

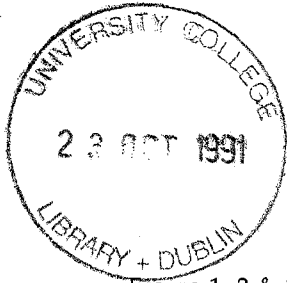
**The
Persistence
of
Modernity**

Essays on Aesthetics,
Ethics and Postmodernism

ALBRECHT WELLMER

Translated by David Midgley

Polity Press



coll
u09105092

English translation © Polity Press 1991
Essays 1, 2 & 3 first published in Germany in *Zur Dialektik von Moderne und Postmoderne. Vernunftkritik nach Adorno* © Suhrkamp Verlag, Frankfurt am Main 1985

Essay 4 first published in Germany in *Ethik und Dialog. Elemente des moralischen Urteils bei Kant und in der Diskursethik*, © Suhrkamp Verlag, Frankfurt am Main 1986

This translation first published 1991 by Polity Press in association with Basil Blackwell

Editorial office:
Polity Press, 65 Bridge Street, Cambridge CB2 1UR, UK

Marketing and production:
Basil Blackwell Ltd., 108 Cowley Road, Oxford OX4 1JF, UK

All rights reserved. Except for the quotation of short passages for the purposes of criticism and review, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

Except in the United States of America, this book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

ISBN 0 7456 05389

British Library Cataloguing in Publication Data
A CIP catalogue record for this book is available from the British Library.

This book is printed on acid-free paper
Typeset in 10½ on 12 pt Palatino
by Photo-graphics, Honiton, Devon
Printed in Great Britain by T.J. Press, Padstow, Cornwall

Contents

| | |
|--|-----|
| Introduction | vii |
| 1 Truth, Semblance, Reconciliation: Adorno's Aesthetic Redemption of Modernity | 1 |
| ✓2 The Dialectic of Modernism and Postmodernism: The Critique of Reason since Adorno | 36 |
| 3 Art and Industrial Production: The Dialectics of Modernism and Postmodernism | 95 |
| 4 Ethics and Dialogue: Elements of Moral Judgement in Kant and Discourse Ethics | 113 |
| Notes | 232 |
| Index | 257 |

86-92

– then relations between people are likely to forfeit a crucial dimension of their language, and their purposes will become ‘speechless’.

The double transgression that Octavio Paz saw in the craft mode of production – passing beyond the cult of utility and beyond the religion of art – cannot be carried into the heartland of industrial production by means of a direct reunion of art and industry as it was envisaged by the founders of the Werkbund. But it might well be conceivable that industrial production could be *made to abide* by purposes established through a process of communicative clarification, and that art and aesthetic imagination could become *implicated* in the communicative clarification of mutual purposes. It would then perhaps be possible for art and industry to come together as moments of an industrial culture through the mediation of a third element, i.e. in the medium of enlightened democratic praxis.

4

Ethics and Dialogue:

Elements of Moral Judgement in Kant and Discourse Ethics

Introduction

The scepticism of moral philosophy on the one hand and revolutionary humanism on the other are natural offspring of enlightenment. This is already true to some extent of the Enlightenment period in ancient Greece, and it is much more obviously true of the modern European Enlightenment. In both instances, Enlightenment means the discovery that seemingly guaranteed norms of correct living, the ‘justification’ for which lay in the order of things, the will of God, or the authority of tradition, have no conceivable foundation other than in the will of men. I imagine that this discovery must have induced a sense of vertigo in those who first made it, a sense of vertigo in which a wide variety of factors may well have been present. They would have felt that their existence was being shaken to its foundations. They would have glimpsed a freedom which was either chilling or exhilarating. Or they may have discovered that the social order in which they lived was founded on coercion, suppression and illusion. And according to the viewpoint or the social position of those who were infected with this enlightened consciousness, one or another of these factors will have dominated. Philosophical scepticism, conservative cynicism, and revolutionary humanitarian pathos are all equally possible as reactions to the discovery of this Enlightenment.

Cynicism (the ‘black’ version of scepticism) is a psychological

and moral problem rather than an epistemological one, and will therefore not be dealt with here.¹ ~~Scepticism and revolutionary (or at least universalistic) humanism on the other hand are, from an epistemological point of view, alternative responses to the discovery of the Enlightenment. Scepticism disputes the possibility of a new foundation for morality, whereas revolutionary humanism sees such a possibility in the united will of rational beings.~~ For the time being I do not wish to elaborate on this distinction, although we shall return to the subject of philosophical scepticism later. What interests me in the first instance is the fate – the philosophical fate – of revolutionary humanism. Not that I want to retell the whole story, rather I wish to examine revolutionary humanism in two of its most advanced versions with a view to drawing conclusions about its possible (philosophical) fate. By ‘advanced’ I mean advanced for their times. And the epithet ‘revolutionary’ is intended to indicate a *philosophical connection* between the humanism we are discussing here and the revolutions of modern times; it does not say anything about the subject of investigation itself. What we are discussing is not the theory of revolution, but (universalistic) ethics.

The two positions I wish to investigate are the formal ethics of Kant and the discourse ethics developed by Habermas and Apel. These are both forms of a universalistic ethics of reason or, as Habermas would say, of ‘cognitivist’ ethics. Both positions are characterized by the pursuit of a *formal principle* as a basis for ethics, which by virtue of its formal character is also a universalistic principle. Moral validity is here grounded in a rational *procedure* which, because it *on the one hand* characterizes some such thing as a core of rationality universally shared by rational beings, *on the other hand* relates to all rational beings as in a fundamental sense free and equal individuals. The question of universal validity is inextricably linked with the universalistic character of the moral principle itself: on this point Habermas and Apel are in fundamental agreement with Kant and, at least as far as the concept of legal ‘legitimacy’ is concerned, with the revolutionary doctrine of natural law. It is precisely in this sense that the authors we are dealing with belong in the camp of Enlightenment humanism.

In what follows I shall not be giving an exhaustive account of the positions in moral philosophy adopted by Kant, Apel or Habermas. The purpose of my analyses and interpretations is

more limited than this. My interest in Kant’s ethics is rather of a heuristic nature. My selective interpretation of Kant is intended to bring out the strengths and weaknesses of his ethics in such a way as to make clear both the motives behind the development of a ‘communicative’ or ‘discourse’ ethics and also those aspects of such a theory which stand in need of demonstration. I am clear from the outset that Kant’s ethics cannot be defended in their entirety, and that is why I have sought to bring out their *strengths* in particular. Some readers might object that I have behaved rather like a temple-robber on occasions, rescuing those items which particularly appeal to me. But in dealing with discourse ethics, on the other hand, I have treated with all seriousness its claim to have developed a system which solves Kant’s problems by means of a universal or transcendental pragmatics and thus ‘sublates’ Kant’s form of ethical universalism in a dialogic form of universalism. I do not believe that discourse ethics has yet achieved this ambition, and that is why I proceed more rigorously towards it than I do towards Kant’s ethics. I use the latter, so to speak, as a yardstick which may itself have become questionable, but which can nevertheless perform useful service when it comes to judging the potential for problem-solving of theories which themselves claim to have ‘sublated’ Kant’s ethics.

