking and acting subjects, we acquire knowledge and how we use it (communicative action) (TCA I, 8). This is important to remember, since it is this conception that is at the base on which his critical social theory rests, and characterizes the way he adapts, or critizes, empirical and philosophical theories in general. Such a conception of rationality has moral, political and psychological implications; it has **normative** implications that go beyond the attitude of an observer (which is the attitude

---

[8] Jurgen Habermas, The Theory of Communicative Action, trans. Thomas McCarthy (Boston: Beacon Press, 1984-1987) 2v. here v.2.,p.375. Later references to pages in this book will be given in the text in parentheses (TCA).

of the "knower"). It presupposes both an attitude of first and second person and an intersubjective understanding and interaction.

From the normative point of view, there seems to be "a paradox of rationality" in that advances in science and technology (which are the result of the rationalization of the worldview and which were supposed to be considered a progress) are now understood to be threatening to destroy the very same lifeworld that produced these advances. The hopes of the Enlightenment that rationality would emancipate from oppression of tradition and politics seem to have become self-defeating and to have backfired.

Habermas argues that this situation is a result of conceiving of 'rationality' solely in terms of instrumental or purposive rationality (which is a subject vs. object relation, where subject manipulates the object in order to achieve desired goals), while ignoring the fact that there is also a different kind of rationality that is related to interpersonal relations and self-understanding (where the relation is subject to subject) and which belongs to the domain of the lifeworld (TCA I, chap. iii).

Habermas' project in his Theory of Communicative Action (TCA) and in his moral writings (Moral Consciousness and Communicative Action (MCCA)) is to show that rationality as conceived by the positivists (cognitive-instrumental rationality) is one-sided and that scepticism in the moral

domain is unwarranted since moral norms can be rationally defended or criticized (cognitive-normative rationality) (MCCA, 43-109). Here there is an essential similarity between Habermas and Kant. With Kant he regards rationality as having two functions: one theoretical (subject vs. object approach), the other practical (first and second person's approach); and also with Kant, he regards practical rationality as having priority over the theoretical. However, while Kant was in search of an a-priori (transcendental) justification of principles of reason, both theoretical and practical, Habermas' regards all knowledge, scientific or moral, as hypothetical.

Rationality for Habermas is embodied in intersubjective relations that are linguisticallly mediated. Language is the medium by which we can arrive at a mutual understanding with regard to things in the natural world, the social world and the self. He considers action aimed at reaching understanding and cooperation in interpersonal relations - "communicative action" - to be fundamental ("undistorted" communication) while conflicts, competition or strategic actions are derivative ("distorted" communication).[9] Undistorted communication implies for Habermas a communicative rationality which is

---

[9] Jurgen Habermas, Communication and Evolution of Society, trans. Thomas McCarthy (Boston : Beacon Press, 1979), p.1. Later references to pages in this book will appear in the text in parentheses (CES).

based ultimately on the central experience of the
unconstrained, unifying, consensus-bringing force of
argumentative speech, in which different participants
overcome their merely subjective views and, owing to the
mutuality of rationally motivated conviction, assure
themselves of both the unity of the objective world and
the intersubjectivity of their lifeworld. (TCA II, 10)

If it is said of Kant that he extended Rousseau's
ideas of autonomy and equality to moral philosophy, then it
can perhaps be said of Habermas that he is trying to extend
Kant's moral ideals and Rousseau's political ideals to social-
cultural philosophy. Habermas' critical social theory tries
to encompass the moral, political and social aspects of
interpersonal relations that he considers to be the basic unit
of society.

As it happened, the social sciences have developed as
separate disciplines each of which is concerned with its own
special domain - such as political science, economics, public
administration, psychology - along the lines of the natural
sciences. However, even the social sciences that deal with
society as a whole (systems theory) seem to deal only with the
functional (self-preservation, or self-maintaining, or system
integration) aspects of society but not with its cultural
norms and values.

Habermas does not reject the positivistic social
sciences as such (in fact, he does rely on knowledge so
obtained and believes firmly that philosophy should cooperate

with them, not compete or show them what to do).[10] Rather,

he considers them as a one-sided approach to the study of

society, especially with regard to modern societies. He is

also critical in the same way of modern philosophical theories

that he does accept (from Hegel's historical consciousness,

through Husserl's lifeword (phenomenology), American

Pragmatism, and others. He is always careful to point out the

limitations of the conception of (instrumental, functionalist)

rationality that is at the root of the above mentioned

theories and the need to supplement them with his conception

of practical reason.

Modern society, according to Habermas, can and should

be studied from two points of view which he regards as

essential: systems theories (TCA II, 235-282) on the one hand,

but on the other hand, he claims that the "lifeworld" (which

the positivistic social sciences try to leave out of their

studies) should not only be studied but should be regarded as

primary. By the concept of "lifeworld" (which he borrows from

Edmund Husserl, although he rejects Husserl's conception of

the transcendental ego because of its "methodological

solipsism") Habermas means all the implicit taken-for-granted

and unquestioned presuppositions of knowledge, norms and self-

understanding, as well as social institutions that constitute

---

[10] Jurgen Habermas, _Moral Consciousness and Communicative Action_, trans. C. Lenhardt and others (Cambridge, Mass. : MIT Press,1991), pp.18-9. Later references to pages in this book will be given in the text in parentheses (_MCCA_)

the solid background of what we call "culture" (or tradition).
It is the so-called "pretheoretical knowledge" of the
individual members which is transmitted from one generation to
another by means of socialization (TCA, II, 119-152). This is
the aspect of social integration.

With regard to social analysis, Habermas considers
system integration (systems theory) and social integration
not to be regarded as merely points of view. They are in fact,
according to him, interrelated and complementary functions
(TCA II, 153-197). Systems theory is concerned exclusively
with the material reproduction of society and its
organization, which has a logic of its own, while the
lifeworld represents the symbolic reproduction of society
(norms, values, etc). The two functions, according to
Habermas, should never be considered separately but always as
interrelated and complementary, and as mutually influencing
one another. It is when there is a one-sided influence of one
of these functions over the other that, according to Habermas,
a crisis in a society arises (TCA II, 153-197).

In systems theories, social processes are understood
to operate "behind the backs", so to speak, of the individual
members of society. Habermas sees in these blind (functional)
social processes (when unchecked by normative considerations)
a damaging effect on the autonomy of the lifeworld and its
individual members. This can be manifested by loss of meaning
and freedom, domination, breakdown of communication, anomie,

a sense of alienation, etc. Habermas speaks of the "colonization" of the lifeworld by these social processes. By "colonization" (TCA II, 332-373) he means that the lifeworld becomes increasingly subordinated to system imperatives which require **strategic** actions of all sorts, while social integration requires **communicative** actions (TCA II, 277-282).

Normally, members of a culture cannot take up an observer's viewpoint in relation to their lifeworld as a whole (since their very identity is drawn from the culture's norms and values). However, Habermas considers a critical (rational) attitude towards one's cultural norms essential in order to avoid domination and loss of control of individuals over their own lives. Such a critical attitude presupposes a distinction between nature and society and an awareness of a distinction between socially (**de facto**) accepted norms and the **worthiness** of norms to be considered valid on their own (regardless of their actual acceptance or not).

Habermas differentiates between traditional and modern societies (TCA I, 43-75). In traditional societies no differentiation is made between "internal connection of meaning and external connection of objects" (TCA I, 49) and there is little perception that the world (whether it be the natural or the social) is linguistically interpreted; therefore, there is less awareness in traditional societies that such an interpretation can be subject to error and open to criticism (TCA I, 50).

By contrast, what characterizes modern societies is, according to Habermas, an awareness that the understanding of the world (natural, social, self) is a linguistic interpretation whose truth or validity can be questioned. This attitude is believed to be less dogmatic than in traditional societies in that there is an awareness of the possibility of the presence of unexamined and unquestioned presuppositions and motives in thought which necessitates a continual critical approach (TCA I, 69-70).

According to Habermas, linguistic utterances or speech acts (a concept he borrows from Austin[11] and Searle[12]) (CES, 1-68; TCA I, 287-328) imply validity claims that we take for granted because they are transmitted to us by tradition. These validity claims refer to the natural world, to the social world and its norms, and to the individual (privileged access) world. Normally, these validity claims are not questioned wholesale; yet, it is because validity claims are understood in modern societies to be open to criticism, that the process of coming to an understanding is important. This is in particular so both with regard to how knowledge is acquired and how it can be used, and with regard to socially accepted norms that have become problematic:

> From the perspective of the participants, coming to
> an understanding is not an empirical event that
> causes de facto agreements; it is a process of

---

[11] John Austin, How to Do Things with Words (Oxford, 1962).

[12] J.R.Searle, Speech Acts (London, 1969).

> mutually convincing one another in which action of
> participants are coordinated on the basis of
> motivation by reasons. 'Coming to an understanding'
> refers to communication aimed at achieving valid
> agreement. It is only for this reason that we may
> hope to obtain a concept of rationality by
> clarifying the formal properties of action oriented
> to reaching understanding - a concept expressing
> the interconnection of those moments of reason that
> became separated in the modern period, no matter
> whether we look for these moments in cultural value
> spheres, in differential forms of argumentation, or
> in communicative practice of everyday life however
> distorted that may be. (TCA I, 392)

Habermas rejects Kant's "philosophy of consciousness"
and Hegel's "philosophy of the collective subject". In both
cases, either in Kant's noumenal, monological thinking, or
"proper" self, or in Hegel's collective acting subject (which
actualizes itself in history) we have metaphysical models of
the "self". Both imply a uniformity of humankind, an
underlying common "nature", and provide a paradigm which is
taken to be the yardstick against which the concrete
historical selves are measured and found ontologically
deficient.[13]

Instead, Habermas accepts Wittgenstein's philosophy of
language and language games (TCA I, 95-97) with the
implications of a plurality of forms of life and human
variety. Consciousness is understood to be culturally,
historically and intersubjectively constituted in an open-
ended process, and not in accord with any pre-established
metaphysical paradigms. The idea of the unity of reason, its

---

[13] Seyla Benhabib, Critique, Norm, Utopia (New York: Columbia
University Press, 1986), pp.327-330; p.393, n. 37.

certainty, its self-evident and unconditional character is being rejected, which amounts to a rejection of an ultimate foundation of reason in any form (TCA I, 2). Rationality is no longer understood to be "pure"; body and mind are no longer understood to be separated:

> Rationality is understood to be a disposition of speaking and acting subjects that is expressed in modes of behavior for which there are good reasons or grounds. This means that rational expressions admit of objective evaluation. This is true of all symbolic expressions that are, at least implicitly, connected with validity claims (or with claims that stand in internal relation to a criticizable validity claim). Any explicit examination of controversial validity claims requires an exacting form of communication satisfying the conditions of argumentation. (TCA, I, 22)

Habermas considers norms of actions to be understandable within their context and thus to be connected with validity claims (MCCA, 45-51). Although these are not facts in the sense of "natural" facts, within their context we can consider them to be either legitimate or illegitimate. This means that actions, or rather acting and speaking individuals, can be judged to be rational or irrational, moral or immoral, sincere or insincere, in the sense that their actions or speech acts are judged to be consistent or inconsistent with commonly accepted norms or beliefs. We can find good reasons to justify or to unjustify behaviour or actions the cause of which may be emotions, feelings, or whatever; a behaviour which we can understand and which does make sense to us (no matter what its cause is) will be rational behaviour; if not, we may consider it to be

irrational (TCA I, 16).

Rationality, for Habermas, is communicative (or dialogical); both theoretical and normative statements are public assertions and their questioning is meaningful only within a public discourse. As with Wittgenstein, there is no private language for Habermas; when conflicts in interpretations occur, they are to be resolved in a meaningful way only after they are publicly discussed, argued and agreed upon by the individuals involved.

Ever since Kant, morality has been regarded as autonomous and the question is whether moral norms can be rationally justified on their own, as Kant believed they could, or whether they are subjective or cultural preferences as the moral sceptics claim. However, if we do not wish to go back to pre-Kantian uncritical (authoritarian) morality then, according to Habermas, we have to establish rules and norms in the light of which we can critically examine our inherited moral norms. It should be remembered that discourse ethics is part of Habermas' critical (normative) social theory and therefore establishing the validity of a non-relative normative standard for social criticism is of prime importance to him.

Kant's ethics is considered to be too formalistic, rigorous and foundationalist, all of which are regarded by Habermas to be related to Kant's monological conception of consciousness; discourse ethics is an attempt to overcome

these weaknesses in Kant's moral theory with a dialogical conception of ethics, which may be implicit in Kant's categorical imperative but is overshadowed by his monological conception of consciousness.

Within contemporary philosophy, Habermas can be placed between neo-Kantians such as John Rawls and Alan Gewirth (among others) on the one hand, and on the other, "communitarians" such as Richard Rorty, Alasdair McIntyre and Jean-Francois Lyotard (among others). Although Habermas is a neo-Kantian ("deontologist", "cognitivist", "universalist", "formalist") he differs from the other neo-Kantians in that he claims that their conception of consciousness is monological (MCCA, 66-67); and although he is a pluralist (plurality of of forms of life), he differs from the communitarians in that he believes that the approach to ethics should be critical and that the normative standard for such criticism or agreement must transcend the particular community (MCCA, 19).

The thesis will be divided into three chapters: in chapter I, I will examine Kant's categorical imperative, its different interpretations and misinterpretations, as well as Kant's methods of validation, both of moral actions and of the validation of the categorical imperative itself. In chapter II, Habermas' theoretical conception of the development (reconstructive science) of moral consciousness in the individual and in the species will be examined. Chapter III will concentrate on Habermas' conception of what constitutes

validity of norms and on his justification of discourse ethics, all in relation to Kant's conception of the categorical imperative.

CHAPTER I

KANT'S CATEGORICAL IMPERATIVE

Kant's categorical imperative has been differently interpreted and it is important to keep this in mind, since the criticism directed towards it depends on the particular interpretation given to it by the critic. This results in a confusion which I think we should be aware of before any discussion in relation to the categorical imperative can begin.

One reason for the confusion is that Kant uses the same term with different senses, such as 'principle' (analytic or normative), 'will' (Willkur=subjective will, Wille=objective will), 'command' as a practical proposition and as an imperative, freedom (negative and positive), without making it explicitly clear, most of the times, that he is so doing. These and other ambiguities in Kant's terminology have given rise to many interpretations and misinterpretations of Kant's intentions. Although it is indeed sometimes difficult to interpret what Kant really meant, it is nevertheless agreed in general that he should not be interpreted in a way that is inconsistent with his overall view known from his different writings. However, even in this regard there is no complete agreement.

In addition, although Kant does claim that there is only one categorical imperative (Fund., 38), it is not always clear what he uses the phrase 'categorical imperative' to refer to. Paton, for example, points to five basic various formulations, all of which, he argues, are in fact interrelated and are contained in the one main formula (the Formula of Universal Law).[14] Yet, there are various formulations, all of which are to be taken into account.

Also, the very conception of what the categorical imperative as a principle is, its status or its place in Kant's thinking, have been differently interpreted as well as his methods of validation of the categorical imperative.

In the following, I will try to clarify some of the confusion. In order to better grasp the unity of Kant's moral theory, I shall present (in Section 1) an analysis of Paton's scheme of the five formulae of the categorical imperative. Paton is not the only one to have done it; Bruce Aune, for example, has worked out a somewhat different scheme.[15] T.C. Williams, who himself suggests four,[16] mentions others who

---

[14] Herbert J. Paton, The Categorical Imperative: A Study in Kant's Moral Philosophy (London : Hutchinson, 1963), pp.129-198.

[15] Bruce Aune, Kant's Moral Theory (Princeton: Princeton University Press, 1979).

[16] T.C. Williams, The Concept of the Categorical Imperative: A Study of the Place of the Categorical Imperative in Kant's Ethical Theory (Oxford : Clarendon Press, 1968), p.25.

suggested different formulations.[17]   John R. Silber claims that   "the number is actually indeterminate because Kant begins with the moral law as a single formal principle and attempts to make its meaning increasingly clear or intuitive by a variety of formulations".[18]   Kant himself speaks of three ways of presenting the main formula (Fund., p.53).

Section 2 will deal with the different interpretations regarding the status of the categorical imperative, and in section 3 I will examine Kant's methods of validating the categorical imperative as the supreme principle of morality.

1.   The Different Formulae of the Categorical Imperative

The five formulae of the categorical imperative are, according to Paton, as follows:[19]

Formula I (Universal Law): Act only on that maxim through which you can at the same time will that it should become a universal law. (Fund., 38)

What is important to note here is to will that one's chosen maxims (principles on which we act) be valid for all rational beings (universal law). This ensures that one's will is not subjective or arbitrary but rationally motivated. It

---

[17] Williams, The Concept of the Categorical Imperative, pp.23-25.

[18] John R. Silber, "Procedural Formalism in Kant's Ethics", Review of Metaphysics, 28, no.2 (1974-75) : 205-206.

[19] Paton, The Categorical Imperative, pp.129-198.

will be rationally self-contradictory, according to Kant, to will that a maxim based on subjective interests become universal law. Here it must be remembered that for Kant rationality is identical for all rational beings, so that a rationally motivated will can not err and will not be in conflict with other rational wills; also, it is important to remember his teleological view (harmony of purposes).

In principle, Kant distinguishes between (will=Willkur, the subjective will) and (will=Wille, the rational objective will) but he often uses the same term for both without making it clear that he is so doing.[20]

> Everything in nature works in accordance with laws. Only rational beings alone have the faculty of acting according to the conception of laws — that is, according to principles, that is, have a will. Since the deduction of actions from principles requires reason, the will is nothing but practical reason. (Fund., 30)

Kant does not say that there are two wills, only that its motivation can come from two different sources, one that is influenced by subjective desires, the other motivated by an objective principle valid for all rational beings.

Wille is the capacity to recognizes what is practically necessary (Fund. 30). Kant differentiates between hypothetical imperatives (valid for all in certain situations) and categorical imperatives (valid for all without exception in all situations)

---

[20] Lewis White Beck, Commentary on Kant's Critique of Practical Reason (Chicago:University of Chicago Press, 1960), p.178.

> When I conceive a hypothetical imperative in general, I do not know beforehand what it will contain until its condition is given. But when I conceive a categorical imperative, I know at once what it contains. For as the imperative contains besides the law only the necessity that the maxims shall conform to this law, while the law contains no conditions restricting it, there remains nothing but the general statement that the maxim of the action should conform to a universal law, and it is this conformity alone that the imperative properly represents as necessary. (Fund., 38)

The universal law in question is not an empirical fact apprehended by reason but an a priori unconditioned objective principle which every rational being would necessarily obey if reason had complete control over desires. It is a formal principle excluding any references to particular ends. There is only one such principle which is called 'the categorical imperative' (Fund., 38). According to Kant, "we must be able to will that the maxim of our action should be a universal law" (Fund., 41).

Formula Ia (law of Nature): Act as if the maxim of your action were to become through your will a universal law of nature. (Fund., 38)

This formula creates some difficulty, since Kant is understood to conceive of morality as free from determination by natural laws. It may, however, mean that when one chooses a morally valid maxim, one wills that such a choice will be made by all without exception as in a natural law. This will in fact happen, according to Kant, if our will were holly or purely rational (Fund, 31). Since humans are only partially rational, their objective will has to compete with their

subjective will so that moral maxims do not come naturally but as commands or imperatives (<u>Fund</u>., 31).

Silber argues that in this formula we can see that Kant did have in mind consequences of actions, although Kant is understood to claim that only actions done for the sake of duty, without regard to consequences, are morally good actions. This formula is the most important, according to Silber, since it makes clear the practical relevance of Kant's moral theory. He points out that critics such as John Stuart Mill, John Dewey (and others) argued that here Kant contradicts himself since he does bring into consideration future consequences. Silber argues that the formula shows that what Kant has in mind is that the moral agent think of the sort of a world which will result if moral or immoral actions become laws of nature, and that it is not Kant who contradicts himself but rather that the critics misunderstood or misinterpreted Kant's intentions.[21]

As Kant writes:

> Some actions are of such a character that their maxim cannot without contradiction be even <u>conceived</u> as a universal law of nature, far from it being possible that we should <u>will</u> that it <u>should</u> be so. In others, this intrinsic impossibility is not found, but still it is impossible to <u>will</u> that their maxims should be raised to the universality of a law of nature, since such a will would contradict itself. (<u>Fund</u>., 41)

---

[21] Silber, "The Procedural Formalism of Kant's Ethics", p.210.