In criticizing Kant, discourse ethics has concentrated on three weak points in his ethics. Firstly it has criticized the formally *monologic* character of Kant’s moral principle which, contrary to Kant’s own opinion, leaves unanswered the question of the possibility of *intersubjectively valid* moral judgements. Secondly it has criticized the *rigorism* of Kant’s ethics with its singularly formalistic hypostatization of the concept of law. And thirdly it has criticized Kant’s attempts at a philosophical justification or *grounding* of his moral principle. Discourse ethics attempts to overcome these three weaknesses of Kant’s ethics by ‘sublating’ Kant’s formally *monologic* universalism within a formally *dialogic* universalism. It does this firstly by reformulating the moral principle in such a way that what is demanded of morally valid maxims is not that *I*, but that *we* can will them to have the validity of universal laws. Secondly it formulates the moral principle in such a way as to permit the question of right action to be understood as a question of rational interaction between individuals, all of whom possess their particular needs and vulnerabilities: this is how it excludes the fixed attitudes of ethical rigorism.

Thirdly and finally, discourse ethics claims, with its reformulation of the moral principle, to have made it possible to achieve a new form of fundamental grounding: Apel and Habermas have tried to show that the moral principle is founded on universal structures of argumentation. Now my own objection to discourse ethics in its current form is, in a word, that it has remained *too Kantian* on the one hand, while departing *too far from Kant* on the other. Discourse ethics has remained too close to Kant in two respects, namely its dependence on assumptions of consensus theory and its programme of an ultimate or fundamental grounding (*Letztbegründung*). On the face of it, of course, these two aspects of discourse ethics have little to do with Kant. But as I shall try to show, in terms of the kind of problem they present us with, the construction of concepts describing *ideal* conditions in consensus theory and the attempt to derive a universalistic ethics, as it were, directly from universal structures of reason, without the mediating instance of a *history* of moral consciousness, are both Kantian. To be sure, I am not saying that it is possible for us today to retrace the path that leads from Kant to Hegel. Although Hegel was the first to show, with the utmost clarity, how Kant's moral philosophy leads to dead ends, any attempt to avoid these dead ends – while not ignoring Hegel's *criticisms* – should steer well clear of Hegel's *system*. In place of an interpretation of a universalist ethics of dialogue in terms of consensus theory (which echoes Kant's notion of a 'kingdom of ends') I propose a *fallibilistic* interpretation; in the place of a strong and one-dimensional claim of justification I propose one that is weak as well as multi-dimensional. Once moral consciousness has become universalistic in character, it really does not need to live in anticipation of a state of reconciliation (however this might be formally described), nor does it need the assurance of an ultimate philosophical grounding. I believe rather that as long as universalistic ethics clings to these two absolute notions it remains just as vulnerable to the objections of Hegel as it does to those of the sceptics. With my criticism that discourse ethics remains too close to Kant, then, I am proceeding on the assumption that ethics needs to advance beyond the false antithesis of absolutism and relativism, which is to say that morality and reason do not stand or fall with the absolutism of ultimate agreements or fundamental groundings.

When I say that discourse ethics is *not Kantian enough* I mean

that it fails to come up to what Kant had already achieved in terms of differentiation. I am thinking in particular of differentiation between problems of morality and problems of law. There can be no doubt that it was Kant's intention to link law and morality together, but at least he distinguished analytically (and with good reason in my view) between problems concerning the *legitimacy* of norms and the problem of morally right action. I am not concerned with the details of Kant's construction of the connection between law and morality (which are often enough problematic in themselves), but rather with the manner in which Kant, through his formulation of the moral principle, distinguishes between questions of morally right action and question concerning the justice of norms. Discourse ethics has not yet matched the degree of differentiation that Kant achieved in this respect; and this is so for reasons connected with its consensus-theoretical premises. Both the fact that discourse ethics remains too close to Kant *and* the relative loss of differentiation it shows by comparison with Kant are, then, connected with the problematic assumptions of a consensus theory of truth.

The basic intuitions of discourse ethics, which I wish to defend, also concern its position towards Kant. I consider the criticism of the formalistic – monologic rigorism of Kant's ethics, as well as the attempt to go beyond this rigid formalism by a dialogic expansion of ethics, to be entirely justified. Finally, like Apel and Habermas, I perceive a connection between the transition from a formalistic to a dialogic ethics on the one hand, and the transition from a philosophy of consciousness to a philosophy of language on the other. To be sure, I believe that we need to determine anew those points in Kant's ethics which may be considered as points of departure for a dialogically conceived ethical universalism. This is what I propose to do in the first part of my essay. Part 2 contains a critique of discourse ethics and its consensus-theoretical premises. In part 3 I wish to show how the basic intuitions of discourse ethics can be reformulated in terms of the 'quasi-Kantian' perspective developed in part 1.

One final prefatory comment on the problem of scepticism in moral philosophy. I believe that there is just as good a case for taking it seriously as there is for not taking it seriously. It should not be taken seriously as a *moral* attitude, but it should be taken seriously as a way of questioning rationalist and foundationalist claims to knowledge. By this I mean that I believe that rationalism

can absorb scepticism and convert it into a catalyst of the enlightenment process. A rationalism enlightened by scepticism would be neither rationalistic nor sceptical, but it might perhaps be reasonable. I believe, then, that our best prospect for continuing the tradition of the Enlightenment, and of revolutionary humanism, is to bid farewell to certain ideals of reason. This does not mean bidding farewell to reason itself; rather it means bidding farewell to a false conception that reason has of itself.

The critique of Apel and Habermas in part 2 of this essay should also be seen to some extent as self-criticism on the part of the author, although I have not taken the trouble to indicate precisely where this is the case. It will be apparent to the reader, however, that while criticizing Apel and Habermas on particular issues I am also indebted to them both for ideas which have had a decisive and lasting impact on my own thinking.

I A Kantian Exposition

In various places, and most recently in his essay 'Diskursethik – Notizen zu einem Begründungsprogramm'² (henceforward cited as DE), Jürgen Habermas has drawn an analogy between the universalization principle in ethics and the so-called induction principle of empirical science. For reasons which should become clear later, I consider this analogy to be problematic, although the force of it immediately becomes apparent if we understand it in a weaker sense as has variously been put forward by M. G. Singer³ and R. M. Hare⁴, namely as a 'generalization principle' which is constitutive of both causal and moral judgements.

This generalization principle expresses the respective general character of causal and normative judgements; and of the corresponding relationships between sufficient reasons and consequences – a character which is part of the logical grammar of the words which we use to formulate causal and normative judgements. For causal explanations, for example, the following rule applies: if a because (causal) b, then – other things being equal – a must always follow b. The identification of a causal relationship means, by implication at least, the identification of a causal *regularity*. But this, I would argue, is the essence of what

has been called the 'induction principle'. Something analogous to what applies in the case of the causal 'because' also applies for the normative 'because': if someone ought to (must, may) do a because conditions b obtain, then – other things being equal – everyone ought to (should, might) do a when conditions b obtain. The conjunction 'because', whether used in its causal or its normative sense, carries an index of generality; it possesses the generality of those linguistic expressions which it conjoins – always with the proviso, of course, of other things being equal. 'Moral judgements are thus analogous to causal judgements and "because" statements generally in possessing this character of implicit generality.'⁵ We might speak of an 'equality principle' instead of a generalization principle, since the generalization principle demands the equal treatment of equal cases. In the case of both causal and normative 'because', an unequal treatment of *prima facie* equal cases requires an explanation (or justification) showing that the cases that have thus been treated differently are in some relevant respect, whether causal or normative, not in fact equal. Presumably the generalization or equality principle has a more general significance comparable in some ways to the logical principle of contradiction, but what interests us here is only the significance it acquires in the context of the logical grammar of causal and normative 'because' statements.