Kant's teleological view of nature should be taken into account.

> We can consider human nature as if there were such a systematic harmony of ends in accordance with a law of nature; and then we ask whether any proposed maxim, if it were made a law of nature, would fit into such a systematic harmony. Some maxims would destroy such a systematic harmony, while others would merely fail to foster it.[22]

As Paton argues, Kant has in mind an ideal coherence of human purposes and wills as a test but not as the characteristic of moral action: by using the law of nature we can decide what we ought to do but it does not tell us about the spirit in which we ought to do it. This is why it is a test of moral action rather than its characteristic (namely, this formula is a subsidiary formula to the first one).[23]

Formula II (the End in Itself): So act as to use humanity, both in your own person and in the person of every other, always at the same time as an end, never simply as a means. (Fund., 46)

This formula enjoins us to consider and respect the rationality of the other person as an end in itself. Every rational being has a will (practical objective principles) but also private ends. Here Kant recognizes a plurality of ends (contents) each of which is to be considered as an independent

---

[22] Paton, The Categorical Imperative, p.150.

[23] Paton, The Categorical Imperative, pp.151-152.

end (because of its rationality and good will) with equal rights to pursue one's own private subjective ends (Fund., 47). Since human action is purposive in content, this formula is a limiting condition on all subjective ends (Fund., 48).

Formula III (Autonomy): **So act that your will can regard itself at the same time as making universal law through its maxims.** (Fund., 48-49)

In a certain sense, the previous formulae do indicate that the will is free. However, there it is understood as a necessity, as a duty against natural inclinations. Here, however, it implies that one's will freely legislates the moral law to oneself; that one obeys one's own legislation. The moral law comes from pure practical reason and not from an outside source and so the will is considered to be free in the positive sense of freedom (self-legislation).

This formula is considered to be logically different from the others, since autonomy is here both a precondition for the categorical imperative itself and also a formula of it, namely, an imperative. It is in fact the most important formula, according to Kant, since he considers the principle of autonomy to be the sole supreme principle of morality (Fund., 57). It is a principle in both senses of the word, namely it is an analytic (descriptive) principle which means it is a precondition and thus goes beyond the first formula (Universal Law) in that it is both the basis of morality, and also a moral principle. Williams claims that Kant's language

is here very confusing (since he uses the term 'principle' in two senses - analytical and ethical) but that it can be made clearer by understanding it in these two meanings.

> That the principle of autonomy in question is the sole principle [**necessary precondition**] of morals can be readily shown by mere analysis of the conceptions of morality. For by this analysis we find that its principle [**moral principle**] must be a categorical imperative, and that what this commands is neither more nor less than this very autonomy. (this is a quotation from Kant (Fund., 57) with square brackets added by Williams)[24]

Formula IIIa (Kingdom of Ends): **So act as if you were always through your maxims a law-making member in a universal kingdom of ends.** (Fund., 50)

Paton claims that this formula is incomplete and must be supplemented by the formula **All maxims which spring from your own making of laws ought to accord with a possible kingdom of ends as a kingdom of nature** (Fund., 50-51). "The use of the word 'kingdom' makes it clear that the laws in question are not to be considered in isolation but as part of a system of laws in both cases". Paton considers this formula to be the most comprehensive, since it does mention form (universal law) and matter (kingdom of ends). It correlates the laws of freedom with the laws of nature. Autonomy is here understood in the sense that the morally good person makes his or her own laws through his or her maxims, yet it has to accord and harmonize with the system of the kingdom of ends.

---

[24] Williams, The Concept of the Categorical Imperative, p.33.

28

It thus ensures objectivity and impartiality of self-legislation. Morality ensures that the rational being is participating in the legislation of universal laws which qualifies it to be a member of a kingdom of ends (Fund., 52).

> In Formula I we recognised that moral action has **one** form (the form of universal law). In Formula II we recognised that it has for its matter **many** objects - or ends. Finally in Formula IIIa we have reached the conception of **all** rational beings as ends in themselves united in one complete system under one universal law.[25]

The concept of a "kingdom of ends" as a counterpart to a kingdom of nature is an expression of Kant's teleological conception of nature. As Kant writes:

> Teleology considers nature as a kingdom of ends; ethics regards a possible kingdom of ends as a kingdom of nature. In the first case, the kingdom of ends is a theoretical idea, adopted to explain what actually is. In the latter it is a practical idea, adopted to bring out that which is not yet, but which can be realized by our conduct, namely, if it conforms to this idea. (Fund., 53, n.18)

Paton sees it in the following way:

> The system of a kingdom of ends governed by self-imposed, objective laws is the framework within which the private ends of ourselves and others ought to be realised. Such a framework by its apparent emptiness leaves room for the creativeness, in a sense the arbitrary creativeness, of human will.[26]

To sum up: When referring to or criticizing the 'categorical imperative' one must have in mind all of these formulae, if the criticism is to be valid. If one concentrates

---

[25] Paton, The Categorical Imperative, p.185.

[26] Paton, The Categorical Imperative, p.187.

on one formula only, one may lose sight of the procedural character of Kant's categorical imperative and its significance.[27]

The unity of Kant's moral theory can be grasped only within the framework of his conception of rational nature as part of human nature, and of his teleological conception of a system of laws of nature and that of freedom. Laws of nature are conditioned, laws of freedom are unconditioned. Freedom is not lawless; it would be absurd, according to Kant, to speak of freedom of the will without immutable laws because otherwise it will not be a free will but an arbitrary one. Free will and a will subject to moral laws are one and the same (Fund., 63-64).

2.   The Status of the Categorical Imperative: The Different Interpretations

In the previous section we were concerned with the meaning of the categorical imperative, with what we think Kant meant what we ought to do in order to act in a moral manner. In this section, I shall present some of the different interpretations of the 'categorical imperative' in terms of its status and its place in Kant's thinking.

---

[27] Silber, "The Procedural Formalism of Kant's Ethics", pp.197-236.

According to Williams, Kant presents two different arguments in the categorical imperative :

(a) The "ethical" strain, in which Kant seeks to set forth moral principles that are practically useful to men as guides to moral conduct; and

(b) the "analytic" strain, in which by analysis of the notion of moral obligation or duty, Kant seeks to set forth the basic or fundamental presupposition on which morality itself must rest; that is to say, the statement regarding the essential nature of morality which must be true if morality is, in fact, something real.[28]

There are different interpretations with regard to the nature of the categorial imperative, each of which seems to find some support in Kant's texts, and each of which in turn is taken to be Kant's main position.

These are: The 'categorical imperative' as a logical principle from which to derive substantive norms; the 'categorical imperative' as a moral criterion for everyday life; the 'categorical imperative' as a theoretical description of what moral consciousness is; the 'categorical imperative' as a procedure of moral reasoning.

The Categorical Imperative as a Logical Principle

This is the so-called "traditional" interpretation which takes the categorical imperative to be a precise or logical guide from which to derive substantive norms. This interpretation led critics to regard Kant's principle as

---

[28] Williams, The Concept of the Categorical Imperative, p.29.

either too rigorous and therefore too detached from everyday practice, or too formal and therefore empty unless some substantive norms are taken into account. By 'rigorism' it is meant that the form is independent of consequences. It is the motive for actions, the intentions, and not the consequences that matter. It is duty detached from inclinations.[29] By 'formalism' it is meant that its validity is independent of content so that any substantive content in it would contradict the form of the universal law.

That the categorical imperative is a logical criterion was understood to be so by Hegel, Schopenhauer, Mill, and others who took and still today take this to be the main position of Kant himself. This interpretation implies that Kant derives the categorical imperative directly from the concept of rationality itself. This, however, does not seem to be the case since Kant derives it from the notion of good will and duty. (Fund., Section 1) Also, this interpretation implies that one can derive the contents of the law from the mere form of law. Again, this does not seem to be what Kant intended, in particular, when one considers Kant's position in the second Critique.

The problem with the logical interpretation (logical contradiction) is that universal law is taken to mean universalization of a substantive law. Let's take Formula I,

---

[29] Williams, The Concept of the Categorical Imperative, p.38.

namely, "Act only on that maxim through which you can at the same time will that it should become a universal law" (Fund., 38). Hegel, for instance, criticizes this formula as being empty, which means that in principle it can be made to be compatible with any substantive system of laws.[30] Kant's examples are taken from everyday life of his society and it may seem as if he meant that it was the particular substantive norm (in the example, Fund., 39) that is in question.

Specific actions are carried out within a specific society with its specific institutions and norms. According to Hegel's interpretation, whenever I am in a certain situation, say, I have a loan to repay and I think that perhaps I will not repay it, then I should have to ask myself whether I would want (without self-contradiction) that the institution of property and its laws become universal law. Hegel then argues that there is no self-contradiction in not wanting the institution of property and its laws to become universal law. Yet, this does not seem to be Kant's point at all. Kant does not have in mind the institution of property (it is an example taken from everyday life in his society; "Kant himself has apologized for the shortcomings of his own examples"[31]); it is rather the will to choose the maxim not to repay a loan

---

[30] G.W.F. Hegel, Philosophy of Right, trans. T.M. Knox (Oxford: Clarendon, 1952), para. 135.

[31] Silber, "The Procedural Formalism of Kant's Ethics", p.227.

(within a system of such laws) that is at issue. By choosing not to repay the loan, I do two things: 1) The institution of loans will lose its meaning and its utility if not to repay loans becomes a law of nature (nobody will want to lend money, and my choice will then be self-defeating); 2) I in fact treat the person I owe the money to as a means to satisfy my subjective ends. I should, therefore, consider if I would want this kind of maxim (not to repay loans for subjective reasons) to become a law of nature (Fund., 39-40). Such an order, if it comes to pass, would contradict both my future needs for loans and formula II, which is that of always treating the other person as an end, not as a means (I myself may also become a victim and a means for other peoples' ends if they chose not to repay me for a loan). As Kant seems to understand it, no rational being would want such a self-defeating universalized maxim to become a law of nature. Every actual situation will be a contingent situation, but the maxim of not defeating a certain useful practice (whatever that may be) or not treating the other as a means, is independent of the contingent situation and as such is valid for all rational beings.

It could be argued that moral conflicts may occur, such as not to repay a loan in order to save a life. Moral conflicts do indeed happen, although Kant was mainly thinking

in terms of duty vs. inclinations.[32]   However, we could

perhaps interpret him in terms of his conception of a harmony

of purposes. No conflicts should occur because there is always

a priority of one purpose over the other, depending on the

relative value of the purpose (for example, the moral

principle to save human life may overrule the principle to

repay a loan, if this were necessary in this particular case

with its particular circumstance).

To sum up: The categorical imperative cannot serve as

a logical guide to substantive norms.   The interpretation

that **universal law** refers to universalizing the contingent

social order is not likely to be what Kant meant. It seems

more likely that he was thinking in terms of universalizing

those maxims which he considered morally valid regardless of

the social order.

The question of whether the institution of property is

just or not, or whether it is necessary or not, is an

important issue, but Kant was not critically analyzing

society, nor did he intend to discover a new morality, as he

himself writes:

> A critic who wished to say something against this work
> really did better than he intended when he said that
> there was no new principle of morality in it but only
> a new formula. Who would want to introduce a new
> principle of morality and, as it were, be its inventor,
> as if the world has hitherto been ignorant of what duty
> is or had been thoroughly wrong about it? Those who know
> what a formula means to a mathematician, in determining

---

[32] Roger J. Sullivan, <u>Immanuel Kant's Moral Theory</u>
(Cambridge : Cambridge University Press, 1989), pp.72-75.

what is to be done in solving a problem without letting
him go astray, will not regard a formula which will do
this for all duties as something insignificant and
unnecessary.[33]

Kant's intention was to articulate the sense of duty,
i.e. the moral consciousness of ordinary every day people, in
terms of pure practical reason, and his effort should be
assessed in the light of this intention. This brings us to a
second interpretation of Kant.

The Categorical Imperative as a Moral Criterion

This interpretation is advanced by Paton. Paton's
interpretation is that the 'categorical imperative' is both an
analytic principle in the sense that it states what principle
is employed in the workings of pure practical reason and also
a moral principle in the sense that it is a criterion or a
test of moral actions. In other words, the 'categorical
imperative' states what the nature of morality is, but at the
same time constitutes the criterion of this morality. It is
both a command and a formula of that command, namely, an
imperative.[34]

Kant's formalism is interpreted by Paton to be not so
rigid as is implied by the traditional interpretation. It is

---

[33] Immanuel Kant, <u>Critique of Practical Reason</u>, trans.
Lewis White Beck (Indianapolis, Bobbs-Merril,,1956), p.8, n.5.
Later references to pages in this book will appear in the text
in parentheses (Second <u>Critique</u>).

[34] Paton, <u>The Categorical Imperative</u>, pp.140-142.

rather that

> the motive of duty must be present at the same time as
> inclinations and must be the determining factor, if our
> action is to be good. It is therefore a distortion of his
> view to say that for him an action cannot be good if
> inclination is present at the same time as the motive of
> duty.[35]

Kant's dichotomy between duty and inclination is a theoretical exercise; it does not mean that Kant ignored the fact that all actions were done for the sake of a certain material end. It is rather that for Kant ends and consequences could not constitute the determining ground of moral action. All Kant is saying is that its distinctively moral value does not depend on the results sought or the results attained. Moral actions occur within a context of activities carried out in everyday life. Pure practical reason spontaneously orders and regulates, in accordance with its own principles, actions which are based on interests. Paton considers the 'categorical imperative' to be a statement of the principle employed in the workings of practical reason, yet he believes that the 'categorical imperative' as a moral criterion has priority for Kant over its descriptive part.

> The formula of the categorical imperative, like any
> other 'expression of essence', must prescribe what we may
> call a logical criterion, whether sound or unsound. But
> because it expresses the essence of the command which is
> supposed to constitute the relation of the moral law to
> an imperfect will, it may be said to prescribe a moral

---

[35] Paton, The Categorical Imperative, p.49.

criterion as well.[36]

The categorical imperative should be regarded as a practically useful principle for a moral agent to adopt, an attitude which will enable him or her to live a moral life.[37] The value of the principle can be demonstrated only in actual situations and the value of its form lies in that it states the nature of moral actions and the principle employed by practical reason in its workings. The value lies in clarifying the nature of moral actions.

To sum up: To say that the moral law must be valid for all rational beings means that it can not be determined by desire; it must be impartial. However, this is not to say that it has to be detached from everyday life, since the very conception of it arises within a context of human activities. It is not a question of wishing to universalize a substantive norm, rather it is the significance of adopting a maxim as a rule which is of moral import. This depends on the will of the agent, not on a substantive norm. According to Paton, the principle is not to be regarded as outside of action but as the principle of the action, embodied in the action, which thus makes it a morally good action.

---

[36] Williams, The Concept of the Categorical Imperative, p.70;
    Herbert J.Paton, "The Aim and Structure of Kant's Grundlegung", Philosophical Quarterly, VIII (1958): 123.

[37] Paton, The Categorical Imperative, p.156.

The Categorical Imperative as a Theoretical Descriptⁱon of Moral Consciousness

This interpretation is advanced by Duncan.[38] He rejects the interpretation that the categorical imperative is an exposition of Kant's own ethics. He claims that Kant's main objective was to carry out a critical examination of the workings of pure practical reason along the lines of the first Critique. He calls Kant's main position "critical", namely, that the 'categorical imperative' is a "supreme principle of morality" in the sense that it is the principle which expresses the functioning of the moral motive.[39] It describes the form, the capacity of a maxim to be universalized. Duncan does not deny that Kant meant it also to be a moral ("prescriptive") criterion but he argues that Kant erred in that he jumped to the conclusion that it could be used as a moral criterion. This carelessness, Duncan argues, on the part of Kant himself is the root of all the misinterpretations of Kant's main position. Duncan also argues that the imperative form (which he says is confusing) was meant to describe the categorial ought.[40]

Although Duncan and Paton seem to be in disagreement, their positions are not in fact really in opposition,

---

[38] A.R.C. Duncan, Practical Reason and Morality (Edinburgh: Nelson, 1957), pp.30+

[39] Duncan, p.71.

[40] Duncan, p.107. Williams, The Concept of the Categorical Imperative, p.88.

according to Williams. Their disagreement turns out to be a result of the unclear use of the terms 'criterion' and 'prescriptive'[41]. Both agree that the 'categorical imperative' expresses the nature of a morally good action. They seem to differ with regard to the "prescriptive" function of it. Duncan rejects the traditional interpretation, but he does not in fact object to Paton's interpretation of it as a moral criterion or a moral ideal. What he does reject is that this should be regarded as the main position of Kant. The disagreement, then, is on the emphasis. For Paton, the moral ideal is Kant's main position, for Duncan it is the statement of the nature of moral actions that is the main position.

If the categorical imperative is indeed a descriptive statement of moral consciousness, then one may ask if we are not faced here with a problem of deriving an "ought" from an "is". However, there might be a confusion here since the description is of an "ought", of a moral consciousness of which we become self-conscious in moral experience. Moral consciousness is precisely this sense of "ought" that Kant intended to describe. However, what he does describe seems to be his preferred morality which he takes to be objectively valid. In this respect, it seems to me that Paton is right in considering the 'categorical imperative' as a moral criterion to be Kant's main position.

---

[41] Williams, <u>The Concept of the Categorical Imperative</u>, p.94.

Kant's descriptive analysis is based on his differentiation between theoretical and practical reason. Theoretical reason is concerned with understanding natural phenomena, with imposing an order (in accord with a priori principles of understanding and of logic) on our sensible intuitions. Practical reason, on the other hand, regulates our actions with its a priori practical principles; it is active compared to the passivity of theoretical reason in that it guides and directs the will.

Theoretical reason tries to understand states of affairs in the world, while practical reason is concerned with how to produce a desirable state of affairs in the world. To do this there are two kinds of **ought** that practical reason has to consider: 1) How to produce a state of affairs; this is the technical, or prudential (in Kant's terminology "hypothetical") **ought** for which theoretical knowledge is required (subject vs. object relation); 2) how to produce this state of affairs in accordance with moral principles; this is the necessary moral **ought** in which account is taken of interpersonal relations and the social world one lives in. For Kant there is only **one** reason which functions in two different manners, theoretical and practical. However,

> in the combination of pure speculative with pure practical reason in one cognition, the latter has the primacy, provided that this combination is not contingent and arbitrary but a priori, based on reason itself and thus necessary. Without this subordination, a conflict of reason with itself would arise... because every interest is ultimately practical, even that of speculative reason being only conditional and reaching

perfection only in practical use. (second _Critique_, 126)

Pure practical reason enables us to be autonomous, self-determining and self-creative. We thus can determine our ends in a moral manner which is intrinsically and unconditionally good. Theoretical reason and knowledge is conditional and helps achieve human material ends but it is the morally good will (pure practical reason) that is unconditionally good and gives human beings an intrinsic value. This view of the primacy of pure practical reason can also be understood as safeguarding against the fact that theoretical knowledge can be used for immoral purposes.

The Categorical Imperative as a Formal Procedure of Moral Reasoning

This interpretation is advanced by Silber. He argues that the act of judgement is important in the application of Kant's moral theory to practice. A moral theory should provide principles for guiding moral judgements. The 'categorical imperative' itself is a principle which specifies the procedure of judgements in the act of moral schematism.[42] The correct application is in the procedure. The moral norm or guide is never given in terms of some substantive goal but is a statement of a procedure that, if followed, can lead to an achievement of the goal.

---

[42] Silber, "The Procedural Formalism of Kant's Ethics", pp.198-199.

In epistemology we start with intuition and move to concepts; in moral philosophy we start with the concept and move towards intuition. The different formulae of the 'categorical imperative' start from the abstract and move towards intuition.

Kant specifies the procedure in the following way: he says that all maxims must have

> 1. A **form**, consisting in universality; and in this view the formula of the moral imperative is expressed thus, that the maxims must be so chosen as if they were to serve as universal laws;
> 2. A **matter**, namely, an end, and here the formula says that the rational being, as it is an end by its own nature and therefore an end in itself, must in every maxim serve as the condition limiting all merely relative and arbitrary ends;
> 3. A **complete characterization** of all maxims by means of that formula, namely, that all maxims ought, by their own legislation, to harmonize with a possible kingdom of ends as with a kingdom of nature.

> There is a progression here in the order of the categories of **unity** of the form of the will (its universality), **plurality** of the matter (the objects, that is, the ends), and **totality** of the system of these... If we... wish to gain an **entrance** for the moral law, it is very useful to bring one and the same action under the three specified conceptions, and thereby as far as possible to bring it nearer to intuition. (Fund., 53)

The moral agent has to try to clarify to him/herself what maxims to choose. In this, pure practical reason enables the agent to know what ought to be done in particular situations. Kant has to be understood as regarding even the commonest of people as being able to know what ought to be done, not automatically (this will happen only to a purely rational being) but by a procedure of reasoning and judgement that is "sharpened by experience" (Fund., 5).

Reasoning is a process of arriving at objectivity, for Kant, and it thus helps to overcome subjective desires. As Kant sees it, the human being has the possibility of acting in a moral manner, if he or she understand themselves as free and at the same time consider other beings as free and equal with themselves. It is this standpoint (or attitude) which Kant believes to be possible for human beings, and it is precisely this standpoint that means that the human being is autonomous, and this is what ensures that human beings are more than animals.

As Silber points out, Kant's rules for reasoning in general apply to moral reasoning as well, these are:

> 1) To think for oneself; 2) In communication with men to imagine (sich denken) oneself in place of every other person; 3) always think in agreement with one's self. The first principle is negative... that of freedom from coercion; the second is positive, that of liberals who accommodate themselves to the concept of other thinkers; the third is the consistent (logical) mode of thinking.[43]

To sum up: That the moral task is to arrive at a judgement based on these rules of reasoning is an expression of Kant's conception of human autonomy and rationality. These rules are formal and have to be applied in each particular case, which will inevitably include substantive norms and a content of sensibility. As Silber points out, individual moral responsibility, for Kant, can not be delegated, therefore the application of ethics can not be left to experts.

---

[43] Silber, "The Procedural Formalism of Kant's Ethics", p.202 (quoted from Kant's Anthropology)

> The problem of relating theory to practice is one that confronts each rational being...
> The only way the will determines what does in fact constitute universal acts of moral volition is by going through a variety of procedures designed (in the categorical imperative) to carry the agent beyond merely subjective conditions of volition.[44]

It is the agent's conscientiousness and faithfulness in carrying out the maxims of moral judgement in practice which counts as moral.

3. The Transcendental Deduction of the Categorical Imperative

The categorical imperative is, for Kant, the supreme moral principle which validates the selection of moral maxims. Kant considers it to be an a priori synthetic moral proposition that should itself be validated by means of a transcendental deduction (Fund., 64).

As it is with other subjects in Kant, here too there is a disagreement with regard to Kant's treatment of the validity of the categorical imperative in the Fundamental Principles of the Metaphysics of Morals and in the Critique of Practical Reason. In particular, there has been a lack of agreement about the difference between the two treatments:

> In the first work Kant seems to desire and develop a theoretical argument for freedom in a sense which is absolute and from which the objective validity of the moral law is to be deduced. In the second work, however, Kant appears directly to reverse himself and to replace this project of a strict

---

[44] Silber, "The Proceduralism of Kant's Ethics", pp.234-235.

deduction [logically sound] with the idea that the moral law (i.e. its validity, not its entire exact formulation and implications) is simply given as an 'a priori fact of reason' (from which alone freedom can then be inferred).[45]

According to Karl Ameriks, most commentators seem to downplay this difference but he himself considers that there is a reversal by Kant in his validation of the categorical imperative. He thinks that in the Fundamental Principles... Kant is dogmatic in that he tries to establish theoretically the reality of freedom (despite Kant's claim in the first Critique that this is impossible) and also uses his doctrine of the two "worlds" ("proper self") to support his deduction, while in the second Critique he abandons this method.[46] Ameriks also claims that Kant himself recognized his ethics to be more dogmatic than his epistemology.

According to Henry Allison, there seems to be "no agreement as to whether the deduction is of the moral law, the categorical imperative, freedom, all three; or even whether there is properly a deduction at all".[47]

Kant's specific intention in analyzing moral consciousness was to separate it from empirical contingent

---

[45] Karl Ameriks, "Kant's Deduction of Freedom and Morality", Journal of the History of Philosophy, XIX (1981): 53.

[46] Ameriks, "Kant's Deduction of Freedom and Morality", pp.77-79.

[47] Henry Allison, "Morality and Freedom : Kant's Reciprocity Thesis", Philosophical Review, 95 (July, 1986): 393.

conditions. He was interested in the a priori elements that are involved in moral experience. He tried to go from the conditioned to the unconditioned, namely, to the first principles. The validity of these principles, according to Kant, lies in their a priori status which makes them objectively valid, and in their not being conditioned by a particular experience but rather in their being the condition of moral experience itself. He is thus trying to look for the ultimate conditions and presuppositions of accepted moral beliefs by a regression or an analysis. The ultimate conditions must be established independently of experience (because they are a priori synthetic) by a transcendental deduction, if we are to avoid circularity (Fund., 64). The problem is: How can we justify the proposition that a rational agent ought to act in accordance with the principle of autonomy, and whether an absolutely good will must also act in accordance with the principle of autonomy? From the concept of a rational will we can not arrive at an obligation to will in a certain way; we connect this obligation to a rational will but it is not contained in it (Fund., 36-37, n.6). In order to make it valid for all rational beings Kant thinks that we need a third term which is necessary to both the rational will and the obligation, otherwise morality is either a prejudice or an illusion. Kant considers positive freedom to be that third term (Fund., 64). In the following, I shall examine Kant's deduction in the Fundamental Principles of the

Metaphysics of Morals, and then the deduction in the second
Critique.

The Deduction in the Fundamental Principles of the Metaphysics of Morals

In section three of his Fundamental Principles...
Kant tries to establish the objective validity (the
possibility) of the moral law from the concept of freedom of
the will (Fund., 64). In contrast to the first Critique where
a priori synthetic judgements are known to be already possible
and have only to be shown how they are possible, here it is
the very possibility of the moral law that has first to be
established. The deduction is hypothetical: if positive
freedom of the will is real, then the categorical imperative
is real, too.

According to Paton, the deduction is questionable:
(a) We seem to be more sure of the moral law than of freedom;
(b) it is inconceivable that the moral law can be justified by
anything except by itself.[48]

In the Critique of Pure Reason (third antinomy) Kant
established that transcendental freedom is thinkable (which
means it is not in contradiction with the concept of natural
necessity), although we can have no theoretical knowledge of
it since we do not encounter it in experience (it is an Idea
of reason). Transcendental freedom is an absolute spontaneity

---

[48] Paton, The Categorical Imperative, p.204.

or complete independence from any determination by antecedent conditions (negative freedom). This is the causality of freedom, the initiation of a causal chain of events. This concept is a theoretical concept and in itself is empty, but it must be assumed in order to justify positive freedom which, for Kant, is the necessary precondition of the moral law. Transcendental freedom means an agent is always autonomous in that he or she may select any maxim whether motivated by natural causes or against such natural causes. As Kant sees it, it is the concept of positive freedom which is the key to explaining the principle of autonomy (Fund., 63).

> The question, then, How a categorical imperative is possible can be answered to this extent that we can assign the only hypothesis on which it is possible, namely, the idea of freedom; and we can also discern the necessity of this hypothesis, and this is sufficient for the practical exercise of reason, that is, for the conviction of the validity of this imperative, and hence of the moral law; but how this hypothesis itself is possible can never be discerned by any human reason. On the hypothesis, however, that the will of an intelligence is free, its autonomy, as the essential formal condition of its determination, is a necessary consequence. (Fund., 78)

The deduction fails (according to Paton) because although morality may imply freedom of the will, it does not necessarily follow that the categorical imperative is the only possible or necessary choice to be adopted by a free will (assuming there is such a will). Although there is a necessary connection, the relation is not reversible. Freedom of the will enables the categorical imperative but it also enables

immoral actions.[49] Autonomy, it seems, is neutral. However, as Allison argues, it can not really be considered neutral since it is precisely the moral law that gives autonomy any standing.[50] Morality, in fact, can not be deduced from a non-moral source such as freedom. Kant is aware of it. It is impossible to explain freedom, he argues, as it is impossible to explain why the human being takes an interest in the moral law. It is the validity of the moral law that causes interest in us; it is not the interest in it that makes it valid.

> This only is certain, that it is not because it interests us that it has validity for us (for that would be heteronomy and dependence of practical reason on sensibility, namely, on a feeling as its principle, in which case it could never give moral laws) but that it interests us because it is valid for us as men, inasmuch as it had its source in our will as intelligence, in other words in our proper self, and what belongs to mere appearance is necessarily subordinated by reason to the nature of the thing in itself. (Fund., 78)

Kant resorts to his doctrine of the two "worlds" or viewpoints — the noumenal self - in order to explain this interest in moral law, and in order to escape the circle in his deduction (Fund., 68- 69).

> It must be freely admitted that there is a sort of circle here from which it seems impossible to escape. In the order of efficient causes we assume ourselves free, in order that in the order of ends we may conceive ourselves as subject to moral laws; and we afterwards conceive ourselves as subject to these laws because we have attributed to ourselves freedom of will; for freedom and self-legislation of will are both autonomy, and therefore

---

[49] Paton, The Categorical Imperative, p.244-245.

[50] Henry Allison, Kant's Theory of Freedom (Cambridge : Cambridge University Press), p.98.

are reciprocal conceptions and for this very reason one
must not be used to explain the other or give the reason
of it, but at most only for logical purposes to reduce
apparently different notions of the same object to one
single concept (as we reduce different fractions of the
same value to the lowest terms). (Fund., 67)

By means of his doctrine of the two "points of view"
Kant tries to escape the circle and resolve the
inexplicability of both freedom and morality. How does he do
it? Opinions differ. What he seems to say is that when we
reflect on ourselves we cannot avoid thinking of ourselves as
being both natural and moral beings. In his last paragraph of
the Fundamental Principles... Kant writes:

It is, therefore, no fault in our deduction of the
supreme principle of morality, but an objection that
should be made to human reason in general, that it cannot
enable us to conceive the absolute necessity of an
unconditional practical law (such as the categorical
imperative must be). It can not be blamed for refusing
to explain this necessity by a condition, that is to
say, by means of some interest assumed as a basis, since
the law would then cease to be a moral law. (Fund., 80)

There are two definitions of the will in the
Fundamental Principles...:

1) Will as a power of a rational being to act in
accordance with principles (Fund., 30);

2) "Will as a kind of causality belonging to living
beings so far as they are rational and freedom would be this
property of such causality that it can be efficient
independently on foreign causes determining it; just as
physical necessity is the property that the causality of all
irrational beings has of being determined to activity by the
influence of foreign causes" (Fund., 63).

The power to act means the power to produce effects, and the power to act in accordance with a conception of a law means that the will must be self-conscious with regard to its causality of actions.

A free will means that it is not to be determined by external causes. Yet, the concept of causality implies the concept of law whether it is determined by nature or whether it is free. This may raise the question of whether one can speak of a free will if it is seen to be subject to laws of freedom (for which Kant takes natural law as a model). How does Kant pass from natural law to laws of freedom? Paton argues that there are several assumptions here: (a) Free will has its special laws; (b) it is self-imposed; (c) a self-imposed law is a form of law as such. However, the principle of autonomy does not assert that there is a necessary connection between causes and effects. The grounds for taking natural law as a model are, according to Paton, inadequate but he thinks that Kant is right in arguing that a lawless will is an absurdity; if 'rational will' means to act in accordance with a conception of a law, then it can not be arbitrary. Laws of freedom are self-imposed: the causality of a free will must take place in accordance with self-imposed laws which are of a different kind from natural laws, but if they are to hold for all rational beings, then they must have a law-like status. However, Kant does not prove that there is such a free

will, only how it should be conceived.[51]

There is also the question of whether only a morally good will is free and a morally bad will is determined. If this is so, then how are we to understand the responsibility of the individual for morally wrong actions? Kant distinguishes between a will under moral law and a will which always obeys a moral law. To be under moral law does not mean that the will necessarily obeys it. Transcendental freedom means a capacity of the will to follow its own self-imposed laws which requires independence from sensible grounds. Although Kant believes that human action is determined by sensuous motives, it is nevertheless according to him, a result of a free choice since, as he claims, human beings act in accordance with a conception of a law. Desires may influence the decision but the choice is free and therefore one is responsible for whatever one does. For Kant, reason "must regard itself as the author of its principles independent of foreign influences" (Fund., 65) and as such must regard itself as free. We are free, according to Kant, to act heteronomously but in so acting we are misusing our freedom; misuse of freedom is itself an act of freedom.

To sum up: Kant tries to establish the reality of the moral law on the basis of the Idea of freedom; Kant cannot prove that a rational agent is in fact free. However, an agent who acts with the Idea of freedom "is bound by the laws that

---

[51] Paton, The Categorical Imperative, p.211.

would oblige a being who was actually free" (<u>Fund.</u>, 65 n.2).
Such a necessary presupposition would for Kant be enough to
justify the moral law but the deduction in section three of
the <u>Fundamental Principles</u>... does not achieve it, since Kant
tries to establish the reality of the moral law on the basis
of a presupposition of positive freedom (autonomy) that can
not be proven (or does not make sense) without presupposing
the moral law itself.

The Deduction in the Critique of Practical Reason

Self-consciousness is a central concept in Kant's
thinking; it means for him understanding the principles
manifested in the activity of reason as such.[52] To the
question: "How is the consciousness of the moral law
possible?" Kant answers that when we abstract the necessary
principles from empirical experience we can then regard them
as the a priori conditions of this activity (second <u>Critique</u>,
29). For Kant, the ultimate precondition of practical reason
is the principle of acting for the sake of universal law as
such, which is the unconditional objective principle of
morality and is possible because of the freedom of the will to
be so determined (second <u>Critique</u>, 30).

In the second <u>Critique</u> Kant takes the reality of the
categorical imperative to be a "fact of pure reason", namely,
that the consciousness of this law "forces itself upon us as

---

[52] Paton, <u>The Categorical Imperative</u>, p.220.

a synthetic proposition a priori based on no pure or empirical intuition" (Second *Critique*, 31). Kant starts from the concepts of good will and duty in the *Fundamental Principles*... but there he is concerned to prove the reality of the categorical imperative. In the second *Critique* there is no question about the reality of the moral law (and therefore no need for a deduction) since it is "given as a fact of pure reason" of which we are **a priori** conscious:

> The consciousness of this fundamental law may be called a fact of reason, since one cannot ferret it out from antecedent data of reason, such as the consciousness of freedom (for this is not antecedently given), and since it forces itself upon us as a synthetic proposition a priori based on no pure or empirical intuition. (Second *Critique*, 31)

This implies a will which is able to act in accordance with this law over and against subjective inclinations. The moral consciousness is manifested in the moral experience of everyday ordinary people, and he is trying to describe this fact in terms of its presuppositions (justifications) such as freedom, autonomy of the will, and in terms of the moral reasoning in the categorical imperative.

We recognize ourselves as free through our being conscious of the binding moral law. He writes:

> To avoid having anyone imagine that there is an inconsistency when I say that freedom is the condition of the moral law and later assert that the moral law is the only condition under which freedom can be known, I will only remind the reader that, though freedom is certainly the ratio essendi of the moral law, the latter is the ratio cognoscendi of freedom. For have not the moral law already been distinctly thought in our reason, we

would never have been justified in assuming
anything like freedom, even though it is not self-
contradictory. But if there were no freedom, the
moral law would never have been encountered in us.
(Second _Critique_, 4 n.1)

The categorical imperative is for Kant the supreme
principle of the moral law whose objective validity is real;
with this the principle of autonomy (Formula III) is also
recognized as a fact. The reality of the moral law implies
that rationality also legislates and that the will is free to
act or not to act in accordance with this law:

The moral law expresses nothing else than the autonomy of
the pure practical reason; i.e., freedom. This
autonomy or freedom is itself the formal condition of all
maxims, under which alone they can all agree with the
supreme practical law. (Second _Critique_, 33-34)

The consciousness of the moral law, the ought, arises
when confronted with the necessity to choose a morally valid
maxim. This ought or command arises spontaneously in pure
practical reason as an Idea which motivates the will, in
addition to the will being affected by desire. It is to be
noted, again, that Kant does not mean to say that the moral
law is apprehended or intuited. The moral law is not an
empirical fact but is originally legislated by pure practical
reason. It is a moral cognition:

In order to regard this law without any
misinterpretation as given, one must note that it
is not an empirical fact but the sole fact of pure
reason, which by it proclaims itself as originating
law. (Second _Critique_, 31)

According to Williams, "consciousness of the moral law" may be ambiguous and therefore interpreted differently: "(a) The appreciation of the objective validity of the purely formal principle of morality, i.e. Formula I; or (b) the consciousness of the manifestation of this principle in the context of particular moral situations". The two meanings may be interrelated, but the acceptance of the first meaning as having primacy over the second entails the traditional interpretation (logical principle) and all the difficulties involved with this interpretation, since it means to "discover" the acts that "fit" this principle. Williams considers the second meaning to be primary for Kant, since it seems to be more compatible with Kant's thinking in general.[53]

As already indicated, Kant distinguishes between two different kinds of ends:

1) An end which is based on a desire (in this case the concern will be whether such an end is physically possible), and

2) an end which is motivated by pure practical reason and which is concerned with whether the achievement of the desired end, or the means to achieve this end, is morally right or wrong (this would not depend on whether the end is or is not physically possible).

---

[53] Williams, The Concept of the Categorical Imperative, pp.103-104.

He calls the second kind of ends an "object of pure practical reason" (second _Critique_, 59). An object of pure practical reason is either good (an object of desire) or **evil** (an object of aversion) (second _Critique_, 60). Such "objects" are not intuited but presented in the consciousness as "cognitions", namely, as what ought to be done in particular situations:

> The practical a priori principles in relation to the supreme principle of freedom immediately become cognitions, not needing to wait for intuitions in order to acquire meaning. This occurs for the noteworthy reason that they themselves produce the reality of that to which they refer (the intention of the will) – an achievement which is in no way the business of theoretical reason. (Second _Critique_, 68)

It is in this sense of being cognized as an **ought** and in being the determining principle of the action that Kant considers a principle to be a morally valid principle.

Kant claims that

> what is required in accordance with the principle of autonomy of choice is easily and without hesitation seen by the commonest intelligence; what is to be done under the presupposition of its heteronomy is hard to see and requires knowledge of the world. That is to say, what duty is, is plain of itself to everyone, but what is to bring true, lasting advantage to our whole existence is veiled in impenetrable obscurity, and much prudence is required to adapt the practical rule based upon it even tolerably to ends of life by making suitable exceptions to it. (second _Critique_, 38)

According to Williams, we should not understand Kant as saying that there is no need for moral guidance, since we spontaneously know what ought to be done; rather, what Kant means is that pure practical reason has the capacity to make

known what ought to be done.[54] Williams also points out that even though there is a certain "subjectivity" in Kant's doctrine (conscientiousness) it is a "subjectivity" that seeks the "objectivity" of rationality. Although this may imply that there can be differences, or disagreement in judgements, such disagreements may be resolved by becoming more acquainted with the contingent conditions involved. As Williams sees it, the formalism of the categorical imperative makes it a practically useful guide precisely because it is impartial and affords a point of view of an observer.[55]

To sum up: Kant is not suggesting a new kind of morality but as he claims, he is trying to give a new formulation to the moral consciousness involved in everyday moral experience. He claims that ordinary people recognize the validity of the categorical imperative even when they do not follow it (Fund., 71; second Critique, 32) This is the sense of duty, the consciousness that we should have or could have followed the moral law. The moral law, according to Kant, is part of our consciousness, of our rationality; in acting, we cannot understand ourselves other than being free or being rational, or as being moral or immoral; we cannot deny our rationality without using rational arguments, nor can we deny our consciousness of the moral law. It is, according to Kant,

---

[54] Williams, The Concept of the Categorical Imperative p.114.

[55] Williams, The Concept of the Categorical Imperative, pp.125-127.

the inevitable nature of being a rational being; if we must accept these principles, namely, if we cannot think of ourselves otherwise, then such thinking cannot be avoided and may be considered to be valid.