The normative equality principle already betokens an elementary concept of 'justice', which means nothing other than the equal treatment of equal cases, and which incorporates a notion of impartiality implying above all the impartial application of established norms. This is the sense in which we call an umpire 'impartial'. In an analogous sense we call a teacher fair when he has no 'favourites', or a judge just when the judgements he pronounces are not 'arbitrary'. The equality principle is admittedly not limited to the application of established norms, but also implies a sense for the precedents which may be created by individual actions and judgements. Just as the causal interpretation of singular events contains an indeterminate indication of a causal regularity, so too does the normatively conceived precedent contain an implicit norm. Both causal and normative precedents contain an implicit rule concerning an equal treatment of equal cases; they limit the freedom for the causal or normative interpretation of *future* cases.

In its normative sense, the generalization principle is really

expressing nothing other than the connection between expressions like 'should', 'must' or 'may' and the concept of a norm. For this reason, of course, the elementary concept of justice which we have just been considering becomes largely inapplicable as soon as we raise the question of how the norms are *grounded* by which standards for the equal treatment of equal cases are defined in the first place, which is to say as soon as the 'justice' of these norms themselves is called into question. After all, the equality principle only concerns the general character of relationships between grounds and consequences; it therefore provides a criterion for the *appropriateness* of causal explanations or normative justifications only in the sense of a principle of *consistency*. When we seek a justification for norms, on the other hand, we are concerned amongst other things with the question of *which* standards for the equal treatment of equal cases are the *right* ones. This is the question which Aristotle discusses in his *Politics*, posing it in the form of asking whether property-owners, freemen or the virtuous should respectively be treated as 'equals' when it comes to the apportionment of the rights and duties of a citizen. The idea that human beings *qua* human beings should be regarded as equals in the matter of fundamental rights is a universalistic principle which only surfaces in *modern* conceptions of morality and justice. We might concede that it is the only principle which could reasonably win the approval of all human beings (without whose acceptance normative principles cannot 'live'), now that traditional justifications of the inequality of mankind have lost their credibility and authority. Thus, as soon as we consider whether it is even possible in principle to justify norms if they can no longer be attributed to a transcendental authority – to justify them, moreover, in the eyes of those who are expected to acknowledge their validity – the logical grammar of basic normative terminology almost inevitably assumes a universalistic significance: we cannot rationally apply these terms in anything other than a universalistic sense. This is how the impression arises that the normative generalization principle and the universalization principle are *one and the same thing*. In my view, however, we should first distinguish between the two layers of meaning, one of which concerns the general character of normative judgements, while the other relates to the universalistic conditions for the possible intersubjective *validity* of norms. The fact that in Habermas's interpretation of the uni-

versalization principle the two layers of meaning coincide is already connected with his consensus notion of practical truth. In other words, it is connected with the fact that for him the *meaning* of normative validity claims coincides with the universalistic *conditions* of their possible intersubjective recognition. In what follows I shall first discuss the basic meaning of the generalization principle, and then proceed to an interpretation of Kant's moral principle as a generalization principle of the second order.

II

Like Singer and Hare I proceed on the understanding that the generalization principle we have been considering is in itself an essential aspect of what Kant formulates as the categorical imperative. But I should immediately add that, for one thing, the 'fact of reason' as Kant understands it is not reducible to this generalization principle, and that secondly it is not possible to derive a universalistic moral principle from this generalization principle even with the help of a supplementary premise (such as Singer's 'principle of consequences').⁶ We might clarify the sense in which the generalization principle is an essential aspect of Kant's moral principle in the following way: the categorical imperative requires that I should only act according to maxims which I can will to have the validity of a universal law. But *as a rule*, what I can will as a universal law will in fact be determined by my own, pre-existing normative convictions, especially by the socially determined normative expectations I have of *others*. In so far as this is the case, what the categorical imperative is ultimately saying is this: 'Do what you think *one* must do,' or even 'Don't do what you think *one* must not do.' In other words, 'Make no exceptions for yourself where normative matters are concerned,' or simply 'Do what you ought to do.' It will not be superfluous here, I think, to point out that the categorical imperative, even in this more or less basic interpretation of it, already presents a demand which is *by no means trivial*, namely that I should act in accordance with previously acknowledged normative obligations, and that I should do so *here and now and without self-deception*. Kant is entirely correct to describe *this* postulate as something simple and readily apparent to all, but

which is nevertheless difficult to fulfil. The requirement to act in accordance with my own normative convictions does not mean, after all, that I stop and work out a suitable justification for each of my actions in turn; nor does it mean that I should act in accordance with what I can *present* as my normative conviction *in each instance*. What it requires incorporates a demand which really is difficult to fulfil, namely that I should not deceive myself as to what I would really expect of *others* in the event that they should find themselves in my situation.

To be sure, the categorical imperative, as has already been emphasized, is not reducible to this basic meaning. The categorical imperative is intended to explain the very possibility of that categorical 'ought' or 'must' which is always implied within any 'normative conviction', and to explain it moreover as the possibility of an 'ought' or 'must' which can be *rationally understood*. It is in this way alone that the categorical imperative can become a universalistic moral principle; the generalization principle as such may be a valid principle for all 'rational beings', but it is not a principle which necessarily distinguishes universalistic norms from others.

We may call the categorical imperative a generalization principle of the second order; only at this point does the notion of a universalization principle become relevant. This universalization principle is no longer something which can be understood as a simple analogue of the induction principle. For here we are no longer concerned only with the general character which belongs to the logical grammar of 'ought' or 'must' statements, but also with the common will of rational beings (and therefore also with the intersubjective *validity* of moral judgements). The categorical imperative is a principle which not only has validity *for* all rational beings, but also refers *to* all rational beings (as is most clearly shown by the so-called 'ends formula').

In what follows I should like to reconstruct the meaning of the categorical imperative to the extent that is necessary in order to show clearly both the motives for moving to a dialogic ethics and also the burden of proof which arises for such an ethics. In my reconstruction I shall proceed selectively; of the various possible interpretations of the categorical imperative I shall adopt the one which appears to me to carry the most substance. The most appropriate of Kant's texts on which to base this interpretation is the *Groundwork of the Metaphysic of Morals*.

My starting-point is the following formulation of Kant's: 'We must be able to *will* that a maxim of our action should become a universal law – this is the general canon for all moral judgement of action.'⁷ The requirement that I should only act in accordance with maxims which I can will to be valid as universal laws is identical in meaning with the requirement that I should only act in accordance with maxims of which I can will that all others should act (e.g. towards me) in accordance with them. Now for Kant it goes without saying that if *I* am unable to will that a maxim should be valid as a universal law, then any rational being will be similarly unable to do so: the test of whether a maxim can be generalized is also a test of whether it can command general approval. Maxims which cannot be generalized are therefore those which I would not, in Gert's phrase, be able to 'publicly advocate'.⁸ This holds in a double sense, moreover: for one thing I should not be able to will that others should adopt such maxims, and secondly I should not be able to expect that the others could approve of such maxims as a universal rule (and of my following such maxims in particular). The maxims which cannot be generalized are therefore the ones upon which rational beings would not be able to agree as rules for common practice.⁹ There is, of course, a problem involved in Kant's assumption that when it comes to the generalization of maxims *my* 'being able to will' or 'not being able to will' must necessarily coincide with that of all other rational beings. And it is precisely this problem which has given rise to the attempt to expand Kant's ethics into a dialogic ethics. But I should like to pass this problem over for the moment and say something more about the connection between the categorical imperative, moral norms, and moral judgements.