(3) On the Metaphysical Nature of the Transcendental Deduction

Kant tries to defend his ethical theory with the idea of a noumenal self which he believes is the only explanation to the moral phenomenon, in that our "proper self" follows its own laws of freedom. There are different views with regard to Kant's dogmatism in his moral doctrine. As we have seen above (see p.43), Karl Ameriks believes that Kant is dogmatic in his ethics. Paton, however, does not believe this dogmatism is to be taken too strictly:

> Kant's ethics - in spite perhaps of occasional lapses - is not based on his metaphysics: it would be truer to say that his metaphysics, so far as we take this to be concerned with a supersensible reality, is based primarily on his ethics. Whatever confusion or error there may be on this topic in the Groundwork [Fundamental Principles] is to a great extent cleared away by the Critique of Practical Reason.[56]

Another claim, by Otfried Hoffe, is that Kant's ethics is not largely transcendental except perhaps his concept of autonomy:

> Kantian justification of ethics is not thoroughly transcendental. Quite the contrary, only a relatively small part, namely, the reduction of the moral law to the

---

[56] Paton, The Categorical Imperative, p.255.

autonomy of the will, possesses a transcendental character. A larger part, including the conceptual determination of the moral law, which leads to form as the sole determining ground of the will, taken by itself, has the character of a problematic of conceptual analysis. At the same time, this part is absolutely necessary for the preparation of transcendental reflection; for freedom as autonomy is the determinate negation of heteronomy, of a concept that is formed in the analytical-problematical part. Over and above this, the theoretical portion characterized by the label 'fact of reason' possesses a still looser connection to a transcendental ethics.[57]

As Hoffe points out, 'transcendental' in Kant means that it is 1) valid before all experience, and 2) that it is necessary for the constitution of experience itself.[58] The a priori shapes experience and thus cannot be understood as being disconnected from experience.

As Paton argues, the form of the categorical imperative is a priori not in the sense that it is prior to experience, namely, that moral judgements can be made before experience. This, he says, cannot be attributed to Kant. Rather, he argues that for Kant "the a priori element must be present in moral judgments... In separating out the a priori element in moral judgements we are determining the condition, the sine qua non, of the validity of moral judgements".[59] The

---

[57] Otfried Hoffe, "Kantian Skepticism Towards Transcendental Ethics", in Seyla Benhabib and Fred Dal. mayr (eds.), Communicative Ethics Controversy (Cambridge, Mass: MIT Press, 1990), p.196.

[58] Hoffe, "Kantian Skepticism Towards Transcendental Ethics", p.193.

[59] Paton, The Categorical Imperative, p.26.

a priori, in Paton's interpretation, does not lie outside the action, "on the contrary, it is the principle of the action, the formal principle which is embodied in the action and in virtue of which the action is good".[60]

According to Bruce Aune,

> if we reject Kant's conception of the intelligible world as an actually existing but largely unknowable domain of rational intelligences, we can preserve a large portion of his moral theory by conceiving of rational beings as ideal beings rather than transcendentally real ones ... Of course, the question can arise, 'Why should we, as imperfectly rational beings, value rational nature in this way?' If we have trouble answering this question, we can take comfort from the fact that Kant himself cannot answer it. As he says 'The subjective impossibility of explaining freedom of the will is the same as the impossibility of finding out and making comprehensible what interest man can take in moral laws; and yet he does take such an interest'.[61]

To sum up: Kant's categorical imperative can be interpreted as a moral principle (or a model for moral reasoning) that guides our actions (what "ought" to be), and at the same time also as a theoretical description of our moral consciousness (what "is").

The structure of the categorical imperative includes form (universality), matter (plurality of ends-in-themselves) and totality of all maxims (kingdom-of-ends). This structure reflects Kant's teleological worldview of harmony of purposes (Kingdom-of-Ends) as well as his conception of the autonomy of the human being's consciousness and its intrinsic value.

---

[60] Paton, The Categorical Imperative, pp.133-134.

[61] Aune, Kant's Moral Theory, pp.102-103.

Kant's formalism has been interpreted to be either logical and empty, and therefore impractical (Hegel), or as an impartial attitude (objectivity of the will, the moral point of view) vs. the subjective (self-interested) will. The plurality of the ends-in-themselves (matter, content) calls for a reciprocal interpersonal attitude (the standpoint of the others, which is also the moral point of view). The Kingdom-of-Ends (the totality of all maxims) means an objective moral order (the achievement of total harmony or reconciliation).

Kant's moral values of autonomy, equality and reciprocity in interpersonal relations, as well as the ideal of a kingdom-of-ends, i.e. a well ordered society, are still "real" moral values today, although Kant's worldview and his methods of justification may not be acceptable today . The question is whether these moral values can be rationally justified at all and in particular, the question is how to understand Kant's categorical imperative with a different conception of rationality and society, and in fact, from a different conception of "reality".

Kant claimed that he was trying to give a new formulation to the idea of moral law; Habermas claims that he attempts to reformulate Kant's formulation. His Discourse ethics is an attempt to find a new reformulation for the same moral values with a differently conceived rationality and society; in other words, from the point of view of a critical social theory.

CHAPTER II

HABERMAS' THEORETICAL CONCEPTION OF THE

STRUCTURE OF MORAL CONSCIOUSNESS

As mentioned earlier (see pp.9-10 above), Habermas'
believes that philosophy should collaborate with the sciences
(natural or social), rather than compete with them (MCCA, 1-
20). As we saw (see p.6 above), his approach is normative in
that his concern is with how we acquire knowledge (in
particular in the social sciences) and how we use it
(communicative action). He relies on theories of
developmental-logic in cognition and moral development, such
as Jean Piaget and Lawrence Kohlberg[62], and claims that his
discourse ethics is more compatible with these theories than
other competing moral theories (MCCA, 117-120).

In this chapter, I will deal with Habermas theoretical
conception of the development of moral consciousness and its
relation to his discourse ethics.

---

[62] Lawrence Kohlberg, "From Is to Ought: How to Commit
the Naturalistic Fallacy and Get Away with It in the Study of
Moral Development", in Theodore Mitchel (ed.), Cognitive
Development and Epistemology (New York: Academic Press, 1971),
pp.151-235.
    Laurence Kohlberg, Essays on Moral Development (San
Francisco: Harper and Row, 1981), 2vs.
    Jean Piaget, The Principles of Genetic Epistemology
(New York, 1970).

64

"Reconstructive sciences" concern themselves, as Habermas does, with uncovering the basic formal structures involved in the intuitive or pre-theoretical knowledge of speaking and acting subjects that are universal within the species, regardless of cultural content. Speaking and acting subjects, in general, seem to know how to perform various acts without necessarily being able to give an account of the concepts and rules which underlie their performance.

> Common to all these disciplines is the goal of providing an account of the pretheoretical knowledge and the intuitive command of rule systems that underlie the production and evaluation of such symbolic expressions and achievement as correct inferences; good arguments; accurate descriptions, explanations, and predications; grammatically correct sentences; successful speech acts; effective instrumental action; appropriate evaluations; authentic self-presentations; etc. (MCCA, 31)

Reconstructive sciences differ from "normal" science in that science deals with perceptible reality (observations) while reconstructive sciences deal with a symbolically structured reality (understanding). Reconstructions relate to pre-theoretical knowledge of a general sort, to universal capabilities (CES, 14). Following Gilbert Ryle, one can speak of know how vs. know that;[63] reconstructive science is concerned with proven intuitive foreknowledge (know-how) (CES, 12). Yet, this know how becomes in fact the object for a know that, since this intuitive knowledge has to be "brought to consciousness through the choice of suitable examples and

---

[63] Gilbert Ryle, The Concept of Mind (New York, 1949), chap. ii.

counterexamples... through a well-thought-out ma[i]eutic method of interrogation" (CES, 19). In this way, reconstructive science relies on indirect corroboration of its theories, while science relies on direct (observable) confirmations.

Reconstructive science, then, is an empirical science but not empirical-analytical. Its status lies between philosophy and normal science. Even non-foundational philosophy is still concerned with criteria of valid knowledge and norms; it is self-reflexive, while science takes naively the objects of cognition as given.

> Philosophy breaks with this naive attitude and questions the constitution of the object domain of the sciences. In the tradition of critical theory, such questioning means analyzing both the context of genesis and the context of application of theories. Self-reflexivity, in the sense emphasized by critical theory, entails critical awareness of the contingent conditions which makes one's own standpoint possible (context of genesis), and an awareness of whom and what the knowledge one produces serves in society (context of application). Such self-reflexivity leads us... to become aware of the "motives of thought"..., and is a constituent of individual and collective autonomy.[64]

Kant understands experience as constituted; his model is an epistemological model, while Habermas' model of investigating processes of understanding is in terms of "deep" and "surface" structures (CES, 24). 'Deep' structure here means the system of rules underlying the competence of any subject to produce meaningful expressions, in contrast to

---

[64] Benhabib, Critique, Norm, Utopia, p.281.

'surface' structure which is the concern of those interested in explicating semantic relations of a language, a culture, or a form of life.[65] Deep structures refer to species competencies, i.e. to universal competencies and not to competencies of a particular individual or a particular group.

Habermas claims that his rational reconstruction has a universal status, yet he also claims that it is hypothetical in the same way that all theoretical knowledge is hypothetical.

> There is always a possibility that they rest on a false choice of examples, that they are obscuring and distorting correct intuitions, or, even more frequently, that they are overgeneralizing individual cases. For these reasons they require further corroboration. While this critique of all a priori and strong transcendental claims is certainly justified, it should not discourage attempts to put rational reconstructions of presumable basic competencies to the test, subjecting them to indirect verification by using them as inputs in empirical theories. (MCCA, 32)

Habermas considers the empirical social sciences "value-neutral" approach to be one-sided. Normative considerations must be subjected to judgements which are not of the cognitive-instrumental type. Theoretical propositions refer to truth, to what is the case, while in studying everyday life one encounters claims to validity which are non-descriptive and these have to be taken into account or explicated within a context of communication processes.

In so far as rational reconstructions explicate the

---

[65] Thomas McCarthy, The Critical Theory of Jurgen Habermas (Cambridge, Mass. : MIT Press, 1978), p.277.

conditions for the validity of utterances, they also
explain deviant cases, and through this indirect
legislative authority they acquire a **critical** function as
well. In so far as they extend the differentiations
between individual claims to validity beyond traditional
boundaries, they can even establish new analytic
standards and thus assume a **constructive** role. And in so
far as we succeed in analyzing very general conditions of
validity, rational reconstructions can claim to be
describing universals and thus to represent a **theoretical**
knowledge capable of competing with other such knowledge.
At this level, weak **transcendental** arguments make their
appearance, arguments aimed at demonstrating that the
presuppositions of relevant practices are inescapable,
that is, that they cannot be cast aside. (MCCA, 31-32)

There are two dimensions to Habermas' reconstructive
science: one is a "horizontal" reconstruction of some basic
interactive competencies and a "vertical" reconstruction of
the development of these competencies at different stages of
a human being's life[66]. These competencies "are defined as
capacities to solve particular types of empirical-analytical
or moral-practical problems" (MCCA, 33). While Piaget's
concern is with empirical-analytical problems, Kohlberg's
concern is with moral judgements. They both "explain the
acquisition of presumably universal competencies in terms of
patterns of development that are invariant across cultures"
(MCCA, 35).[67]

However, Kohlberg's theory is more problematic since
"he has to prove (a) that a universalist and cognitivist
position can be defended against moral relativism...and (b)

---

[66] McCarthy, The Critical Theory of Jurgen Habermas,
p.278.

[67] Kohlberg, "From is to ought" , pp.163-180.

that a formalistic ethics...can be demonstrated to be superior to utilitarian and contractarian theories" (MCCA, 35-36).

What interests us in this chapter is Habermas' conception of moral cognition, i.e. the basic (intuitive) competence to make moral judgements; the ability to differentiate moral judgements from evaluative judgments and the degree to which an individual can integrate conflicting claims, namely, to take into account the claims of all others in order to resolve these conflicts; the intuitive knowledge of what it means to justify a norm; the ability to argue in defense of a norm; all of these are part of what Habermas considers to be basic interactive competencies of autonomous individuals.

Since Habermas accepts Kohlberg's theory of moral development, as well as Piaget's theory of cognitive development (which together are referred to by Habermas as "developmental-logic"), it may be useful to present a summary of these theories. However, as we shall see a little later, he only uses these theories in order to supplement them with (not to derive from them) his reconstruction of two basic competencies that he believes are related to moral development and which are not dealt with in the above mentioned theories.

According to Kohlberg, the "distinctive characteristic of the moral is that it involves active judgement... Judgement is neither the expression of, nor the description of,

emotional or volitional states, it is a different kind of function with a definite cognitive structure". A rational being may not necessarily be moral, but morality necessarily involves rationality.[68]

Moral consciousness develops, according to Kohlberg, in stages as the individual grows from early childhood to maturity; it is an acquired competence, as is Piaget's cognitive capacity.

> The cognitive structures underlying the capacity of moral judgment are to be explained neither primarily in terms of environmental influences nor in terms of inborn programs and maturational processes. They are viewed instead as outcomes of a creative reorganization of an existing cognitive inventory that is inadequate to the task of handling certain persistent problems. (MCCA, 125)

As it happens, Kohlberg (who relies on Piaget's theory of cognitive development) explicitly describes his theory in terms of Western moral philosophies ("formalistic theories... from Kant to Rawls"[69]), which means, in effect, that one theory tries to corroborate itself by the theories of the other, thereby opening itself to the criticism that it is circular. Habermas argues that although there is a circle here, it is not vicious. One theory can adjust itself to the other when corroboration demands it, and anyway, even in the natural sciences the natural world in question is presupposed.

---

[68] Kohlberg, "From is to ought", pp.185-187.

[69] Lawrence Kohlberg, "The Claims to Moral Adequacy of a Highest Stage of Moral Judgment", Journal of Philosophy, 70 (1973): p.633.

He claims that data used to test an empirical theory cannot be described independently of the paradigms furnishing their basic concepts.

> On the meta- or inter-theoretical level, the only governing principle is that of coherence. We want to find out what elements fit together... It is only a question of seeing whether the descriptions produced with the aid of several theoretical spotlights can be integrated into a relatively reliable map". (MCCA, 118)

According to Kohlberg, the moral criteria of the adequacy of moral judgement help to define a standard for the study of psychological advance. Psychological studies show that individuals prefer the highest stage of reasoning they can comprehend. However, if the studies showed an inconsistency with moral theories, the philosophical claims would become questionable:

> Science, then, can test whether a philosopher's conception of morality phenomenologically fits the psychological facts. Science cannot go on to justify that conception of morality as what morality ought to be.[70]

There are three basic levels, according to Kohlberg, that are involved in the development of moral judgements; these represent in fact three different systems of moralities (namely, each level constitutes a structural whole). Each level, in turn, has two stages. The first level (early childhood: stages 1 and 2) is the "preconventional" level which is characterized by hedonism and obedience to

---

[70] Lawrence Kohlberg, "The Claim to Moral Adequacy of a Highest Stage of Moral Judgment", p. 634.

authorities, and which is expressed in particular concrete actions. Obedience to authorities occurs because of fear of punishment or expectations of reward and not because of respect for a moral order (namely, not because it is believed to be "wrong" or "right"). The second level (stages 3 and 4) is the "conventional" level which is characterized by conformity to culturally accepted norms. Actions are understood in terms of their accord with culturally defined roles. The third level is the "postconventional" level which is characterized by legalism, social contract theories and utilitarianism (stage 5), and universal ethical principles (stage 6) (MCCA, 123-125).[71]

Each stage incorporates the preceding stage in that it reorganizes its structure; in other words, it supersedes it while retaining its elements and moves forward by means of a learning process. This learning process is seen as an increase, in each stage, of reflective thinking. The first level is "concrete thinking"; in the second level, the reflection is on the first level, and in the third level the reflection is on the second. The structure in the higher level means an increased capacity of the individual to accommodate more roles, perspectives and attitudes (integration), a capacity that will bring the individual into more equilibrium (MCCA, 123-125). The transition from one stage or level to the next indicates an increase in adaptations and active

---

[71] Kohlberg, "From is to Ought", pp.195-212.

constructions of the subject. Variations in the development can be seen across as well as within societies. The development of these stages is linear (irreversible) and invariant and is considered by Kohlberg to be universal for the species.

In cognitive development (Piaget)[72], these levels are characterized in terms of the forms of attitudes towards the world (natural, social or the self) beginning with childhood. The development begins, as preconventional, with the symbiotic attitude (no distinction is made between the outside world and the self); then it becomes ego-centric (first person's point of view). On the conventional level, it becomes sociocentric-objectivistic (one's own culture's point of view, awareness of norms of actions, ability to see oneself apart from society, yet the social world is seen as natural); then finally, the postconventional universalistic and decentered stage. A 'decentered' attitude means the adoption of a hypothetical (fully reflective) attitude towards the natural world, the social world as well the self (MCCA, 138).

This developmental-logic means an increasing mastery of a complex of structures that enable the maturing individual to take part in increasingly complex systems of actions (role behaviour, argumentation, etc.). According to Habermas

> Discourse Ethics is compatible with this constructivist notion of learning in that it conceives discursive will

---

[72] Jean Piaget, Principles of Genetic Epistemology (New York, 1970).

formation (and argumentation in general) as a reflective form of communicative action and also in that it postulates a **change of attitude** for the transition from action to discourse. A child growing up, and caught up, in communicative practice of everyday life is not able at the start to effect this attitude change [emphasis in the text]. (MCCA, 125)

There are two basic interactive competencies (social perspectives) which are, Habermas claims, essential to moral consciousness (moral point of view) that develop naturally but which, he argues, are not sufficiently analyzed by Kohlberg. This approach is in line with his criticism of any theory he accepts but which he regards as one-sided, namely, that it does not pay enough attention to practical reason. Habermas relies on theoretical knowledge but his real concern is with its implications for his conception of practical reason (in Kant's terminology **pure** practical reason).

The social perspective that Habermas has in mind are:

1) **Behavioral expectations** that develop from particular actions (preconventional), to social roles and norms (conventional), to principles of choosing between roles and other normative expectations (postconventional). This is achieved through increasing reflexivity from one stage to another (CES, 86);

2) **Reciprocity** that develops as the maturing individual learns to participate in increasingly complex forms of interaction due to an increasing ability to imagine the standpoint of others (mutual recognition). According to Habermas, this reciprocity is not a norm but a basic

interactive knowledge of speaking and acting subjects, that develops naturally and makes interaction possible (CES, 88).

At level one, the interaction is incomplete (parent-child, teacher-pupil, etc.); at level two, the interaction is complete from the point of view of cultural norms and role behaviour, but incomplete because there is an appeal to a higher authority with regard to these norms; at level three, the interaction is complete in that social norms lose their "natural" authority, and there is a mutual recognition of the autonomy of the other and the need for a discourse in order to solve moral conflicts.

Thus, a mature moral actor possesses, according to Habermas, a complex structure of acquired competencies that include speaker, hearer and listener's attitudes, namely, first, second, and third person's (objectifying) perspectives.

Habermas is thinking in terms of acquired competencies that depend on the contingent society within which an individual is socialized. The human being is conceived of as having basic competencies, including behaviour expectancy and reciprocity, which here sound neutral but seem to correspond to Kant's moral terms of duties and obligations. As Habermas understands it, they represent basic competencies that enable both a social interaction (the basic unit of society) and a moral development which can be characterized as either authoritarian, conventional, or postconventional, depending on which society the individual has been socialized in.

Habermas also sees a parallel development in terms
of authoritarian societies (first level morality), conformist
societies   (second   level   morality),   and   autonomous
communicative action/discourse   (third   level   morality)
societies. This is Habermas' theory of the evolution of
societies (reconstruction of historical materialism (CES,
chap. iv)) which is very controversial and is considered by
critics to be inspired by Hegel's philosophy of history and
Marx's historical materialism. This criticism is interpreted
to mean that Habermas has not completely freed himself from
the conception of the "philosophy of the collective subject"
which is the model for both Hegel and Marx, and which Habermas
himself rejects as an unacceptable metaphysical model (see
p.13 above).[73]

While the theoretical description of the first two
levels and their stages are more or less accepted by critics
without debate (i.e. they may be seen as a natural development
of moral consciousness by means of a process of learning), the
postconventional level is controversial for various reasons.

First, there is a question of whether the transition
to   the   postconventional   level   describes   a   universal
phenomenon, or whether it is a description of the development
of a modern Western scientist or moral theorist. As Thomas
McCarthy argues, Habermas has not been so critical of Piaget's
and Kohlberg's theories as he has been with regard to other

---

[73] Benhabib, Critique, Norm, Utopia, p.343.

theories. Critics point to the fact that most of Piaget's and Kohlberg's research was done in modern Western societies. Both seem to take modern Western societies as the norm for measuring cognitive and moral development.[74]

Yet, McCarthy does not reject Habermas' claim that differentiation of attitudes towards separate realities (nature, society, self) indicates a cognitive development and progress. Rather, McCarthy suggests that this differentiation may need to be further developed in the direction of a "unity in difference".[75] Habermas seems to accept this.[76]

There is, however, the question of whether level three is superior to level two as a standard of resolving moral conflicts. As Stephen K. White argues, reciprocity (a viewpoint that constitutes, according to Habermas, the moral point of view) is already available and mastered by competent actors in level two:

> Clearly this makes reciprocity a viewpoint available to mature individuals in all societies; but just as clearly there is no natural necessity for this viewpoint being consciously accepted as the standard for resolving moral conflicts. The reflexive use of the viewpoint of reciprocity as a requirement in moral deliberation can only be conclusively justified when one has already

---

[74] Thomas McCarthy, "Rationality and Relativism: Habermas's 'Overcoming' of Hermeneutics" in John B. Thompson and David Held (eds.) Habermas: The Critical Debates (London: MacMillan Press, 1982), pp.68-70.

[75] McCarthy, "Rationality and Relativism", p.78.

[76] Jurgen Habermas, "The Unity of Reason in the Diversity of Its Voices", in his Postmetaphysical Thinking: Philosophical lectures, trans. William Mark Hohengarten (Cambridge, Mass.: MIT Press, 1992), pp.115-148.

assumed the superiority of the moral point of view associated with a decentered, postconventional consciousness [emphasis in the text].[77]

Another question (by McCarthy) is whether at the third level the transition from stage 5 (legalism, social contract theories) to stage 6 (universal ethical principles) can be considered a "natural" development, since at this level there is no distinction between the reflective attitude of the moral theorist or psychologist and the ordinary subject under investigation. Differences between their moral attitudes and theories have to be worked out through philosophical debates and argumentation.[78] Habermas does accept this criticism:

> He or she is not merely using a special competence in a naive way, but is incipiently already involved in reconstructing. The manner in which a question is resolved already betrays an implicit theory regarding what it means to ground a normative proposition. But then competing views of this kind can just as little be placed in a hierarchy from a developmental-logical standpoint as can the corresponding "higher" forms of moral philosophy. McCarthy's thesis is all the more plausible to me as it agrees with the empirical evidence".[79]

The transition from "conventional" to "postconventional" thinking means a breaking away from conventional thought. When this happens, two possibilities lie ahead, according to Habermas:

---

[77] Stephen K. White, The Recent Work of Jurgen Habermas: Reason, justice and Modernity (Cambridge: Cambridge University Press), 1988, p.64.

[78] McCarthy, "Rationality and Relativism", pp.68-75.

[79] Jurgen Habermas "A Reply to My Critics", in John B Thompson and David Held (eds.), Habermas: The Critical Debates (London: MacMillan Press, 1982), p.260.

1) One possibility is that while the de facto norms have lost their "natural" character, one can still preserve the sense of validity of norms, relativize the devalued norms and judge them in the light of rationally grounded principles of normative validity; this is the necessary condition to the transition to the postconventioanl level;

2) a second possibility is that one is freed from the conventional way of thinking by gaining an insight into the illusory character of conventional consciousness, while still preserving the moral intuitions that determine unreflected thinking. Instead of a postconventional moral consciousness one arrives at a meta-ethical explanation of moral illusion. The results are either ethical scepticism a la Max Weber (a reconciliation of theoretical scepticism with existing value attachments), or ethical naturalism (emotivism) (MCCA, 184-187).[80]

> If one takes McCarthy's argument seriously, one is no longer able to order cognitivist ethics from a developmental-logic standpoint; one has to return them to the debate of the philosophers and thus to the historical context of this debate as well.[81]

> The debate among the moral philosophers cannot be settled with the psychological assertion that Kantians have better, structurally privileged access to their moral intuitions than do rule-utilitarians or social-contract

[80] Habermas,"A Reply to my Critics", pp.260-261.

[81] Habermas, "A Reply to my Critics", p.261.

theorists in the Hobbesian tradition.[82]

The way Habermas sees it, at levels one and two (preconventional and conventional) basic social competencies are being acquired by means of a learning process that becomes increasingly reflexive, while at the postconventional level, learning processes can proceed only if the reflective abstraction previously operative as a learning mechanism is sublimated, as it were, into the procedure of rational reconstruction, however ad hoc and unmethodically that procedure may be pursued.

> All those who make moral judgments at the postconventional level, whether they be psychologists, research subjects, or philosophers, are participants in the joint venture of finding the most appropriate possible explanation of a core domain of moral intuition to which they have access under fundamentally equal social cognitive conditions.[83]

The point is that at this level moral judgements do not represent merely a pre-reflective use of an intuitive know-how but a beginning of an explication of a moral theory.

Finally, there is the question of ethics of conviction and ethics of care and responsibility, or in terms Habermas prefers, justice and solidarity. Carol Gilligan claims that there seem to be two types of moralities, one of which is ethics of justice (with regard to which psychological studies

---

[82] Jurgen Habermas, "Justice and Solidarity", in Thomas E. Wren (ed.), The Moral Domain: Essays in the Ongoing Discussion between Philosophy and the Social Sciences (Cambridge, Mass.: MIT Press, 1990), pp.226-227.

[83] Habermas, "Justice and Solidarity", p.227.

show men test higher than women in Kohlberg's scale of stages) while women test more than men in ethics of care (which is lower in Kohlberg's scale of stages). There seem to be therefore two different moralities which are differentiated by gender.[84]

Discourse ethics as a cognitive theory has been regarded by critics to be concerned with questions of justice rather than with questions of care and responsibility. In this, critics claim, Habermas follows Kant. Habermas, however, claims that discourse ethics does include ethics of care, or rather ethics of solidarity which includes ethics of care. Principles of responsibility and care are already contained in the meaning of his term 'normative validity' (MCCA, 181). It should be remembered that for Habermas, what constitutes the moral point of view are the social perspectives of behaviour expectancy and reciprocity.

At the postconventional level, the social world (institutional) becomes detached from the lifeworld and is compelled to justify itself. As the natural world becomes theorized (science), so the social world becomes moralized

---

[84] Carol Gilligan, In a Different Voice: Psychological Theory and Women's Development (Cambridge, Mass.: Harvard University Press), 1982, pp.164-165.

Carol Gilligan, "Do Social Sciences Have an Adequate Theory of Moral Development?", in Norma Haan and others, Social Science as Moral Inquiry (New York: Columbia University Press, 1983).

Carol Gilligan, "New Maps of Development: New Visions of Maturity", Annual Progress in Child Psychiatry and Child Development (New York: Brunner/Mazel, 1984), p.4.

(judged according to moral principles) (MCCA, 177). Practice of everyday communication splits into two parts: norms and values (morality and the good life). Moral questions differ from evaluative questions in that the latter are connected with the particular lifeworld that shapes the ego identity of its members, while moral questions are taken "out of their context in such a way that moral solutions retain only the rationally motivating force of insights" (MCCA, 178). Moral questions, however, do not occur in a void; they always presuppose a form of life with its values.

Habermas does not accept Kant's conception of a human being as "deficient" (natural inclination) trying to overcome its deficiency by acting in a moral manner (duties and obligations). Habermas regards the individual human being as a socialized being whose basic competencies develop into competencies within a specific form of life. Kant's conception of basic given needs and interest of the individual is rejected by Habermas who regards the needs and interests of a socialized individual as culturally interpreted.

In Kant, what makes morality possible is the freedom of consciousness from the causal laws of nature. Moral consciousness is a "fact of reason" that cannot be explained nor denied. In the final analysis, autonomy for Kant means the capacity (which is at the same time also an ideal) of the individual mind to emancipate itself from natural inclinations and traditional unquestioned norms of thinking and acting.

In Habermas, autonomy means first of all having an acquired (that is also an ideal) set of competencies and attitudes (cognitive, interactive, reflective) that constitutes ego and moral autonomy . It is the ability to adopt simultaneously a performative attitude (interactive, standpoint of the others, reciprocity) and an objectifying attitude (hypothetical-reflective).

There is a similarity with Kant in that mutual recognition or reciprocity (the formula of the end in itself) is a key feature or an ideal in both Habermas and Kant. As we saw, Kant's moral agent is to consider the social order that will result if all concerned will act in a selfish manner. For Habermas, social interaction can be characterized either as **strategic** interaction (one party will try to overcome the other by force) or as **communicative** interaction (conflicts will be resolved by force of arguments). It is the second type which implies, for Habermas, autonomy of the individuals involved and recalls Kant's formula of the end in itself.

The fundamental difference between Kant and Habermas is in the relation of the autonomous individual to his or her own natural needs. For Kant, autonomy means overcoming or suppressing **given** natural desires by the force of reason. For Habermas, this concept of autonomy vs. heteronomy is rejected (MCCA, 203). Instead, autonomy means the competence to adopt the standpoint of the other and at the same time also a reflective-hypothetical attitude towards one's own individual

needs and interests. In this sense, autonomy in Habermas is a complete decentering, namely, a capacity to reflect on everything concerning human life, including "natural" inclinations, which for Habermas are always culturally interpreted or even created and therefore could and should be publicly debated and discussed. The abstract human individual of the Enlightenment with its given interests and desires is removed and replaced in Habermas by the concrete individual whose needs are culturally interpreted or created.

Habermas' conception of ego autonomy as decentered and self-reflective and therefore self-transparent has been interpreted by critics to signify a suppression of one's own individual "inner" (non-rational, bodily subjective) nature. Michel Foucault is a case in point. He asks: "In what is given to us as universal, necessary, obligatory, what place is occupied by whatever is singular, contingent and the product of arbitrary constraints".[85] This question is important to Habermas whose concern is with the freedom of the individual from external and internal constraints and oppression, such as ideology, politics, and all other forms of social domination and psychological subjugation. The model offered by Foucault is the aesthetic self-formation of the body that is supposedly free from the attention of the social and political worlds.[86]

---

[85] Michel Foucault, "What is Enlightenment" in Paul Rabinow (ed.), Foucault Reader (New York: Pantheon Books, 1984), p.45.

[86] White, The Recent Work of Jurgen Habermas, p.146.

Habermas' conception of authentic self-expression has a double meaning, according to Stephen K. White:

a) Sincerity of the speaker towards the hearer, which can only be demonstrated by the speaker's actions, rather than by declared intentions;

b) self-expression of authentic needs which can also only be demonstrated by the individual's behaviour i.e., whether his or her interpretation of needs are self-deceptive or self-debilitating.[87]

Habermas' conception of authentic self-expression entails the adoption of the same hypothetical-reflective attitude that we already have with regard to science and postconventional morality. The decentered modern structure of consciousness enables the hypothetical-reflective attitude to be directed also towards "inner" (bodily, non-rational) nature and thus makes authentic aesthetic experiences possible, precisely because what was taken previously to be "natural" now collapses under the hypothetical gaze.

> At the same time, this decentering indicates an increased sensitivity to what remains unassimilated in the interpretive achievements of pragmatic, epistemic, and moral mastery of the demands and challenges of everyday situations; it effects an openness to the expurgated elements of the unconscious, the fantastic, and the mad, the material and the bodily - thus to everything in our speechless contact with reality which is so fleeting, so contingent, so immediate, so individualized, simultaneously so far and so near that it escapes our

---

[87] White, The Recent Work of Jurgen Habermas, p.147.

normal categorical grasp.[88]

The conception of aesthetic self-formation that Foucault offers conflicts with Habermas' conception of intersubjective otherness (reciprocity):

> It becomes difficult to see how Foucault could possibly bring his aesthetics of existence into any coherent relationship with his endorsement of some forms of collective political action. His antipathy to anything juridical leads him inexorably into a conceptual cul de sac in relation to ethics and politics.[89]

"Autonomy" for Habermas means both recognition and respect of the social "other" (reciprocity, mutual recognition), as well as self-recognition of one's own inner (non-rational, bodily, subjective) "other".

Kant's moral ideal of the autonomy of the individual is a standard assumed for Habermas' social and moral theory, as well as for all his critics who try to deny the importance of rationality in human life. The debate between the "rationalists" (such as Habermas is) and the "non-rationalists" (such as Foucault) is in fact a debate about what human autonomy is and how it can be better achieved.

To sum up: Kant's moral ideals were envisioned to be a-historical and universal. Habermas is aware of a culturally conditioned moral consciousness as a "fact of historical reason" that develops genetically, culturally and across the

---

[88] Jurgen Habermas, "Questions and Counterquestions", in Richard J. Bernstein (ed.), Habermas and Modernity (Cambridge, Mass.: MIT Press, 1985), p.201.

[89] White, The Recent Work of Jurgen Habermas, p.151.

species.

Kant's conception of consciousness is of an inner rational creativity; in Habermas, it is structurally differentiated and intersubjective. Ego autonomy is an artistic conception in that it regards the individual as mastering a complex structure of competencies that enable him or her to interact socially and thus to become morally constructive, as well as creative in personal self-expression. This means that Habermas' conception of the individual's moral consciousness is richer than that of Kant's rigid conception of moral consciousness. In Habermas, various aspects of concrete individual and social situations are reflected in his conception of moral consciousness.

CHAPTER III

HABERMAS' DISCOURSE ETHICS:
METHODS OF VALIDATION IN RELATION TO KANT'S

The purpose of this chapter is to examine, in more detail, Habermas' discourse ethics in relation to Kant's categorical imperative and Kant's methods of validation (as was examined in the first chapter) and against the background of Habermas' theoretical conception of moral consciousness (as was examined in the previous chapter). Habermas considers discourse ethics to be a procedural version of the categorical imperative, a reformulation of it in terms of pragmatism and contemporary philosophy of language. The question I will deal with in this chapter is how to understand discourse ethics as a reformulation of the categorical imperative.

As we saw (pp.29, 41-43), the categorical imperative is a procedure of moral reasoning (which can be said to show or describe moral consciousness in action. As Kant understands it, the aim of the procedure is to regulate the choice of our maxims in everyday actions. What characterizes the categorical imperative is that it is a monologue of moral reasoning. This monologue is in fact an inner rational dialogue (in that the moral agent should reflect from the perspective of the other - the end in itself formula) that is to lead each individual to a moral insight which will then guide his or her choice of maxims of actions; this is the

validating aspect of our moral actions.

The elements in Kant's conception of practical reason, such as reciprocity and autonomy in the sense of freedom and as self-legislation in the kingdom of ends, also play a role in Habermas' conception of validity and validation, although they are given a different interpretation. What in Kant is understood to happen in the individual consciousness (which is identical for all) is in Habermas understood to be a public (intersubjective) linguistic discourse. Kant's ethics of dialogue has become in Habermas a dialogic ethics.[90]

Kant's moral order was part of a harmonious cosmic order and although Kant's ethics can be interpreted to be a social ethics, Kant does not have a critical social theory in Habermas' sense of it of the term. He seems to accept the social order as a "given", which means that society is not an issue for him, as are moral actions of individuals. Morality is an inner self-reflection of the individual's mind which is a sign of autonomy. In Habermas, the social order (namely, institutions) is no longer considered to be a "given" (i.e. unchangeable) but has to be judged (or to justify itself) according to universal moral principles. Morality is no longer an inner-reflection but an intersubjective reflection.

---

[90] Albrecht Wellmer, "Ethics and Dialogue", in his The Persistence of Modernity: Essays on Aesthetics, Ethics and Postmodernism, trans. David Midgely (Cambridge,Mass.: MIT Press, 1991), p.142.

Discourse ethics includes two main principles (set of preconditions) that a valid norm would have to fulfil:

1) The principle of universalizability (U) which states that "for a norm to be valid, the consequences and side effects of its general observance for the satisfaction of each person's particular interests must be acceptable to all";

2) the principle of discourse (D) which states that "only those norms may claim to be valid that could meet with the consent of all affected in their role as participants in a practical discourse" (MCCA, 93).

Habermas regards U as the only universal moral principle that holds for any moral discourse in any context, regardless of cultural content, which means that he considers its validity to be transcultural. This principle, he maintains, is to be distinguished from all other substantive moral principles or basic norms because the latter "can only be the subject matter of moral argumentation" in discourse, not the preconditions for the validity of the discourse. It is also to be distinguished from principle D "which stipulates the basic idea of a moral theory but does not form part of a logic of argumentation" (MCCA, 93). This means that discourse is always a factual discourse (within a form of life), while U is applicable to all forms of life.

As with Kant, we find in Habermas both (1) a justification of principle (U) itself as a validating factor of norms, and (2) a justification of the discourse as a

validating factor of substantive norms. This corresponds, in the first case, to Kant's transcendental deduction of the categorical imperative, and in the second case, to the categorical imperative as a procedure for validating the choice of maxims.

I shall examine Habermas' methods of justification in relation to Kant after clarifying, first, certain conceptions of Habermas that are related to his conception of discourse ethics.

According to Habermas, discourse ethics

stands and falls with two assumptions: (a) That normative claims to validity have cognitive meaning and can be treated like **claims to truth** and (b) that the justification of norms and commands requires that a real discourse be carried out and thus cannot occur in a strictly monological form, i.e. in the form of a hypothetical process of argumentation occurring in the individual mind (MCCA, 68).

Habermas distinguishes between **truth** of descriptive statements and **validity** of normative sentences. He argues that moral sceptics think in terms of assimilating or reducing normative sentences into truth statements, something that cannot be done, with the result that norms are believed to be subjective preferences (MCCA, 53-54). According to Habermas, normative validity claims should be regarded not in the same sense as truth claims but as "analogous to truth claims". Propositional statements, he argues, relate to facts in the same way that regulative statements relate to "legitimately ordered interpersonal relations". The truth of a sentence implies an **existence** of a state of affairs, while the

rightness of actions refers to an observance of norms. Norms, however, depend on a continual reestablishment of legitimately ordered interpersonal relationships. Reference to norms assume actors who follow them and actions that fulfil them, i.e. a social world with accepted norms (MCCA, 45-48).

However, the existence of socially accepted norms (social facts) does not mean that they are worthy to be recognized as valid on their own right (regardless of whether they are accepted or not). The judgement with regard to the worthiness of a moral norm to be so recognized depends more on a practical discourse (moral argumentation) than empirical judgements depend on theoretical discourse. There is a connection between the "existence" of norms and the anticipation of the justifiability of these norms, while there is no such connection for the existence of states of affairs; states of affairs exist, whether we anticipate them or not (MCCA, 61-62).

In theoretical rationality we rely on "perceptions" (observations) which are the basis on which we can arrive at a consensual validation. Habermas claims that the role of "perceptions" in theoretical rationality is filled in practical rationality by a "web of feelings and attitudes", a sense of justice and injustice, of duty or of guilt, a certain normative expectation which is the basis on which norms of actions can be judged to be valid or not (MCCA, 50-51). Social norms and beliefs may become questionable, however,

when there seem to be sufficient reasons to doubt their validity. When this happens, a discourse should be called for, according to Habermas, in which an attempt would be made to arrive at a new mutual understanding or an new agreement with regard to the validity of intersubjectively shared norms:

> I shall speak of 'discourse' only when the meaning of the problematic validity claim conceptually forces participants to suppose that a rationally motivated agreement could in principle be achieved, whereby the phrase 'in principle' expresses the idealizing proviso: if only the argumentation could be conducted openly enough and continued long enough. (TCA I, 42)

It is important to clarify here two points which seem to cause confusion: One is that 'discourse ethics' should be understood to deal exclusively with norms whose validy have become questionable, and not with everyday communicative action. The other is the question of what is involved in a rationally motivated agreement.

With regard to the first one: What we have here is a reflective level which is concerned with the validity of norms and not a reflection within communicative action (everyday) level (MCCA, 67). This should be emphasized, since critics of Habermas' discourse ethics sometimes fail to distinguish between the two levels, with the result that their criticism does not always apply to Habermas' discourse ethics as a concrete critical reflection on the validity of norms of actions and on the methods of their justification.