We might say that the real point of the categorical imperative is that it enables the categorical 'ought' or 'must' or 'may' of moral norms (i.e. of 'categorical imperatives' in the plural) and of moral judgements to be seen as rationally understandable derivatives of the categorical 'ought' of a single meta-principle. The only thing which then remains to be explained is this categorical 'ought' itself. As long as we assume that Kant succeeded in elucidating this fundamental categorical 'ought', then it is also possible, with recourse to his arguments, to elucidate the 'ought' or 'must' or 'may' of our ordinary moral judgements and norms. In contrast to some of Kant's assertions, but in agreement with Ebbinghaus¹⁰ and Singer,¹¹ I argue from the premise that the

obligation implied by the categorical imperative is 'transferred' to substantive moral norms and judgements primarily by way of a *prohibition* of actions (or maxims) which are not capable of being generalized.¹² Let us take as an example the maxim that if need be I shall get myself out of difficulties by making an insincere promise.¹³ Like Kant, I assume that *we* – as rational beings – cannot will that such a practice should become universal. Now the categorical imperative evidently says that under these circumstances I (like X or Z) *may not* act according to the maxim 'if need be make an insincere promise'. In the concrete situation we are imagining this means that if *p* signifies the act of making an insincere promise and not-*p* signifies not making such a promise, then I *may not* do *p*, or I *ought to* (should) do not-*p*. The 'ought' in 'I (or one) ought to do *p*' would thus result from the fact that I *cannot* will of a particular maxim that it should have the validity of a universal law. The 'must' or 'ought' of our ordinary moral convictions could be 'derived' from the categorical imperative only *via negationis*, as it were.

According to the interpretation I am offering here, on the other hand, the generalizability of maxims would merely mean that it is morally *permitted* to act in a corresponding way. Now it is impossible to overlook some formulations of Kant's in which he says that maxims which can be generalized *are eo ipso* practical laws.¹⁴ In order to resolve this contradiction we must clarify the concept of 'generalizability', which is to say the meaning of the expression 'being able to will that a maxim should be valid as a universal law'. Now it is no accident, as we shall shortly see, that Kant himself uses examples of a *negative* nature. He demonstrates, for example, that I *cannot* (reasonably) will of certain maxims that they should be valid as universal laws, either because I cannot even conceive of them as universal laws, or else because my will would then be 'in conflict with itself'¹⁵ (for instance if I simultaneously *wanted* other people to help me and *did not want* them to help me). At this point it is important to remember that the categorical imperative, in so far as it contains a 'procedure of examination', does not relate to random propositions, but to maxims which someone – the performer of an act – 'has'. This is what gives rise to a characteristic asymmetry: someone who might, for sake of argument, have the maxim not to tell the truth if to do so would result in his own disadvantage can readily comprehend that he cannot will that this maxim should be valid

as a universal law. But someone who has the maxim to tell the truth on all occasions even if this results in his own disadvantage *does eo ipso* will that this maxim should be valid as a universal law (and by the same token he also *can* so will it). But this is something which we cannot straightforwardly assert about the person with the bad maxim. (In what sense could we possibly say of him that he 'can will' the maxim of truthfulness to be a universal law? For his own part, he wants to follow a different rule, and where others are concerned, perhaps it is good enough for him if they are just truthful towards *him*.) Let us take as another example someone whose maxim is never to show weakness however difficult that might be. *If* this is his maxim, then he will also be able (and perhaps even want) to will it as a universal law. But the same is true of a person whose maxim is to show weakness in preference to always playing the strong man. From these few examples we can already see that the answer to the question whether someone is able to will a particular maxim to be a universal law depends on what maxims he does in fact have. Thus the question of whether someone is able to will a maxim to be a universal law cannot decide the issue of whether this maxim is a 'practical law'. But if I assert that I *cannot* will my maxim to be a universal law, then it necessarily follows from this that to act in accordance with this maxim would be morally reprehensible (because I am making an 'exception' for myself): I 'must' not act in accordance with this maxim. Now if I were to adopt the maxim *not* to do what is prohibited by a maxim of the 'non-generalizable' type, then we could call this new maxim the 'negation' of the original one. (In this sense, the maxim 'I shall always tell the truth even if it is to my disadvantage to do so' is the negation of 'I shall not tell the truth if it is to my disadvantage to do so'.) If this new maxim really is the maxim by which I act, then I am also willing it to have the validity of a universal law. But in this instance the fact that I can will my maxim to be a universal law has a special significance by virtue of the logical process out of which my 'being able to will' has arisen, namely the fact that my maxim is the negation of a non-generalizable maxim. A maxim which can be generalized in *this* sense expresses a moral obligation, at least for *myself*.

What I would propose, therefore, is that we distinguish between a 'weak' and a 'strong' notion of the generalizability of maxims. The *weak* notion of generalizability is adequate for the purposes

of eliminating maxims of the non-generalizable type, but it is not adequate to support the thesis that generalizable maxims are practical laws (i.e. moral norms). For this purpose, a *strong* notion of generalizability is necessary, bearing in mind that the generalizability of maxims in this strong sense has to be understood in terms of a negatory relationship between these maxims and the non-generalizability of the maxims which they negate. Note, however, that *this* negatory relationship with non-generalizable maxims must be distinguished from the negatory relationship which *every* generalizable maxim (in the weak sense) has with its 'negation'. If my maxim is never to show weakness even in the utmost extremity, and if I will this maxim to be a universal law, then this means, of course, that I *cannot* will the maxim to show weakness rather than always playing the strong man to be a universal law. But in *this* case, my 'not being able to will' is dependent upon the fact that an opposing maxim is already *my* maxim (so that the 'not being able to will' is here secondary to the 'being able to will'). The 'not being able to will' in the case of a maxim which is *mine* is quite different, for the non-generalizability of such a maxim is independent of other maxims which I might have in addition.

Thus it is only possible to assert that generalizable maxims are *eo ipso* practical laws if we limit our understanding of 'generalizable' to those maxims whose negation is non-generalizable, even assuming that such maxims were *mine*. This is what provides the philosophical justification for the emphasis on negation when putting the question about the generalizability of maxims. The problem of the intersubjective *validity* of moral norms, however, cannot be definitively solved even in this way. It is far from clear (contrary to what Kant believed) that the moral obligations which I acknowledge must necessarily be acknowledged by every other rational being (and vice versa). This is a problem to which I shall return. For the purposes of the deliberations which follow, meanwhile, I shall assume for simplicity's sake that the moral norms distinguished by the categorical imperative do have intersubjective validity.

III

Next I should like to discuss Kant's thesis that the norms distinguished by the categorical imperative are *universally* valid, that is, they are binding 'practical laws' which admit of no exceptions. It is possible – using Kant's own arguments against him – to justify this thesis if it is formulated with sufficient caution. Let us take for example the maxim, 'If it *suits* my purposes I shall tell an untruth,' which we may safely assume to be non-generalizable. When we say that the categorical imperative forbids us *once and for all* to act in accordance with this maxim, we mean that a categorical prohibition is placed on the telling of any lie that might be justified (or motivated) by this maxim; and this prohibition applies in the sense of strict universality (in the sense of Kant's '*universalitas*').¹⁶ As a moral norm this might be formulated as follows; 'One may not lie' or 'Thou shalt not lie'. But we must not overlook the fact (even if Kant himself overlooked it) that the universal validity (*universalitas*) of this norm results from the non-generalizability of a *particular type* of maxim (or from the inadmissibility of a *particular type* of reason for acting). The strict prohibition applies to a class of *reasons* for acting; it *cannot* apply as a strict prohibition to the corresponding *actions* (in this case, lying). Kant's polemic against the possibility of exceptions is quite correct if it is applied to the right sort of exceptions. The norm 'Thou shalt not lie', if understood in the sense we have outlined above, admits of no exception whatsoever. But this is not to deny the possible existence of reasons for acting which do constitute *generalizable* exceptions by virtue of the fact that they can be 'publicly advocated' or, to keep closer to the terms of our own argument, that the maxims to which they correspond can be generalized. It appears that Kant has confused two distinct categories of 'exception'; provided that we ourselves avoid this confusion, we shall have no difficulty in viewing derived moral norms as universally valid in the strict sense *while also* admitting the possibility of exceptions.