Habermas does distinguish between de facto agreement in everyday communicative action, and genuine or legitimate agreement achieved under ideal conditions for communicative action. 'De facto' agreement in this case will be an agreement to coordinate actions according to accepted norms, so that those involved can freely pursue their private goals. In actual situations this may involve strategic actions, forced agreement, manipulation or deceit, in short a "constrained" agreement, a situation which is all too familiar to us from social and political life. An ideal condition would in this case be an 'unconstrained' agreement that was arrived at by all concerned without coercion. Such an agreement could be achieved, according to Habermas, only by means of a debate and convincing arguments, since we can arrive at a genuine agreement only by way of a linguistic communication that requires convincing arguments and mutual understanding, and not by means of Kant's methods of a solitary thinker.

However, in everyday actions time may be a restricting factor when a decision has to be made in order not to bring everyday activities to a standstill. Thus, a compromise of some sort may be required in order to be able to go on. It is different with regard to the question of an agreement concerning the validity of the norms themselves, which have wider moral implications. Validating the norms requires more time and elaboration and more understanding of the wider consequences involved. Strategic actions and constrained

agreements in communicative action may in fact reflect an imbalance of power and a privilege of one group over another, something that may at times be inevitable; yet, this cannot be tolerated, at least not in principle, when it comes to the validation of general norms of actions that should apply equally to all.

The specific concern of discourse ethics is, therefore, to be understood with regard to the problem of justifying those socially accepted norms of actions that have become problematic, or new norms that require validation. As we saw in the previous chapter, justification of socially accepted or traditional norms and principles may be required, since for one reason or another they do become questionable to a postmetaphysical, postreligious or postconventional consciousness. The problem is to determine what it means in principle to justify a norm and how to justify it in a manner which avoids both recourse to an external authority such as a dogmatic religion or a metaphysical tradition, on the one hand, and moral scepticism on the other.

One reason for the confusion of levels seems to be that in Kant the categorical imperative is a moral principle to be applied in everyday actions. Principle (U) is a moral principle in that it sets the precondition which validates the discourse (principle (D)). In other words, principle (D) is the validating debate that is guided by principle (U).

Kant's categorical imperative is a procedure of moral

reasoning which will guide us in choosing our maxims of actions. A maxim is a subjective principle of action and the procedure of the categorical imperative is intended to test its moral permissibility. Kant's moral agent is not a naive agent but one who already has some moral knowledge and intuitions, so that he or she becomes aware of the need to use the categorical imperative as a test for the universalizability of the maxims. Otherwise, it would be difficult to imagine how an agent will know when, or if at all, to use the categorical imperative as a test of what maxim to choose.[91] It is the self-understanding of the agent as part of a community and an awareness of the other as an end-in-itself that implies for Kant a moral consciousness (or the moral law).

In Habermas' case, however, it is not a question of what maxims to choose in everyday life but of determining whether a moral norm that was formerly accepted as part of a moral system can be justified. This is not an individual matter but an application, so to speak, of the "categorical imperative" as a test of the validity of a norm which will be accepted by all and therefore binding for all. There are, therefore, two levels which should be kept in mind: In Kant, a moral order is assumed to exist prior to the moral agent and the aim of the procedure of the categorical imperative is to

---

[91] Barbara Herman, "The Practice of Moral Judgment", Journal of Philosophy, 82 (1985): 414-436.

arrive at the right moral insight. In Habermas, the existing moral order itself (or rather part of it) may be understood to require a correction, or to be shaped, by means of a discourse. It should be remembered that Habermas' critical social theory covers political, social and moral aspects which for him always intermesh within a concrete form of life.

The other important point is the question of a rationally motivated agreement: The paradigm of rationality is, according to Habermas, the discourse. This kind of rationality is procedural, not substantive. The concern is with validity claims that social members can react to with a "yes" or "no" response, and when required may support their claims or reactions with good or convincing reasons or arguments. It is in this sense of providing good reasons for or against a validity of a norm that Habermas considers norms to have cognitive meaning and which allows the possibility of arriving at a rationally motivated consensus.

A distinction should be made between 'ethical cognitivism' and 'ethical rationalism'. By 'ethical cognitivism' it is meant that good reasons can be provided for or against norms. This is

> the view that ethical judgments and principles have a
> cognitively articulable kernel, that they are neither
> mere statements of preference nor mere statements of
> taste but that they imply validity claims... 'X is
> right'... [means] 'I can justify to you with good grounds
> why one ought to respect, uphold, agree with X'.[92]

---

[92] Benhabib, Communicative Ethics Controversy, p.355.

As such, ethical cognitivism is opposed to ethical decisionism or ethical emotivism. By 'ethical rationalism' it is meant

> a theoretical position which views moral judgments as the core of a moral theory and which neglects that the moral self is not a moral geometer but an embodied, finite, suffering, and emotive being.[93]

Ethical rationalism, according to Benhabib, does not take into account (or rather abstracts from) the contingent origins of moral development and views ethics as a preceding background to such a moral development. In the case of communicative ethics, Benhabib argues that it has been presented as an ethical rationalism while in fact it is a form of ethical cognitivism. The emphasis in discourse ethics seems to be on questions of justice ("rationalism") rather than on moral emotions and character. However, as we saw in the previous chapter, Habermas does include in his conception of 'normative validity' questions of (contextual) emotions and character, to the extent that these have a moral import for the community in question. There may, however, be a problem of deciding which emotions have or do not have a moral import. Again, this could be established only by means of a discourse.

The terms 'cognitivism' and 'rationalism' seem to be used in an interchangeable manner; this should be kept in mind, in particular because 'rational' behaviour can be

---

[93] Benhabib, Communicative Ethics Controversy, p.356.

interpreted also as "instrumental/ purposive". This, too, causes a confusion. In an attempt to avoid any confusion, I will use the term 'rational' and 'cognitive' interchangeably, referring in both cases, either to normative cognitivism or normativ. behaviour, and 'purposive', or 'strategic' rationality for instrumental behaviour.

Discourse ethics presupposes freedom to express doubts and a willingness to enter into a rational discussion (which are the signs of a postconventional attitude). It also presupposes that social members know intuitively what it means to justify a norm (how to provide good reasons). As with Kant, Habermas is concerned with common everyday life and persons, and he shares Kant's belief in the competence of common speaking and acting individuals to deal with moral questions in a rational manner:

> Validity claims are in principle open to criticism because they are based on formal world concepts. They presuppose a world that is identical for all possible observers, or a world intersubjectively shared by members and they do so in an abstract form freed of all specific content. Such claims call for the rational response of a partner in communication. (TCA I, 50)

The proper domain of discourse is

> the formal conditions of rationality in knowing, in reaching understanding through language and in action, both in everyday contexts and the level of methodically organized experience or systematically organized discourse. The theory of argumentation thereby takes on a special significance; to it falls the task of reconstructing formal-pragmatic presuppositions and conditions of an explicitly rational behaviour. (TCA I, 2)

The purpose of a practical discourse, then, is to arrive at an agreement concerning validity claims of norms of actions that have become publicly problematic. It presupposes autonomous individuals who share the same worldview and who are at the same time rationally motivated to come together in a discourse in order to resolve (by the force of the better argument) the problems at hand.

Borrowing from George Herbert Mead, Habermas' conception is that society is prior to the individual: "Individuation is a product of sociation" (TCA II, 3-42).[94] "An autonomous ego organization is by no means a regular occurrence, the result, say, of nature-like processes of maturation" (which is un-Kantian). However, autonomy has more than a descriptive meaning in that it is also a symbolic organization which implies an ideal (which does sound like Kant) (CES, 70).

What is crucial in Habermas' conception of autonomy is that in contrast to Kant, the "other" (who is also an autonomous individual) is not another (identical) "consciousness" monologically arriving at the same rational insight, but a concrete different social "other" with whom one communicates linguistically, whereby individual differences are revealed. Yet, since consciousness is culturally constituted, the consciousness of the "other" is not a

---

[94] George Herbert Mead, Mind, Self, Society (Chicago, 1962).

complete stranger either. Linguistic communication implies both a common understanding (lifeworld) and individual differences.

> Human action is linguistically mediated, both for the actor and for others, who formulate their intentions and the definitions of what they do in linguistic terms. Such formulations are essentially contestable, by ego as well as by alter. The interpretive indeterminacy of social action is not an ontological shortcoming, but its constitutive feature...
> Social action always entails linguistic communication...
> Interpretive indeterminacy is a constitutive feature of social action.[95]

Autonomous individuals may each interpret situations differently but only by means of a discourse can each one reveal his or her differences and arrive at a common understanding through a process of argumentation. By the exchange of arguments, individuals may become convinced by others to modify their original interpretation, or they may convince others to modify theirs; or a new interpretation may arise and be accepted by all. This means that social action or interaction is not constrained by a metaphysical model of one sort or another and is not imposed by force, ideology, religion, etc.. Rather, autonomous individuals may transform certain aspects of their lifeworld or society as a result of their newly gained interpretation which for Habermas means a "rational" consensus.

Habermas' conception of 'rational' consensus requires further elaboration. As already mentioned (see p.96 above),

---

[95] Benhabib, <u>Critique, Norm, Utopia</u>, p.243.

Habermas' conception of 'rationality' should be understood in terms of cognitivism. Principle (U) is a discursive principle which in fact presupposes the democratic principles of personal autonomy, equality and impartiality. The principle would be meaningless, however, if the organization of society were not already based on democratic principles, such as parliamentary and other democratic institutions. Yet, there is a question as to whether democratic principles can in any way be seen to be related to rationality. To answer this question we shall have to examine, in still more detail, Habermas' conceptions of 'rationality' and its role in discourse ethics.

For Kant, "reason" or "understanding" or what we now call "rationality" means a cognitive capacity which is a capacity of rules (understanding), a capacity of principles (reason), and a priori synthetic judgements that constitute the necessary form of descriptive or normative assertions. To be rational means to actualize a potential that gives the person control over his or her behaviour. This conception that rationality gives a person control over his or her behaviour has important moral implications for both Kant and Habermas. One can distinguish in this case between nature (functional behaviour) which is morally neutral, and rational behaviour (which implies freedom, responsibility and control over one's life and actions).

Rational behaviour, in turn, can be understood as

being guided by purposive-instrumental or strategic norms, and by practical rationality (Kant), or communicative rationality (Habermas). Instrumental or purposive rationality is a means-ends rationality, where what is in question is to discover the most efficient way to achieve a desired material end (which involves theoretical knowledge). Practical rationality involves interpersonal relations, which means it is guided by social norms.

Kant's conception of rationality (both theoretical and practical) can be said to be dogmatic in the sense that it presupposes that transcendental rules and principles predetermine our way of thinking and acting. These rules are unchangeable and a-historical and therefore rational behaviour will be identical for all, at all times and in all societies. This conception of rationality does have a metaphysical connotation: Monological reasoning leads to the same insights for all, although Kant's conception of rationality is that it does enable the human being to be free and responsible (in the sense of obligation under the laws of freedom).

By contrast, Habermas' conception of rationality, purposive or practical, is that it is both communicative and contingent, and what is more important, it is an open-ended process[96]. It is not, for him, a capacity of consciousness

---

[96] Herbert Schnadelbach, "Remarks about Rationality and Language", in Seyla Benhabib and Fred Dallmayr (eds.), The Communicative Ethics Controversy (Cambridge, Mass. MIT Press, 1990), pp.272-280.

that can be described in terms of fixed rules but a linguistic communicative competence. Linguistic communication competence cannot be grasped by logical means alone. It is a basic interactive competence that is mediated by language and is therefore subject to different interpretations, due to the fact that language games (or presuppositions of forms of life) differ one from the other as well as the concrete situations which are always contingent.

Linguistic communication seems to be the precondition for a philosophical reflection, which is that of discourse.

> The meanings of 'reflection' and 'self-reflection' change. These no longer refer to the cogitative activities of a Cartesian ego or to the laboring activity of the making self but to processes of communication between selves. They designate the activity through which controversial validity claims, as well as rules of argumentation, are disputed, debated, and adjudicated in discourse.[97]

It is important to note that Habermas distinguishes between 'linguistic competence' and 'linguistic communicative competence'. Linguistic competence means the ability to generate grammatically correct and comprehensible sentences or utterances (CES, 26-28). Linguistic Communicative competence, on the other hand, means the "ability of a speaker oriented to mutual understanding to embed a well-formed sentence in relations to reality" (namely, to nature, society, self) (CES, 29).

> A grammatically well-formed sentence satisfies the claim to comprehensibility; a communicatively successful speech

---

[97] Benhabib, *Critique, Norm, Utopia*, p.282.

action requires, beyond the comprehensibility of the linguistic expression, that the participants in communication be prepared to reach an understanding, that they raise claims to truth, truthfulness, and rightness and reciprocally impute their satisfaction. Sentences are the object of linguistic analysis... speech acts of pragmatic analysis. (CES, 31-32)

The competence that makes communication possible presupposes an understanding (agreement) in language itself.[98] The concept of linguistic competence, however, may imply "a mastery of an abstract system of rules, based on an innate language apparatus, regardless of how the latter is in fact used in actual speech". This involves an assumption that "the sharing of identical meanings... is a pre-established code that is supposed to make communication possible". Habermas regards this conception of linguistic competence to be monological in that language (grammar, phonetics and semantics) is understood, according to this conception, to develop "independently of the pragmatic dimension of language performance". It is for this reason that Habermas considers Noam Chomsky's theory of linguistic competence to be conceived along the lines of the "philosophy of consciousness" or "methodological solipsism".[99]

Linguistic communicative competence includes the concepts of action (strategic rationality) and interaction (communicative rationality). Competence for communicative

---

[98] Schnadelbach, "Remarks about Rationality and Language", p.282.

[99] Jurgen Habermas, "Towards a Theory of Communicative Competence", Inquiry, 13 (1977) : 361-362.

action would require both linguistic competence and at the same time also a competence to make oneself understood with regard to the natural world, society, and self, within a context of a lifeworld. This is in fact a metacommunication competence.

> If one operates with the model of communicative competence in the sense indicated, then one must indeed construe rationality as a complex unity of the possession and application of rules that is distinguished by interaction between its elements, and in which there is no room for one-sided, a priori-deductive relationship between understanding/reason and judgment, and performance. Viewed in this way, rationality can at least be seen as an open system of communicative competence that is not completely represented in rules - a system that, though bounded by rules, nonetheless contains rules of a type which permit intentional changes.[100]

Linguistic **communicative** competence does not mean following "correct" (rational) rules blindly but of using or disposing of them and even creatively transforming them. Rationality, it seems, can be represented only as a "basic stock of communicative rules of competence", and not as a-priori fixed rules, which means that rationality is not "pure", as Kant believed it should be, but is "infected" by empirical and contingent contents. For Habermas, what it means to be rational, or what it means to rationally justify a norm, differs from Kant's conceptions of it.

Rationality is an elusive concept and circular. In order to understand what rationality is we always (already) make use of it, which means that we can never really transcend

---

[100] Schnadelbach, "Remarks about Rationality and Language", pp.283-284.

it. Rationality is an intuitive pre-theoretical competence
that we try to reflect upon. It cannot be "pure" as in Kant
because the validity claims are historically and culturally
influenced.

There are two basic questions with regard to
rationality:

(1) Is it desirable (i.e. does it have a normative value
in addition to its instrumental value) as Kant believed it
had, and is so with Habermas (in a modified form), or is it to
be considered undesirable and oppressive as we have seen in
the case of Foucault (see p.82 above)? The doubts about the
desirability of rationality (its moral implications) began
with Schopenhauer and Nietzsche and these doubts continue to
be expressed even today by modern philosophers;

(2) the second question is with regard to a rationally
motivated consensus in relation to normative validity claims,
namely, whether this is possible at all and if so, whether it
is superior to other forms of consensus, such as is arrived at
in conventional or authoritarian cultures. This raises, among
other questions, the problems of moral relativism.

Habermas is not in search of a first foundation for
rationality or knowledge that would enable us once and for all
to solve our moral conflicts. In fact, he argues against this
conception of rationality, yet he also argues that we cannot
criticize rationality, let alone deny its desirability or
superiority, without precisely making use of it by means of

convincing rational arguments in the very process of denying its importance. In this respect, those who try to deny the importance of rationality in social, moral and personal life already participate in a reflective (rational) debate about rationality itself and do so by engaging in rational arguments ("radical self-criticism").[101] This is all the more obvious when moral relativism is defended by means of rational arguments that are not necessarily the preferred form for a conventional or authoritarian consciousness.

Reflections on the validity of norms of rationality and rational questioning of the validity of traditional norms are indeed a typical unavoidable activity of a postconventional consciousness that does seek rational answers to its questions or at least tries to get the answers by means of rational arguments.

Habermas is a pluralist in the sense that he accepts the fact of contextual rationality (which means that rational behaviour can be explained or understood in terms of socially accepted norms). He nevertheless recognizes the limitations of a contextual rationality, since it cannot provide us with a normative standard that will transcend the particular social context and by means of which we could measure relations of power and domination within society itself. In other words, a critical social theory has to go beyond the social context in

---

[101] Jurgen Habermas, _Philosophical Discourse of Modernity_, Chap. xi.

order to reflect critically on socially accepted norms. This can be achieved, according to Habermas, by means of discourse ethics.

Habermas' Justification of Principle (U)

Because we depend on linguistic communication that is subject to different interpretations, a practical discourse requires, according to Habermas, certain "unavoidable" presuppositions of argumentation that will render the discourse "valid".

Habermas borrows from Robert Alexy[102] in specifying the formal unavoidable conditions for argumentation as follows.

1. Every subject with the competence to speak and act (which for Habermas means every potentially rational subject) is allowed:

1(a) To take part in a discourse (this rule defines the participants; Habermas does include advocates for those who are not in a position to argue, but whose best interests are to be taken into account, such as children, mentally handicapped, etc.);

1(b) to question or introduce any assertion (speech act) whatever (which includes providing justification for doing so, or providing grounds for avoiding giving a

---

[102] Robert Alexy, "A Theory of Practical Discourse", in Seyla Benhabib and Fred Dallmayr (eds.), Communicative Ethics Controversy, pp.163-175.

justification). Alexy terms this the "general rule of justification;

1(c) to express his or her attitudes, desires, and needs. (Rules 1(b) and 1(c) grant equal opportunity for all to present their arguments);

1(d) no one may be prevented by internal or external coercion from exercising the foregoing rights (this rule specifies conditions to be prevented, which can be in any form, ranging from using lies in argumentation to the exclusion of some subjects from the discussion, in order to achieve an "agreement" or "consensus" with regard to a contested norm) (MCCA, 89-90).

The above mentioned rules are termed by Alexy "rules of reason"; they specify the "most typical preconditions for discourse theory's concept of rationality". These correspond to Kant's first rule of reasoning (see p.42 above), that of thinking for oneself (freedom from coercion, autonomy), except that in Habermas there is no private language (monologue), no predetermined moral insight to arrive at the end of such a moral reasoning, and each speaker must be allowed freedom of expression also for his or her wishes, desires and thoughts.

Habermas argues that these "rules" are not constitutive rules (such as in chess); rather, these are unavoidable presuppositions (intuitive knowledge, or a pre-understanding of the game of argumentation) that each subject who is competent to speak and act brings with him or her to

the process of argumentation. What this means is, that whoever enters into the argumentation game must assume these rules as conditions to be realized approximately for the purpose of argumentation. For whoever tries to contest or contradict or contravene these rules (conditions) is caught up in a "performative self-contradiction", in that by the very act of contesting (arguing against) these rules of the argumentation game, one is contradicting one's own assumptions in the very process of arguing against them (MCCA, 79-82).

The second set of unavoidable presuppositions of argumentation is related to the procedure of argumentation. Here the presuppositions concern a special form of interaction, that of mutual recognition of all the participants (responsibility, reciprocity, truthfulness) who are engaged in the genuine search for a normative validity. Some of Alexy' rules (conditions) in this regard are:

2(a) Every speaker may assert only what he or she really believes, namely, be sincere about what he/she asserts (although the assertions themselves may be hypothetical).

2(b) a person who disputes a proposition or norm not under discussion must provide reasons for doing so.

These conditions correspond to Kant's second rule of reasoning (see p.43 above), that of recognizing the other as an autonomous subject (the formula of the end-in-itself, reciprocity, impartiality). This "unrestrained competition for better arguments...[is] irreconcilable with traditional

ethical philosophies that have to protect a dogmatic core of fundamental convictions from all criticism". (MCCA ,88) Here we are reminded that discourse ethics applies to modern postconventional moralities and not to conventional, authoritarian or religious ones.

Alexy terms these rules "the rules of the burden of argumentation".

A third set of conditions is the requirement of minimal logical and semantic consistency, such as that one should not contradict oneself; that one should apply the same predicate to all resembling objects. This corresponds to Kant's third rule of reasoning (see p.42 above), that of logical consistency, which is a basic rule for "any linguistic communication in which what is in question is correctness or truth".[103]

Habermas considers the above mentioned conditions to be not conventions or definitions for prejudging the arguments and their conclusions but inescapable presuppositions for conducting a meaningful and valid discourse (theoretical or practical), and that accepting these presuppositions as unavoidable together with knowing what it means to justify a norm (providing good reasons under conditions of a legitimate procedure) entails a recognition of his universalizable principle (U). As mentioned above, U stipulates that a contested norm cannot meet with the consent of all

---

[103] Alexy, "A Theory of Practical Discourse", p.163.

participants "unless all affected can _freely_ accept the consequences and the side effects that the _general_ observance of a controversial norm can be expected to have for the satisfaction of the interests of _each individual_" (MCCA, 93).

Principle (U) is considered to be justified by Habermas on the grounds that rational argumentation presupposes equality of all concerned (equal opportunity) and impartiality (reciprocity, mutual respect). These are normative presuppositions that he does not try to justify but which are considered by him to be unavoidable.

> Demonstrating the existence of performative contradictions helps to identify the rules necessary for any argumentation game to work; if one is to argue at all, there are no substitutes. The fact that there are _no alternatives_ to these rules of argumentation is being proved; the rules themselves are not being _justified_. True, the participants must have accepted them as a 'fact of reason' in setting out to argue. But this kind of argument cannot accomplish a transcendental deduction in the Kantian sense. (MCCA, 95)

In this respect, there is a similarity with Kant but the 'fact of reason' has become a 'fact of historical reason' that Habermas believes to be, in effect, "a fact of natural history" (developmental-logic, as examined in chapter II) and which he regards as a weak transcendental justification of U (namely, the status is hypothetical).

The fact that principle (U) contains an unavoidable normative content reminds us of Kant's difficulties regarding the transcendental deduction in section 3 of the _Fundamentals Principles_... As it was pointed out there (Chapter I, pp.48-

49), the moral law could not be deduced from, nor justified by, a non-moral source such as freedom, and Kant arrived at the conclusion (in the second Critique) that our moral consciousness is a "fact of reason" that cannot be explained nor denied and is therefore objectively valid. Likewise, it seems, the justification of U would not be possible on the grounds of a non-moral source. We cannot avoid a moral content in the justification of the supreme moral principle, since this content itself constitutes what we mean by 'moral consciousness' . The modern (postconventional) consciousness is to be critical with regard to all moral norms but the criticism itself cannot be carried out without being guided by moral principles of equality and reciprocity, if the criticism is to make sense and be considered morally valid.

The term 'postconventional' is misleading to a certain degree. It is meant to suggest a critical attitude towards certain traditional or conventional norms, but it is not really a complete breakaway from the conventional attitude or from the lifeworld (as it may seem to imply). The 'postconventional' attitude itself becomes or may be said to have become "conventional". In the "conventional" stage, social norms appear to be "natural", so we can say that for a postconventional consciousness certain moral norms are not questioned because in Habermas' terms they are "unavoidable". This seems to be the case with regard to the democratic principles and the autonomy of the individual.

The principle of universal respect and egalitarian reciprocity are our philosophical clarification of the constituents of the moral point of view from within the normative hermeneutic horizon of modernity. These principles are neither the only allowable interpretation of the formal constituents of the competency of postconventional moral actors nor are they unequivocal transcendental presuppositions which every rational agent, upon deep reflection, must concede to. These principles are arrived at by a process of 'reflective equilibrium' in Rawlsian terms, whereby one, as a philosopher, analyzes, refines, and judges culturally defined moral intuitions in light of articulated philosophical principles. What one arrives at the end of such a process of reflective equilibrium is a 'thick description' of the moral presuppositions of the cultural horizon of modernity.[104]

Habermas' Conception of Normative Validity

There are two validating factors in Habermas' conception of normative validity: (a) Consideration of "consequences" as stated in principle (U); (b) the general or common will (consensus) in principle D. In the following, I shall examine each of these two factors.

1. **"Consequences" as a Validating Factor of Norms.**

On the face of it, it seems strange that a Kantian universalistic moral philosophy should consider normative validity to be in any way linked to the consequences that its observance will have on those concerned. This coulo be interpreted as a form of utilitarianism on the part of Habermas, something which contradicts Kant's position. The term, however, may be interpreted in two ways: 'Consequences' may be regarded as either

---

[104] Benhabib, The Communicative Ethics Controversy, p.339.

(1) **short term** consequences which will satisfy the immediate self-interest of an agent; or

(2) the **long-term** consequences of a moral order (or disorder) that will result when a norm is adopted as valid.

From the point of view of the long-term consequences, a short-term self-interest may indeed backfire if everyone were to follow his or her own self-interest regardless of others. As we saw (pp.32-33 above), although Kant is generally understood to mean that it is the agent's intention in his or her action that counts rather than the consequences, he can also be interpreted to mean that the agent should think of the social order (long-term consequences) that will result if everyone acted from self-interest alone (short-term).

In Kant, the will is motivated by two different sources, one source is subjective desires, self-interest, etc. (Willkur, the empirical will), the other is objective moral principles (Wille, the moral will). It is the moral will that thinks of the long term results of one's actions. Thus, 'consequences' in Habermas should not necessarily be interpreted as strictly utilitarian. Even for Kant, it is because there is a Willkur (self-interest) that it is important that Wille (moral considerations) should restrict it.

According to Seyla Benhabib, "consequences" in U causes confusion and is therefore redundant; for example,

116

masochists may consider inflicting suffering as desirable[105].
It seems to me that there is a confusion here of the levels of
the debate. In the case of Benhabib, she seems to have in mind
everyday communicative action with short-term solutions
(although even in this case, masochists will probably be a
minority). However, Habermas seems to have in mind the long-
terms consequences for all those affected. In a universal
debate, it seems more reasonable that the long-term
consequences be considered the motivating reason to agree to
accept as valid and observe certain norms. Although it is
important to keep in mind that discourse is a continuation of
everyday communicative action, the issue in discourse ethics
is the **validation** of norms of actions that have real
consequences for all.

2.   **The Common Will (Consensus) as a Validating Factor of
      Norms.**

        According to Habermas,

        What is expressed in normative validity is the authority
        of a general will shared by all concerned, a will that
        has been divested of its imperative quality and has taken
        on a moral quality. This will invokes a universal
        interest, which we can ascertain through discourse, that
        is, grasp cognitively from our perspective as
        participants [emphasis in the text]. (MCCA, 74)

        The general will in question here is not a pre-
determined, Kantian rational (common) will. It is the task of

---

        [105] Benhabib, The Communicative Ethics Controversy,
pp.342-344.

argumentation in discourse to bring about new moral insights which will then help to form a common will or consensus. Such a common will or consensus cannot be formed by each participant trying to reason monologically because the solitary thinker can be mistaken about his or her assessment of what the standpoint of the other might be[106]. What is crucial here is that impartiality is conceived by Habermas to be both an initial impartial will on the part of the **individual** (which was also in Kant) and an **intersubjectively shared** impartial consensus (in Kant, self-legislation for the kingdom-of-ends) that in Habermas can be arrived at only as **the outcome** of a real discourse. Thus, Habermas' procedure is open-ended in that it is designed to help bring about (not uncover) mutually accepted new moral insights by means of the more reasonable and convincing arguments.

It should be pointed out that this is an ideal situation in which a norm should be validated. There is no moral obligation to join a debate but the validity of the debate and its outcome is questioned if not all have participated. The ideal situation can be used as a yardstick to measure the extent of the validity, namely, the extent to which it is an expression of the general will.

It is important to understand Habermas' theory of argumentation since it is crucial for the validating procedure:

---

[106] Wellmer, "Ethics and Dialogue", p.141.

> We use the term __argumentation__ for that type of speech in which participants thematize contested validity claims and attempt to vindicate or criticize them through arguments. An __argument__ contains reasons or grounds that are connected in a systematic way with the __validity claim__ of a problematic expression. The 'strength' of an argument is measured in a given context by the soundness of the reasons; that can be seen in, among other things, whether or not an argument is able to convince the participants in discourse, that is, to motivate them to accept the validity claim in question. Against this background, we can also judge the rationality of a speaking and acting subject by how he behaves as a participant in argumentation, should the situation arise. (__TCA__ I,18)

It should be hamphasized that it is not a question of "persuasing" but of "convincing" arguments. 'Persuasion' may imply threatened force or misleading arguments, while 'convincing' implies that good arguments have been made for or against accepting a validity claim.

Following Aristotle, Habermas recognizes three levels of argumentation which he considers to be interconnected in practice: namely, argumentation as a **process**, as **production** of cogent arguments, and as a **procedure** (__TCA__ I, 25-26).

By a **process** of argumentation Habermas means the structure of the conditions presupposed to be satisfied when a competent speaker enters into argumentation. These conditions he terms an "ideal speech situation" where the purpose or motivation is to reach an understanding that is arrived at solely by means of the more convincing arguments, rather than by physical force or authoritarian power. This recalls Kant's Kingdom-of-Ends:

> Participants in argumentation cannot avoid the presupposition that... the structure of their communication rules out all external or internal coercion other than the force of the better argument and thereby also neutralize all motives other than that of the cooperative search for truth. (MCCA, 88-89)

Habermas considers this an "action oriented to reaching understanding". It should be pointed out that the "ideal speech situation" is a presupposition and not an actual situation. This distinction is important; without this presupposition, "validity" of a norm will not make sense (according to Habermas). One has to remember that this is an ideal, a desirable situation. In reality, one can speak of approximationg such a situation rather than achieving it; it is not a "truth" but a presupposition.

By **production** of arguments Habermas means the construction of the cogent argument (the logic of the argument), namely, the act of redeeming the validity claim in question. The logic of the argument in Habermas is an informal logic since discourse means, for Habermas, a real discourse which is contingent, and the logic of the argument will depend on the context (the web of feelings and attitudes, behaviour expectancy, or moral intuitions) of the lifeworld in which it is carried out.

Habermas argues that the concern for rational argumentation in ethics seems to focus either on the process of argumentation or the production of arguments. For example,

Steven Toulmin[107] focuses on the logic of the argument (production), and Wolfgang Klein[108] on the process (rhetoric) of argumentation (TCA I, 25-38). The result is that their analysis is not complete, according to Habermas, since it is the **intersubjective** exchange of the arguments that is crucial to determine the rationality of normative validity (i.e. whether there are good reasons for accepting a norm as valid). This implies that the perspectives of the participants in the discourse are to be taken into account, as well as the fact that participants intuitively know what it means to justify a norm. What is at issue here is that it is not only the attitude of the third person who observes and makes judgements, but also the first and second person's viewpoints which are unpredictable and are being formed in the very process of the give and take of arguments. This can only be achieved by the **procedure** of argumentation which, for Habermas, means that the different arguments freely compete for acceptance, and it is the resulting consensus or the common will that can be qualified as 'rational'.

It is here where practical or communicative rationality can be understood to give persons control over

---

[107] Steven Toulmin, The Uses of Argument (Cambridge, Eng., 1958);

Steven Toulmin, Human Understanding (Princeton, N.J., 1972).

[108] Wolfgang Klein, "Argumentation und Argument", Zeitschrift fur Literaturwissenschaft und Linguistik, 38/39 (1980): 9ff.

their fate, instead of being led by functional forces (system imperatives or uncritically accepted traditional norms, political domination, etc.) which are at work, so to speak, "behind the backs" of individuals and groups.

Principle (U) is only a rule of argumentation whose real content is to be provided in a real discourse (D) within a real lifeworld. What the general will is at any given time can only be discovered in a real discourse (in concrete situations) by means of arguments and not in advance of it.

> Discursively redeemable norms and generalizable interests have a non-conventional core; they are neither merely empirically found already to exist nor simply posited; rather they are, in a non-contingent way, both formed and discovered. This must be so if there can at all be anything like a rational will.[109]

In this conception of the "general will" one can discern Habermas' borrowed philosophical ideas, such as Hegel's historical consciousness, Husserl's phenomenology, American pragmatism, Wittgenstein's philosophy of language, etc. The idea of practical discourse is perhaps made clearer in that for Habermas it is (historical) moral consciousness in the making. The general will is not determined in advance but is formed through the debate. Morality is phenomenological (the web of feelings and attitudes, reflection on norms). The pragmatic influence is seen in Habermas' recognition of the timely concrete

---

[109] Habermas, as quoted in McCarthy, The Critical Theory of Jurgen Habermas, p.327.

(objective) problems involved in the validation of norms. What he criticizes in the above mentioned theories - such as Hegel's conception of the "collective subject", Husserl's conception of the transcendental ego, and the functionlist conception of rationality of pragmatism - is their limitations in that they all presuppose an instrumental/purposive rationality. Habermas tries to adopt these theories to his conception of practical reason (the moral point of view).

The demand in Kant's categorical imperative is: "Act only on that maxim through which you can at the same time will that it should become a universal law" (Fund., 38). The will in question is a moral-rational will (Wille) which means for Kant that it is objective and universal, in contrast to Willkur which is the expression of a subjective or an arbitrary will. It is action motivated by a rational (lawful) will which makes it, according to Kant, a valid moral action.

By a 'general will', Habermas means not "what each can will without contradiction to be a universal law" (Kant) but what all can will in agreement to be a universal norm.

> Whether a norm is universalizable, capable of rational consensus, can be ascertained only dialogically in unrestricted and unconstrained discourse. From this point of view Habermas's discourse model represents a procedural reinterpretation of Kant's categorical imperative: rather than ascribing as valid to all others any maxim that I can will to be a universal law, I must submit my maxim to all others for purposes of discursively testing its claim to universality. The emphasis shifts from what each can will without

contradiction to be a general law, to what all can will in agreement to be a universal norm. A rational will is not something that can be certified and secured privatim; it is inextricably bound to communication processes in which a common will is both "discovered" and "formed".[110]

In Kant, the rational will (Wille) is set in opposition to the subjective will (Willkur) because only in this way can the rational will be universalized. This is indeed required for a monological consciousness. In Habermas, the purpose of the discourse is also to arrive at a consensus with regard to what the generalizable interests and needs may be. In Kant, autonomy means self-legislation. For Habermas, autonomy means also self-realization for all and the recognition of all by all of their generalizable needs. What these neralized needs and interests are cannot be known in advance and for this reason a discourse is required. The general will means, in fact, both a willingness to participate in a discourse concerning the validity of a contested norm and also what the common interest of the participants is.

There may seem to be a paradox here in that discourse assumes both a reconciliation and a conflict. If there were no conflict, or contested norms, there would be no need for a discourse. Yet, discourse means that there is a general willingness to enter into a discourse in order to resolve conflicts by means of rational arguments and not by violence.

---

[110] McCarthy, _The Critical Theory of Jurgen Habermas_, pp.326-327.

Habermas' conception of the generalized will can easily be misinterpreted. The Enlightenment's conception of the individual's given self-interests is still influential today, with the result that the 'moral point of view' (impartiality) is understood by some moral theorists to be more in the sense of a fairly negotiated agreement, or a compromise, or a balance of power achieved between conflicting interests.

According to Seyla Benhabib, there are three ways of interpreting the concept of the "generalized will" or "general interests":

1) One is a minimal interpretation, which would describe a set of procedures such as "not taking interest in each other's interests, i.e. limited altruism" (Rawls) which Habermas rejects because it involves strategic rationality, in addition to being monological;

2) the other is a maximal interpretation, which would describe an actual social situation in which conflicts of interests among individuals disappear (Rousseau, Kant's Kingdom-of-Ends);

3) or an interpretation which has critical implications and helps to reveal partial or ideological interests that claim to be universal. It is this third interpretation of the "generalized will" that applies to discourse ethics[111].

---

[111] Benhabib, <u>Critique, Norm, Utopia</u>, p.311.

As we saw, there are two main supreme principles in Kant: One is the first formula (universal law), the other is expressed in the third formula (the principle of autonomy) where autonomy is conceived of as self-legislating (positive freedom) of the rational will that is in harmony with the kingdom-of-ends. Autonomy in Kant means both that it is a precondition of the moral law (an analytic principle) as well as a command of that law (a moral principle). Both imply the moral responsibility of the individual. Something similar seems to occur in Habermas with the difference that for Habermas the main moral principle is not Kant's first formula but the second (the end-in-itself) formula. The pluralities of the ends-in-themselves and their mutual recognition in a dialogue carried out by means of language is the first moral principle, and it contains both, autonomy as a precondition for discourse (rule of argumentation) and as a moral principle guiding the discourse (ideal [domination-free] speech situation).

Universality in Habermas is not imagined or imposed (as in Kant) but is to emerge by means of a universal dialogue carried out by all the autonomous "ends-in-themselves". It is the universal participation in the dialogue, and the agreement arrived at as a result, that makes it universal. As Benhabib puts it, "participation precedes universalizability. The old adage 'no taxation without representation' is now reformulated

as 'no universalizability without participation'".[112]

As it was indicated (see p.42 above), Kant's moral reasoning starts from form (universal law), to matter (plurality of ends), to the totality of all maxims (kingdom-of-ends). In Habermas, the procedure starts with matter (plurality of ends) from which the form (unity or universality emerges). Since the process is open-ended, there is no totality in Habermas in the sense of Kant's kingdom-of-ends. The ideal [domination-free] speech situation, which does resemble Kant's kingdom-of-ends, is an ideal condition presupposed or anticipated in the debate, not the end result of the debate. Kant's ethics is an ethics of negative duties but his kingdom-of-ends does suggest a state of positive duties (self-legislation, achieved reconciliation).

To sum up: Habermas' justification of discourse ethics is understood by him to be in terms of "unavoidable" presuppositions; indeed these presuppositions are unavoidable for a modern post conventional consciousness. Habermas does not justify them but regards them as a "fact of historical reason". His conception of normative validity includes two factors: 'consequences' (principle (U) and the 'common will' (principle (D)). These are also found in Kant, but Habermas gives them a different interpretation, namely, the common will cannot be known in advance, and the consequences must be acceptable by all those affected. In both cases it can be

---

[112] Benhabib, Communicative Ethics Controversy, p.315.

achieved only by means of a **valid** discourse whose presuppositions are, according to Habermas, unavoidable.

Kant's categorical imperative has been characterized as rationalist, universalist, and formalist. Habermas' discourse ethics can be characterized in the same way but it has to be differently interpreted. In what follows, I shall present some of the arguments in defense of discourse ethics advanced by sympathetic critics.

On the Rationality of Discourse Ethics

The argument against the possibility of a rational justification of discourse ethics is as follows: although democracy is based on a certain consensus of basic norms, this can not be said to constitute a rational justification or a rational consensus, so that Habermas' conception of an idealized lifeworld (unconstrained rational consensus, ideal [domination-free] speech situation) is meaningless.

In defense against such criticism, Albrecht Wellmer, for example, argues that what the critics do **accept** is enough to justify Habermas' position. The basic agreement on norms implied in the democratic principle (equality and reciprocity) can be considered to be rational, Wellmer argues, so long as no arguments (good reasons) are brought against it and individuals experience this form of life to be good or adequate. The question is whether communicative action, or consensual action, within a democratic framework is one of

many other possibilities for actions or not. Wellmer thinks that it is not, since once we believe that rational argumentation about norms is possible, granting equal rights and liberties to everybody is accepting a principle of consensual coordination. If the principle of rationality is combined with the democratic principle then, according to Wellmer, a principle of consensual coordination becomes unavoidable in so far as we believe in the possibility of rational argumentation. If there is a normative disagreement, then some kind of agreement arrived at (by means of argumentation) without force will be considered to be "fair" or "just" or "rational".[113]

The question Wellmer poses is whether, when we accept the principle of rational consensus within the democratic framework, we can draw a line beyond which we can say that the idea of a rational agreement does not make sense any more. If we did draw such a line, then to deny rationality to the basic consensus of the democratic principle itself would be arbitrary (or dogmatic). Wellmer considers that the basic consensus of the democratic principle deserves to be called rational if nothing in the principle is exempt from the possibility of critical examination.