The distinction we are dealing with here is, incidentally, not identical with the distinction between 'egoistic' and 'altruistic' motives. I may not lie, *regardless of whether* the advantage I hope to attain thereby is for myself or my child or my friend. What is prohibited by the norm we are discussing is the telling of a lie

for *private* ends (whether these are egoistic or altruistic in nature). Saving an innocent person from the Gestapo, for example, would not be a private end in this sense, but rather one which could be justified by means of a *different* moral norm, namely that we may not deny help to innocent victims of persecution. The transgression against a moral norm ('Thou shalt not lie') could be defended in this case with reasons which can be 'publicly advocated' (even if under the implied circumstances I should not be able, in a phrase of Kant's, to 'openly express' them). In other words we might say that the corresponding maxim, 'I shall try to save innocent victims of persecution, and if need be I shall tell a lie in order to do so,' is generalizable. A difficulty arises at this point, however, which Kant was excused by his rigoristic interpretation of moral norms from having to answer. It is this: whereas it is self-evident that the maxim we considered earlier, 'If need be (i.e. if it suits my purposes) I shall tell an untruth,' *cannot* be generalized, the generalizability of the maxim we have now arrived at is, on closer inspection, by no means self-evident. This maxim is too *vague* to permit a clear decision. The difficulty we have might be summed up as follows: I am only able to will *this* maxim to be a universal law if I can be certain that all mankind possesses sufficient judgement and good will to decide correctly when the 'needs' of the situation really justify a lie. If I could be certain of *this*, however, there would no longer be any need for our maxim, because then there would be no persecution of innocents. We must therefore acknowledge that Kant is being consistent when he rejects such maxims as an unsuitable basis for law-making in any conceivable kingdom of ends. In order to justify any exception to the rule 'Thou shalt not lie', we should clearly have to discuss the particular circumstances of a concrete situation. The reasons which I could 'publicly advocate' in justification of this exception might in principle be capable of expression in the form of a generalizable maxim, but in the process we are confronted with the dilemma that the more *precisely* I describe the type of situation involved, the *smaller* the *field of application* for such a maxim becomes, and that the more *general* I make my description, the more *indeterminate* it becomes. What this means is that the justifiable exceptions to moral norms are not subject to rules in the same sense as the actions which those norms prohibit (or command), and that is why the *faculty of judgement* plays a much more fundamental part in the appli-

cation of moral norms than Kant wished to admit. This is also why, in matters of moral controversy, it is not as a rule the basic moral norms that are at issue, but the descriptions of situations or of types of situation. As soon as we have agreed upon these descriptions (and thus on the 'facts' in the broadest sense of the word), the moral controversies as a rule resolve themselves; in this sense we might say that moral demands reside in the concrete circumstances of action.

I should like to conclude my analysis of the example we have been looking at by modifying my assumptions about the moral problems it presents us with in one further important respect. This modification amounts to 'dissecting' the problem into two component parts. If we look at the question in terms of moral *norms* rather than maxims for action, then the problem consists precisely in a conflict between two norms, the first of which requires that I assist persecuted innocents, while the second requires that I should not lie. Now if I consider the way in which these two norms are related in a negatory fashion to non-generalizable maxims for action, then it immediately becomes clear that our hypothetical situation represents a case of the *direct* application of the *first* norm, but only the *indirect* application of the *second* norm. In other words, the command to lend assistance results from the non-generalizability of the maxim 'I shall only assist persecuted innocents if it is not to my disadvantage to do so'; in this instance the *aim* of my action (to assist persecuted innocents) is positively *commanded*. The prohibition against lying, on the other hand, results from the non-generalizability of a maxim which – under the conditions we are assuming – is not even at issue in the situation we have described, namely the maxim 'I shall tell falsehoods if it appears advantageous to me to do so'. This is just another way of saying that what is at issue here is not lying as a means of achieving a 'private end', but as a means of realizing an end which is morally *commanded*. Thus if we take account of the inherent relationship of each of these norms to non-generalizable maxims, and if we also take account of the specific nature of the situation in which the action takes place, our conflicting norms are revealed to be different in kind. For cases of this sort we should therefore be able to agree with Kant's thesis that there is in reality no conflict between different kinds of moral duty, even if this thesis now acquires a somewhat un-Kantian twist.

My second description of our exemplary situation shows up an aspect of that situation which in my first characterization – with the help of a quasi-generalizable maxim for action – remained obscured. Conversely it is also true to say that the second description obscures an aspect which the first had highlighted, namely the problems concerning the application or ‘concretization’ of moral norms. We only have to modify the extreme case suggested by our example very slightly in order to see that a lie can by no means always be a legitimate means of assisting persecuted innocents. This means, however, that in spite of the general nature of the argument we have just developed in order to resolve the apparent conflict between norms, the solution we have proposed can only be valid in extreme cases. The full implications of this aspect of the problem become clear if we try to subsume the exception to the command to be truthful under a generalizable maxim, i.e. if we try to formulate a kind of law of permission. As we have seen above, it is not possible, strictly speaking, to formulate generalizable maxims of this kind because they would have to contain some indexical element. A ‘law of permission’ could therefore only take the form ‘In situations such as *this* it is permitted to lie’. The general validity of exceptions, unlike the general validity of moral commands themselves, can ultimately only be demonstrated in the sense of a justification of particular actions in concrete situations. This is where ‘situationist’ or ‘existentialist’ ethics contain an element of truth. We cannot but marvel at the immense astuteness of Kant who, because he had made no provision for such a ‘situationist’ element (implying the exercise of judgement) in his ethics, adhered with the utmost consistency to the only possible alternative, namely a rigoristic ethic of duty.

Let me now return to the thesis that ‘derived’ moral norms – as strictly general norms which nevertheless admit of possible exceptions – arise by a process of negation from non-generalizable maxims. This thesis does not imply that, when considered in the light of the categorical imperative, all basic moral norms take the form of prohibitions – along the lines of ‘Thou shalt not lie’, or ‘Thou shalt not kill’, or even ‘*Neminem laede*’.¹⁷ It is much rather the case that even norms like ‘Help the needy (as far as you are able)’ – which, according to Kant, correspond to ethical duties that are ‘of broad obligation’¹⁸ – can, in the same way as moral norms of prohibition, stand in a negatory relationship to

non-generalizable maxims such as, for instance, ‘I shall help nobody if it is of no advantage to me to do so’. The difference between these ‘positive’ moral norms, which come close to what Gert calls ‘moral ideals’,¹⁹ and moral norms of prohibition is that the latter prohibit actions, while the former prohibit the *omission* of actions (or *attempts* to act). But whereas the prohibition of an action means the same as a command *not* to perform *this* action, the prohibition of ‘remaining inactive’ (in particular situations) does not as a rule mean the same as a command to perform *a particular* action. Broadly binding ethical duties leave, as Kant remarks, ‘a latitude within which we may do more or less without being able to assign definite limits to it’.²⁰ We might say that positive norms command us to act in a particular way (making the happiness of others my aim, to put it in Kantian terms), but they do not command us to perform a particular act.

Excursus R. M. Hare has tried to solve the problem of exceptions to moral ‘prima facie principles’²¹ in a somewhat different way from the one I have proposed here. Hare distinguishes between two levels of moral thinking, which he calls the ‘intuitive’ and the ‘critical’.²² On the intuitive level of moral thought we are concerned with prima facie principles which are both *general* and more or less *unspecific*, i.e. they admit of exceptions. It is only in situations of moral conflict that we are forced to switch to the *critical* level of moral thought, which implies the formulation of critical moral principles which can be of ‘unlimited specificity’.²³ For Hare, prima facie principles are merely a means of disencumbering ourselves of complexities, so to speak (or of developing moral ‘habits’), in the moral practice of everyday life. If we had the intellectual capacity of an archangel we would be able to exercise critical moral judgement on all occasions and thus allow our actions to be determined by moral principles which consistently did justice to the particularity of the situations we found ourselves in.²⁴

Hare’s ‘critical moral principles’ are prima facie principles which have been modified by exception clauses in some such manner as this: ‘One ought never to do an act which is G, except that one may when it is necessary in order to avoid an act which is F, and the act is also H; but if the act is not H, one may not.’²⁵ (This example of course provides only the starting-point for the critical specification of a prima facie principle.) The reason why

this method of solving the problem of 'exceptions' seems to me cock-eyed is that we do indeed require the notion of an archangel (or a god) in order to combine the universality of principles with the particularity of situations in the manner which Hare is describing; which is to say that we need the image of an infinite intelligence which would be capable of 'sublating' the particular completely within the universal. Only if we presuppose such an infinite intelligence as the ultimate goal of our finite mental efforts can we attempt to solve the problem of moral exceptions or conflicts (i.e. moral *problem* situations) with reference to the unlimited specifiability of moral norms. In using this image, Hare is importing a figure of thought which is in a certain sense legitimate in the field of the natural sciences (i.e. the idea that there are no limits to the possibility of specifying causal laws) into the realm of practical, historical phenomena. If for no other reason, this importation should be disqualified on the grounds that in the field of history (to which all action belongs) the notion of an 'ultimate' language, of an 'exhaustive' description, does not even make sense as a regulative idea.

There is a way of formulating the objection to Hare's notion of the unlimited specifiability of moral principles which is both more specific and more precise. We have seen that, proceeding on Kantian premises, it is possible to distinguish clearly between maxims which are (in the strong sense) generalizable and from which it is possible to derive universally valid moral norms, and 'quasi-generalizable' maxims for which the equivalent moral norms would have to take the form of 'laws of permission'. We have also seen that, because of their imprecise nature, it is not in fact possible to formulate such laws of permission as laws in the Kantian sense. This is because they cover an indeterminate number of cases in which what they seem to permit would be morally prohibited. In Hare's example of a norm that is specifiable (within limits), quoted above, the first 'except' clause represents a law of permission of the kind we mean, whereas the second 'except' clause signifies a limitation of the permission. Now Hare, too, argues from the premise that as creatures of finite intelligence we must get by with *finite* specifications of moral principles. But if the critical moral principles upon which we base our judgements include among their components 'laws of permission' which, because of the limited specification of our principles, can, as it were, only be furnished with a basic minimum of necessary

qualifications, then these moral principles are themselves, almost of necessity and certainly predictably, false.

Let us consider for example the following principle: 'It is forbidden to kill people except when it is necessary in order to perform an act of mercy towards a terminally ill patient who asks you to release them from their sufferings.' In advocating such a principle we have particular situations in mind in which it may be defensible, and perhaps even commendable to act in such a way. But it is obvious that not only could such a principle, if it were to become enshrined in law, be subject to infinite abuse (this is the *exoteric* side of the problem), but also that in this general form it is probably morally wrong, even presupposing that it were applied in a spirit of *good will*. In the vagueness of the formulation (*when* is an action necessary as an act of mercy?) there also lurk a thousand counter-arguments which might occur to us in concrete situations (and even at our desks), and I mean counter-arguments against the principle itself in its general formulation. In concrete situations, however, we have to assume responsibility and act as best we can. Now if we were to reconstruct the logic of moral judgements or justifications as Hare understands it, our moral justifications in problem situations would be necessarily false because they were based on false principles. But in reality, the fact that in concrete situations we have principles (albeit not yet sufficiently specified ones) by no means prevents us from having reasons for doing the right thing in these situations. Whether the reasons we have are watertight depends, it seems, more on our apprehension of (these) situations than on the availability of generally valid principles. Or to put it another way, even if reasons and descriptions of situations always have something general about them, in the cases which we are considering here the justifications which we are able to formulate invariably include an indexical element, and that (if anything) is what decides whether they are watertight. Our apprehension of the situation always exceeds what is made explicit in our descriptions and justifications; that is why even our (critical) principles, if we formulate them, contain an implicit reference to paradigmatic situations with which we have to be familiar in order to apply these principles. The principles themselves contain an indexical element; otherwise they would not be able to play any part in moral thinking without having been adequately 'specified' in advance. The same is not true, however,

of prima facie principles if we understand them as moral norms whose universal validity resides in the negatory relationship in which they stand to non-generalizable maxims.

It is my view, then, that whatever else we might think of Kant's ethics, they do allow us to formulate a categorical distinction between moral norms and 'laws of permission' which shows up something of the 'fine structure' of moral judgements and justifications, a fine structure which is, moreover, lost to view if in the formulation of 'critical' moral principles we interpolate moral commands or prohibitions on the one hand and 'laws of permission' on the other, in the way that Hare does. It was for this reason, and because of the rationalistic implications of his approach (as described above), that I called Hare's solution to the problem of moral exceptions 'cock-eyed'.

Now although these ideas of Hare's which we have criticized belong to the same rationalist tradition as Kant's philosophy, it appears that Kant, in contrast to Hare, has retained an element of Aristotelian caution when transferring the concept of law into the realm of moral philosophy. It is precisely for this reason that he would have been bound to find it impossible to mediate between general principle and particular situation in the manner in which Hare attempts it. Kant's philosophical rigorism in moral questions is the rationalistic conclusion he had drawn from this very difficulty; this conclusion was the price he paid in order to be able to define moral action generally in terms of moral laws. On the other hand we saw that it is perfectly possible to preserve the '*universalitas*' of fundamental moral norms (i.e. the 'prima facie principles' in Hare's sense) if we relate them in a negatory fashion to non-generalizable *maxims* for action. The problem of 'exceptions' then appears in a different light, namely as one that can ultimately only be solved in concrete situations, and then by means of *reasons*, not by means of an unlimited specification of *principles*.

I do not want to exaggerate the differences I have with Hare. In a certain sense it is possible to see Hare's distinction between 'intuitive' and 'critical' moral principles as a translation of my own proposed solution to the problem of 'exceptions' (working with a Kantian perspective) into a different philosophical language; the structural homologies are plain to see. Looked at like this, Hare is closer to an Aristotelian tradition than Kant. But when Hare talks of 'principles' where strictly speaking it is no

longer possible to formulate them, I consider this to be, to say the least, misleading. Even if the reasons which we might adduce for moral judgements in concrete situations always carry an index of generality, in cases of moral conflict they nevertheless remain 'linked' to the situations in question in such a way that they can at best be converted into ad hoc principles, i.e. into rules, the correct application of which is tied to a capacity for judgement which can in turn only be formed through a process of (moral) familiarization with corresponding exemplary situations. This is not true of those moral principles which 'correspond' to non-generalizable maxims; here we really are dealing – in the sense which I have explained above – with *universal* principles. I therefore believe that my own proposed solution to the problem of moral exceptions – at least within a Kantian perspective which, in a broader sense, Hare also shares – is more convincing than Hare's solution.