The unavoidable presuppositions that Habermas puts forth as a justification demand both 1) recognition of all

---

[113] Albrecht Wellmer, "Reason, Utopia, and Enlightenment", in Richard J. Bernstein (ed.), Habermas and Modernity (Cambridge, Mass., MIT Press, 1985), p.59.

speaking and acting subjects as participants in a discourse (Benhabib calls it "the principle of universal moral respect", and 2) freedom for all to initiate various topics and arguments including challenging the presuppositions of discourse ethics itself (Benhabib calls it "the principle of egalitarian reciprocity"). As she argues, "the racist, the sexist, the bigot can challenge" discourse ethics but they will have to convince the opponents with good reasons why they want to exclude some from participation in the discourse. The rules of argumentation are pragmatic rules that are required in order to continue the discursive procedure. They cannot be given up altogether without resorting to violence, coercion or suppression[114]. This argument can also be advanced against the charges of the dogmatism of discourse ethics, since the "dogma" itself can be challenged within discourse ethics itself. Thus, these rules can not be said to be dogmatic.

No line, then, can be drawn between internal agreement (within the system) and external agreement (judging the system as a whole). It should be remembered, Wellmer points out, that we do not deal here with geometrical reasoning but with practical (basic interaction) reason. The utopian perspective inherent in a democratic tradition is a center of gravitation, a force that becomes stronger since mutual recognition is already embodied in consensual coordination.[115]

---

[114] Benhabib, Communicative Ethics Controversy, p.340.

[115] Wellmer, "Reason, Utopia, and Enlightenment", p.57.

On The Universalism of Discourse Ethics

There is a question of whether discourse ethics is only one possibility among other universalistic moral principles at the postconventional level (neo-Kantians on one side, and ethical decisionism on the other). The dilemma here is, as Benhabib sees it, that the ideals of discourse ethics are usually seen as either a "fact of reason" (naturalism, ultimate justification) or they imply a sheer choice (decisionism). That there is another possibility, that of a fallibilist communicative rationality, is not considered in the above arguments. As she sees it, ethical decisionism seeks to "eliminate the burden of choice of the particular individual", an attitude which may in fact result in an ethics of authority (of the moral theorists) or dogmatism.

> Critical theory, precisely because it assumes that the moral individual is as autonomous as the theorist, that theoretical enlightenment bestows no prima facie moral and political authority, and that moral insight can be shared by all, cannot, in principle, answer the demand of decisionism that it eliminate the burden of choice of the individual. To think otherwise would be 'bad faith'. Thus, a weak justification of communicative ethics, based on a discursive and fallibilistic conception of reason, would not need to fall into decisionism; for it is shown that what decisionism asks of normative theorizing, as well as the search for 'ultimate' grounds, is itself unreasonable.[116]

Universalistic ethics seems to be a characteristic of a postconventional moral consciousness. Since the traditional

---

[116] Benhabib, Critique, Norm, Utopia, p.327.

(religious, metaphysical) grounds for justifying moral norms have become questionable, there seems to be only one possibility left and that is a universalistic ethics that does in fact relativize all socially accepted norms of all cultures. A universalized norm, according to Habermas, is one that has been recognized as such by all and it is U which provides the test for the universal validity of the norm (it should be noted that U excludes all norms whose observance may have adverse consequences for some). The problem is how to reconcile the idea of plurality of concrete forms of life (cultural relativism) with the idea of a universalistic principled ethics.

In this respect Kohlberg makes a relevant observation, namely, that it is important to make a distinction "between the idea that 'everyone has their own values', and the idea that 'everyone ought to have their own values'". The former is a factual statement that can be termed "cultural relativism", while the latter is a value statement which can be termed "ethical relativism". This confusion can also be expressed as "there are no universal values" and then to infer from it that "there ought not to be any universal human values; every person or culture ought to do its thing".[117] The latter is a principle which can be considered to be as dogmatic and universalistic as any other absolute universal moral principle.

---

[117] Kohlberg, "From Is to Ought", pp.156-158.

Universalistic ethics enjoins that all people of all cultures be considered equal and that they are entitled to universal moral respect, yet this does not mean that all moral systems are equally just or justified. One culture may subscribe to universalistic ethics, while others not, but this does not mean that universalistic ethics is relative to the culture. As Seyla Benhabib argues, the traditional opposition between universalism and historicism may no longer be compelling; she suggests that we should think in terms of a 'historically self-conscious universalism'.

> All human communities define some 'significant others' in relation to which reversibility and reciprocity must be exercised - be they members of my kin group, my tribe, my city-state, my nation, my coreligionists. What distinguishes 'modern' from 'premodern' versions of universalistic ethical theories is the assumption of the former that the moral community is coextensive with all beings capable of speech and action, and potentially with all humanity. In this sense, communicative ethics sets up a model of moral conversation among members of a modern ethical community, for whom the theological and ontological basis of the inequality among humans has been radically placed into question.[118]

On The Formalism of Discourse Ethics

The "ideal speech situation", or an ideal community of communication (Apel, 1980), means that the conditions of symmetry (equal opportunity for all) and impartiality (mutual recognition and respect) are fulfilled. As already mentioned, the ideal speech situation recalls Kant's Kingdom-of-Ends. In

---

[118] Benhabib, _The Communicative Ethics Controversy_, pp.339-340.

Habermas' case, however, the ideal speech situation is also a normative standard for social criticism whose task is to unmask mechanisms of constraint, domination, etc. that are actually at work in society. For him, the ideal community of communication is a **form** of life of unconstrained and undistorted intersubjectivity. In this sense it is a **regulative** idea. What this means is that "only the _form_ of the process of reaching agreement and consensus is made the criterion of normative rightness", namely, "conditions are now normatively distinguished under which a rational, i.e., discursively orientated, consensus would be factually possible".[119]

As Wellmer argues, with this idea Habermas transcends both Kant and the social contract theory. Habermas transcends Kant in that the moral principle is dialogical and open-ended. In this it is recognized that in the monological model it is possible to be **mistaken** in one's assessment of the standpoint of the other, and that this assessment should be tested in a real interpersonal dialogue. Habermas transcends the social contract theory in that discourse ethics

> does not establish in advance the contents of a possible rational consensus and thereby, at the same time, substantive criteria for a distinction between what is 'reasonable' and what is 'unreasonable', but rather, conversely elevates the discursive procedure itself into

---

[119] Wellmer, _Communicative Ethics Controversy_, p.303.

the criterion for the rationality of any possible consensus".[120]

The only normative content that principle (U) contains is a relationship of a domination-free reciprocal recognition. As Wellmer sees it, "a fundamental Kantian intention is better realized than Kant himself was able to do in his ethics: namely, the elaboration of a _formal_ determination as the solely a priori realizable content of moral consciousness".[121] In this way, Wellmer claims, the form of the principle does not allow socially accepted norms to be regarded as a content that should not be questioned because it takes place behind the backs of the participants. Rather, all socially accepted norms become the subject-matter for a possible critique and possible discursive transformation. This means that Hegel's criticism of Kant's formalism (that any substantive norm may fit into it) does not apply to Habermas' discourse ethics.

However, as Wellmer also argues, the formal structure of the ideal speech situation does not in itself constitute an independent normative standard for rationality and he claims that such an independent standard is not necessary. A 'rational consensus' means that it was achieved by moral insight (good reasons) and not by means of deception or

---

[120] Wellmer, _The Communicative Ethics Controversy_, pp.311-312.

[121] Wellmer, _The Communicative Ethics Controversy_, p.312.

coercion. An agreement can always be only a factual agreement; the ideal speech situation always presupposes a certain substantive pre-understanding of 'rationality'. To have doubts about the rationality of a factual consensus means that either specific substantive counterarguments can be advanced against it, or that doubts will be expressed as to the rationality of those consenting.

> But what we mean by the 'rationality' or 'autonomy' of speakers cannot be sufficiently grasped with the aid of the structural features of an ideal speech situation. Therefore, in hidden doubts of this kind there is always a hypothesis which can ultimately only be tested by producing a new 'rational' consensus - a state of affairs out of which those consenting can recognize as such their previous lack of rationality.[122]

Thus, while principle (U) is a formal principle from which nothing direct can be deduced, it is nevertheless not "empty", since it does contain a normative content: Universal respect and reciprocity. This normative content is the filtre, so to speak, which does not allow immoral norms to be validated by a concrete discourse. What we do seem to have, in fact, is "an ethical orientation toward structures of inequality such that those structures are, at least initially, always to be brought under interpretations which illuminate them as possible structures of power"[123]. In this case, it seems that we can say that discourse ethics is at least an ethics of negative duties which is also that of Kant.

---

[122] Wellmer, <u>Communicative Ethics Controversy</u>, p.324.

[123] White, <u>The Recent Work of Jurgen Habermas</u>, p.77.

Another reason why discourse ethics is not entirely empty is that the degree of reciprocity that is necessary for discourse is not externally imposed but rather internally presupposed by the participants themselves when they enter into a discourse. This in itself may not avoid clashes between individuals and the demands of the form of discourse but it cannot be claimed that discourse ethics is so abstract as to be illegitimate.[124]

As Benhabib argues, it is the form of the procedure which determines the legitimacy of the agreement, rather than the consensus itself. One of the problems of U, she claims, is that for Habermas it has the effect of guaranteeing consensus, yet

> consent alone can never be a criterion of anything,
> neither of truth nor of moral validity; rather, it is
> always the rationality of the procedure for attaining
> agreement which is of philosophical interest. We must
> interpret consent not as an end-goal but as a process for
> the cooperative generation of truth or validity. The core
> intuition... [is] that these principles have been adopted
> as a result of a procedure, whether of moral reasoning or
> of public debate... Consent is a misleading term for
> capturing the core idea behind communicative ethics:
> namely, the processual generation of reasonable agreement
> about moral principles via an open-ended moral
> conversation.[125]

Still, one can argue that the very purpose of a discourse is to arrive at a genuine consensus, and the presupposition of the participants must be that it is, at

---

[124] White, The Recent Work of Jurgen Habermas, p.75.

[125] Benhabib, The Communicative Ethics Controversy, p. 342.

least in principle, possible, otherwise the very meaning of the discourse and the motivation for it will become questionable[126]. As Wellmer argues, what is meant by a 'genuine' consensus is "that individuals know that their common interest is recognized in the institutions of society and thereby at the same time recognize one another reciprocally as free and equal persons; but we could not say what its concrete content was".[127]

What can be concluded from the above, however, is that it is the form of the exchange of the arguments under conditions of equality and reciprocity which is the basis for considering a consensus to be valid, genuine or legitimate. What seems to be more important is, in my view, the understanding that a consensus which undermines or contradicts the form of the procedure of discourse cannot be considered to be 'legitimate' or 'rational' or 'just'.

[126] McCarthy, The Critical Theory of Jurgen Habermas, p.306.

[127] Wellmer, The Communicative Ethics Controversy, p.317.

CONCLUSION

The aim of Habermas' critical social theory is to
uncover relations of power and oppression which are a threat
to individual and moral autonomy. His discourse ethics is a
formal moral principle which is to serve both as a guide for
practical discourse and as a criterion for social criticism.

Discourse ethics can be seen as an attempt to bridge
the gap between radical universalism and radical pluralism.
A worldview of radical plurality that would not allow for
universalistic principles seems to be as dogmatic (and
universalistic) as that of radical universality that does not
recognize contingent conditions and plurality.

Although discourse ethics does not supply us with a
standard that we can apply directly to reality, it does offer
an ethical direction. The contingent conditions must always be
taken into consideration but the form should ensure that the
priority of practical reason holds over system imperatives, so
that individuals can and do have some control over their
behaviour and fate.

The aim of this thesis was to bring out similiarities
and differences in Kant's and Habermas' conceptions of
validity and validation moral norms. Discourse ethics can be
seen as a critical social scientific interpretation of the
categorical imperative. Kant's main ideals of morality, such

as practical reason (a capacity for moral judgements in all human beings), personal autonomy, plurality of ends-in-themselves, a formal structure of human interrelations (regardless of cultural content), are "unavoidable" ideals for a postconventional consciousness.

The tension that characterizes Kant's categorical imperative, in the sense that it can be understood to be both a theoretical statement about our moral consciousness and a moral criterion, is also evident in Habermas. Kant was anxious to prove the objective validity (the "is") of the categorical imperative, even while he believed it to be a valid moral criterion. Likewise, we see in Habermas a normative approach (utopian, what "ought to be") combined with his theoretical conception of moral development (reconstructive science, what "is").

What is involved, in Kant and in Habermas, is a basic conception of what it means to be human. As Kant and Habermas see it, the human being has self-interests but also a moral consciousness. The question for both is not whether actions should be guided solely by moral consciousness. They both recognize that human actions are motivated by self-interest but they also believe that actions should, at the same time, be guided by a moral consciousness as well. However, what differentiates Habermas' view from Kant's is his conceptions of consciousness and rationality. For Kant, consciousness is a-historical, the moral order is given and so is human

"nature". In Habermas, counsciousness is culturally constituted, rationality is contingent and linguistically mediated, and human needs and interests are culturally interpreted or created.

What is common to both Kant and Habermas is that the moral point of view requires an impartial attitude when it comes to moral reasoning. What differentiates between them is how the impartial attitude is applied. In Kant, one is to consider oneself as a legislating member of the kingdom-of-ends. This is what is meant to be autonomous. In Habermas, one is a member of a shared concrete historical culture; to be autonomous means to be self-critical (with regard to one's own interpretation of needs) and critical with regard to the validity of social norms. It is in this sense that participation in a discourse makes one a "co-legislator". In both cases, however, it is meant that human beings can have some control over their behaviour and fate.

BIBLIOGRAPHY

Allison, Henry. <u>Kant's Theory of Freedom</u>. Cambridge: Cambridge University Press, 1990.

Ameriks, Karl. "Kant's Deduction of Freedom and Morality", <u>Journal of the History of Philosophy</u>, 19 (1981): 53-79

Apel, Karl Otto. "The Problem of Philosophical Fundamental Grounding in light of a Transcendental Pragmatics of Language", in Kenneth Baynes and others (eds.). <u>After Philosophy</u>. Cambridge, Mass: MIT, 1987.

Apel, Karl Otto. "The a priori of the Communication Community and the Foundations of Ethics: The Problem of a Rational Foundation of Ethics in the Scientific Age", in his <u>Towards a Transformation of Philosophy</u>. London: Routledge and Kegan, 1980.

Aune, Bruce. <u>Kant's Theory of Morals</u>. Princeton University Press, 1979.

Baynes, Kenneth. <u>The Normative Grounds of Social Criticism: Kant, Rawls, Habermas</u>. New York: State University of New York Press, 1992.

Beck, Lewis White. <u>A Commentary on Kant's Practical Reason</u>. Chicago, University of Chicago Press, 1960.

Benhabib, Seyla. "Communicative Ethics and Current Controversies in Practical Philosophy", in Seyla Benhabib and Fred Dallmayr (eds.). <u>The Communicative Ethics Controversy</u>. Cambridge, Mass.: MIT, 1990, pp. 330-367.

Benhabib, Seyla. <u>Critique, Norm, Utopia : A Study of the Foundation of Critical Theory</u>. New York: Columbia University Press, 1986

Bernstein, Richard J. <u>Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis</u>. Philadelphia: University of Pennsylvania Press, 1983.

Bernstein, Richard J. (ed.). <u>Habermas and Modernity</u>. Cambridge, Mass.: MIT Press, 1985.

Clement, Grace. "Is the Moral Point of View Monological or Dialogical: The Kantian Background of Habermas' Discourse Ethics" <u>Philosophy Today</u>, 33 (Sum., 89): 159-173.

Dallmayr, Fred and Benhabib, Seyla (eds.). The Communicative Ethics Controversy. Cambridge, Mass.: MIT, 1990

Habermas, Jurgen. Communication and the Evolution of Society, trans. Thomas McCarthy. Boston: Beacon Press, 1979.

------. "Justice and Solidarity: On the Discussion Concerning Stage 6", in Thomas E. Wren (ed.). Moral Domain: Essays in Ongoing Discussion Between Philosophy and the Social Sciences. Cambridge, Mass: MIT Press,1990, pp.224-50.

------. Moral Consciousness and Communicative Actions, trans. C. Lenhardt and S.Weber Nicholsen. Cambridge, MIT, 1990.

------. Philosophical Discourse of Modernity: Twelve Lectures, trans. Frederick Lawrence, Cambridge, U.K.: Polity Press, 1987.

------. The Postmetaphysical Thinking: Philosophical Essays, trans. William Mark Hohengarten. Cambridge, Mass.: MIT Press, 1992.

------. "A Reply to my Critics", in John B. Thompson and David Held (eds.). Habermas: The Critical Debates, Cambridge, MIT Press,1982.

------. Theory of Communicative Action, trans. Thomas McCarthy. Boston: Beacon Press, 1984. 2v.

------. "Towards a Theory of Communicative Competence", Inquiry, 13 (1970): 360-375.

Honneth, Axel and Joas, Hans (eds.). Communicative Action: Essays on Jurgen Habermas' "Theory of communicative Action, trans. J.Gaines and D.L.Jones. Cambridge, Mass: MIT Press, 1991.

Kant, Immanuel. The Critique of Practical Reason, trans. Lewis White Beck. Indianapolis: Bobbs-Merrill, 1956.

------. The Fundamental Principles of the Metaphysics of Morals. trans. Thomas K Abbott. Indianpolis: Bobbs-Merrill Co. 1949.

"Kant's Practical Philosophy", Monist, 72 no.3 (1989), 311-483 (a series of articles about Kant's practical reason).

Kohlberg, Lawrence. "From Is to Ought: How to Commit the Naturalistic Fallacy and Get Away With it in the Study of Moral Development", in Theodore Mischel (ed.).

<u>Cognitive Development and Epistemology</u>. New York: Academic Press, 1971, pp.151-235.

------. <u>Essays on Moral Development</u>. San Francisco: Harper and Row, 1981, 2v.

McCarthy, Thomas. <u>The Critical Theory of Jurgen Habermas</u>. Cambridge: MIT, 1978.

Moore, Jennifer. "Kant's Ethical Community", <u>Journal of Value Inquiry</u>, 26 (Jan.90): 51-71.

Paton, H.J. <u>The Categorical Imperative: A Study in Kant's Moral Philosophy</u>. London: Hutchinson of London, 1963.

Rawls, John. "Kantian Constructivism in Moral Theory" <u>Journal of Philosophy</u>, 77 no.9 (Sept., 1980): 515-572.

Rehg, William. "Discourse and the Moral Point of View: Deriving a Dialogical Principle of Universalization", <u>Inquiry</u>, 34 No.1 (March, 91): 27-28.

Silber, John R. "Procedural Formalism in Kant", <u>Review of Metaphysics</u>, 28 no.2 (1974): 197-336.

Thompson, John B. and David Held (eds.). <u>Habermas: The Critical Debates</u>. Cambridge, MIT Press, 1982.

Wellmer, Albrecht. "Practical Philosophy and the Theory of Society", in Seyla Benhabib and Fred Dallmayr (eds.). <u>The Communicative Ethics Controversy</u>. Cambridge, Mass.: MIT Press, 1990, pp.293-329.

------. "Reason, Utopia, and the Dialectic of Enlightenmnet", in Richard J. Bernstein (ed.). <u>Habermas and Modernity</u>. Cambridge, Mass.: MIT Press, 1985, pp.35-66.

------. "Ethics and Dialogue", in his <u>The Persistence of Modernity: Essays on Aesthetics, Ethics, and Postmodernism</u>, trans. David Midgley. Cambridge, Mass.: MIT Press, 1991, chap. iv.

White, Stephen K. <u>The Recent Work of Jurgen Habermas: Reason Justice, and Modernity</u>. Cambridge, Cambridge university Press, 1988.

------. "Habermas' Communicative Ethics and the development of Moral Consciousness", <u>Philosophy and Social Criticism</u>, 10 no.2 (1984): 25-47.

Williams, T.C. <u>The Concept of the Categorical Imperative: A</u>
<u>Study of The Place of the Categorical Imperative in</u>
<u>Kant's Ethical Theory.</u> Oxford: Clarendon Press, 1968.