IV

Kant evidently never conceived the individual will that is capable of willing a maxim to be a universal law as anything other than the expression of a will that is *common to all mankind* as rational beings. The 'cognitivism' of Kant's ethics, i.e. the claim of moral judgements to universal (in the sense of *intersubjective*) validity, stands and falls with this presupposition (leaving aside for the present the problem of 'fundamental' grounding). The problematic nature of this presupposition is plain to see: the expression 'capable of willing' contains an irreducibly 'empirical' element, and we must therefore reckon with the possibility that different people are capable of willing *different* ways of acting as universal ones. I have shown above that this problem can be alleviated *to a certain degree* if we keep the concept of non-generalizability to the forefront of our minds whenever we consider the concept of a generalizable maxim (in the strong sense). For when the matter is defined in terms of 'not being able to will', then the individual exercising moral judgement is placed in a privileged position: if *I cannot* will something to be the case, then *we cannot* will it.

This is not to say, however, that the intersubjective validity of moral judgements is thus secured, for what we can or cannot will as general ways of acting is indubitably decided from case to

case by the matrix of concepts through which we interpret social reality and our own needs. Let us consider, for instance, the authoritarian maxim of not beating about the bush but immediately taking stern measures when confronted with insubordination (of pupils, subjects, persons under my command). Whether or not I judge this maxim to be generalizable depends on whether I, as an authoritarian teacher or officer, interpret social reality with the help of a matrix of concepts in which obedience and insubordination represent, as it were, the positive and negative poles of an ordered existence, or whether, as a democrat, I interpret them with the help of concepts related to a normative scale that runs from self-determination at the one extreme to dependency at the other. A moral principle like the categorical imperative can never operate in a 'vacuum' where norms are concerned. But if this is so, then following the categorical imperative in concrete situations does not *by itself* guarantee the intersubjective validity of corresponding moral judgements. And it is initially *quite impossible* to see how the categorical imperative could help us to secure a moral consensus. And if we also bear in mind the problems surrounding the question of 'exceptions', or rather of the situational aspect of moral judgements, which we saw earlier, then we are confronted with a tangle of difficulties which Kant was only able to hide beneath his 'formalistic' interpretation of the categorical imperative. This formalistic interpretation comes to the fore in the *Critique of Practical Reason* where, at decisive points of the argument, what can be 'willed' is revealingly replaced with what can 'hold good', the criterion for which is what can be 'conceived'.²⁶ The following passage is typical:

Therefore, either a rational being cannot conceive his subjective practical principles, that is, his maxims, as being at the same time universal laws, or he must suppose that their mere form, by which they are fitted for universal legislation, is alone what makes them practical laws.²⁷

By making the 'form' of maxims as 'laws' the criterion for morally correct action, Kant – seemingly – preserves the objectivity of morals. But if we follow this thought to its logical conclusion we find that the fruitful ambiguity of Kant's moral philosophy has been abandoned in favour of a formalistic ethic of duty which

can scarcely be of serious interest to us today.

My own view – which is surely compatible with nearly all contemporary philosophies of morals which take Kant as their starting point – is rather that the productive element in Kant's formulation of the moral principle in the *Groundwork* resides precisely in the reference it makes to the empirical will of the person performing the act, even if it conceives the object of that will to be general ways of acting rather than individual goals. We might say that *if* there is anything correct in Kant's reconstruction of our moral intuitions, then it must be the way that the rationality of moral judgements is anchored in a particular relationship between what is (empirically) 'willed' and what 'ought' to be the case. A large part of contemporary moral philosophy can be understood as the attempt to emphasize this basic Kantian intuition over against the decline of Kant's ethics into formalism; indeed, this is the common ground between rule-utilitarianism and communicative ethics. And if we take seriously the problem that lies behind Kant's construction of a connection between rationality and the will of beings who have the capacity to act, then crudely speaking there appear to be three possible alternative solutions for an ethics based on Kant.

The *first* of these alternatives consists in conceding that different rational beings might be able to will quite different ways of acting as universal ones. In this case we are denying the necessary condition that the rational will of all beings capable of action should coincide; the moral universe disintegrates – potentially – into a plurality of moral worlds, as is the case at least in the earlier work of Hare.²⁸ Hare incidentally eliminates the 'problem of fundamental grounding' by deriving the universalism of ethics directly from the logical grammar of basic moral terminology ('ought', 'must' etc.). He anchors ethics, so to speak, in a fact of *our* (modern) reason.

The *second* alternative consists in the attempt to establish a minimal ethics, the contents of which more or less correspond to moral norms which we also encounter in Kant if we consider examples of non-generalizable maxims. The negatory relationship between moral norms and non-generalizable maxims reappears in this second variant on 'Kantian' ethics, but in an altered form: moral norms are in the first instance prohibitions on ways of acting on which rational beings would not be able to agree as admissible ones. The word 'rational' should here be understood

in a *weak* sense: the 'rational' will is a self-interested one which takes account of the consequences of alternative possibilities for regulating action – just as in the corresponding points of Kant's argument. I believe that the theory of B. Gert,²⁹ some thoughts of G. H. von Wright,³⁰ and to a certain extent also the theories of Singer and Rawls³¹ belong to this second alternative. Since in this second alternative the moral 'ought' is reconstructed from its basic contents it is true in a sense that the unity of the moral universe remains intact, but here it is the concept of moral obligation that presents a problem instead. Once the rational sense of the moral 'ought' has been called into question, it is no longer possible to resort to a Kantian coup in order to justify the demand that – here and now – I *ought* to act according to generalizable maxims or, in Gert's terms, according to reasons which can be 'publicly advocated' – in other words, that I should act *morally*.³² The reconstruction of intersubjectively valid moral contents in Kant's sense has the (paradoxical) consequence that the immediate identity of the *rational* will with the *moral* will breaks down.³³ It can only be restored – not by means of an ultimate grounding, but in weaker forms – if the categorical 'ought' is linked once more with an (empirical) 'willing'. The second alternative thus reveals the deficit in Kant's ethics regarding a *philosophical grounding*; the more convincing it becomes as a reconstruction of basic ideas in Kant's ethics, the more clearly it reveals that the unconditional moral 'ought' of the categorical imperative can scarcely be an 'ought' of pure practical reason.

The *third* and last alternative consists in the expansion of Kant's moral principle into an ethics of discourse as has been proposed not only by Apel and Habermas, but in a different form also by the 'constructivists' of Erlangen and Constance.³⁴ Much as in the case of the alternative we have just been looking at, valid moral norms are here equated with those rules upon which we should be able to agree in a rational dialogue. The decisive point in which this alternative *differs* from the previous one is that it abandons the claim to a philosophical grounding of substantive moral norms and replaces it with a principle of dialogic agreement which *takes the place* of Kant's moral principle. By taking this approach, it also becomes possible once more to tackle the problem of fundamental grounding. At least Apel and Habermas have tried to show that a principle of uncoerced dialogue for settling normative validity claims is 'built into' the conditions

for communicative action as a constitutive principle, which must therefore always have been acknowledged, implicitly at least, by beings that have the capacity for speech and action. Of the three alternatives for developing an ethics based on Kant, or on 'Kantian' precepts, which we have outlined here, only the third represents a serious attempt to rehabilitate an emphatic concept of practical reason in Kant's sense of the term, that is to defend *both* the possibility of justifying moral norms *and also* the rational sense of an unconditional moral 'ought'. I shall examine the difficulties which arise from this third alternative at a later stage.

V

The selective reading of Kant's moral principle which I have been propounding so far corresponds structurally more or less to the *second* of the three alternatives outlined above. Accordingly I have so far omitted any discussion of whether it is possible to provide a philosophical grounding for the categorical imperative, which also means that I have not addressed the question of the rational sense of the moral 'ought' in the light of the categorical imperative. I find myself in agreement with other critics of Kant when I say that Kant himself gave no satisfactory answer to this question. But as we have seen, this is not the only weakness in Kant's moral philosophy. If I have so far emphasized it *strengths*, then I did so because on the one hand I wanted to show that Kant's reconstruction of moral judgement can be made entirely plausible for a *limited* but *fundamental* class of moral problems, and also because, by emphasizing the strengths, I wanted to throw the *weaknesses* of Kant's argument into sharper focus. In this way I hope both to have established a certain standard by which we can measure the achievement of an ethics that attempts to go beyond Kant, and also to have characterized the problem areas in which it is possible to discern the motives for the development of a dialogic ethics. For the problem areas of Kant's ethics, as characterized above, prompt the thought that we might situate the formal principle of ethics that Kant was looking for one step lower, as it were, and try to identify it in the connection between (intersubjective) *validity* and (rational) *grounding*. This is what Apel and Habermas try to do, replacing Kantian formalism with a 'procedural' formalism. The intention is that the

formulation of a moral principle in procedural terms should not only open the way for philosophical investigation of those problem areas which remain as blank spots on the map of Kant's ethics, but that it should at the same time also enable us to solve the problem of 'fundamental grounding' without relapsing into metaphysics. These two basic intentions of dialogic ethics are interconnected in a systematic way, as we shall see later. But in order to clarify these intentions a little further, and in the process to show exactly what claims are being made by discourse ethics and the points that remain for it to demonstrate, I should first like to pursue the question of whether it is not possible to find in Kant's own writings points from which to develop a dialogic expansion of ethics.

John R. Silber has attempted to show that the formalism of Kant's ethics must *in itself* be understood as a 'procedural' formalism.³⁵ What he means by this is not the 'procedure' of a real dialogue, however, but the procedure of forming a moral judgement. Silber tries to explain the 'procedural formalism' of Kant's ethics by interpreting the categorical imperative in the light of what Kant calls 'maxims of common human understanding'.³⁶ In our context it is the second of these maxims ('Think from the standpoint of everyone else') that is especially important. In the light of this maxim, says Silber, the moral examination of a maxim for action requires a hypothetical shift of perspective, since it is only by thinking from the standpoint of others, and particularly of those who would be affected by our actions, that we can arrive at a considered judgement as to whether – as rational beings – we can reasonably will a maxim to be a universal law. 'In order to respect the humanity of all rational beings the moral agent must put himself into the place and point of view of others. In this way he will understand the values and needs of other beings and by moving out beyond himself will limit his tendency to concentrate upon the fulfilment of his own needs to the neglect of the needs and legitimate desires of others.'³⁷

Silber's interpretation of the categorical imperative in the light of Kant's maxim from the *Critique of Judgement* might seem to suggest that Kant's ethics contains the inherent possibility of a link to a dialogic ethics. For if I am only able to give an adequate answer to the question whether I can (reasonably) will a maxim to be a universal law if I also let the others speak in my own mind, as it were, and give due weight to the perspective – the

values and needs – of the others (and this can only mean others *as they really exist*) in my own considerations, then it seems to me that two things follow from this:

- 1 that a hypothetical element is inherent in moral judgements (it is possible for me to be *mistaken* in my assessment of the standpoint of others), and
- 2 that it is in the nature of moral thinking that it points to real dialogue (since I should only be able to test my apprehension of the standpoint of others by means of real communication).

In other words, if 'thinking from the standpoint of everyone else' presupposes, in the case of moral thinking, an apprehension of the standpoint of others, then the idea of moral insight guided by the categorical imperative represents a *problem* which I can only ever solve in a hypothetical and provisional sense. The question of the generalizability of maxims then becomes a question of whether *we* are able to will a maxim as a universal law; and this is a question which it is *ultimately* only possible to answer by means of real communication among those concerned.

It is true that Silber himself has not completed this step that leads from a 'procedural' to a dialogic ethics. What Silber is trying to show with his repudiation of 'formalistic' interpretations of Kant's ethics is precisely that a 'monologic' application of the categorical imperative in the examination of maxims is perfectly adequate, as Kant believed, for producing agreement between individual and general will. Looked at in this way, Silber's reference to the maxims of common human understanding should be understood as an attempt to show that the categorical imperative represents the precise *specification* of those maxims by which reason constitutes itself *as practical*. It is in this precise sense that Silber says, 'The moral law is itself to be understood as a principle which specifies the procedure of judgement in the act of moral schematism.'³⁸ But what Silber leaves unclear is the question of how a 'monologic' application of the categorical imperative can be reconciled with the requirement that we must give the perspective of others – their needs and values – due weight in our moral thinking, for *this* requirement seems to point to the necessity for a transition from solitary reflection to a real dialogue.

Silber admittedly concedes the *fallibility* of moral judgements. In this connection he points to the demand, which Kant cites

with approval, that we must 'work out our own salvation with fear and trembling'.³⁹ But what is actually implied there is the infinite potential for self-delusion, in the sense that we can never be entirely certain that our *basic way of thinking* is morally good. On the other hand, the question of correctly apprehending the needs and values of others, in the way that Silber presents it, is not in the first instance at all concerned with the problem of moral self-delusion, but much rather with the problem of the appropriate apprehension of situations, including the precise manner in which those affected by a particular action are involved in the situation. As far as this problem is concerned, Silber's demand that in making a moral judgement we must put ourselves in everyone else's position appears less as a proposal for *solving* the problem than as a – somewhat misleading – reformulation of the problem itself. But then if Silber were correct with his thesis that the procedural formalism of Kant's moral principle at least *aims at* a solution of precisely this problem, we should also have to concede that it is in the nature of the categorical imperative to require a transition to *real* dialogue, for only in the medium of real communication and discourse would it be possible to clarify whether I have put myself in the position of others *in the correct manner*. Silber's thoughts would then have to be understood as pointing to an inherent 'dialogicity' in Kant's moral principle. The question is, can we concede that Kant's ethics – implicitly – contain such a dialogic principle without simultaneously calling into question Kant's own grounding of his ethics in a 'monologic' moral principle?

For the purposes of answering this question, I should like to distinguish between a 'dialogic ethics' and an 'ethics of dialogue'. By 'dialogic ethics' I mean an ethics in which the dialogue principle *is substituted for* the moral principle; by 'ethics of dialogue' I mean an ethics in which the dialogue principle occupies a central position among the *derived* moral principles. My thesis is that it is not possible to move to a dialogic ethics while remaining true to Kant's thinking, but that it may very well be possible to expand Kant's ethics into an ethics of dialogue. It is precisely such as expansion of Kant's ethics, which we might call 'dialogic' in a narrow sense, that is suggested by Silber's reflections. It would be an expansion of Kant's ethics which acknowledged the existence of a plurality of needs and values, as well as the necessity of mediating between these transsubjec-

tively, and took account of them as *problems* which Kant had neglected. The decisive thought is this one: in so far as it is *possible* for individuals to resolve differences in the way they understand themselves and the situations they find themselves in by means of dialogue, and in so far as it is *possible* for them to achieve a mutual understanding about their respective needs and values, then the Kantian ethic also *demand*s that this should be done. For a maxim of refusing to enter into dialogue in situations where incompatible claims, needs or interpretations collide with each other is not (in the Kantian sense) generalizable. A dialogue principle that is derived in this fashion will not, however, be primarily concerned with the question of the generalizability of maxims, but above all with the question of an adequate apprehension of situations and the way individuals see themselves. It will be particularly effective in cases where a correct understanding of the needs and values of others is required. What we are dealing with here is, as it were, the 'communicative substructure' of Kant's ethics, i.e. that dimension of practical reason that has to do with problems concerning the 'commonality' of our world and the adequacy of our understanding of situations and of ourselves. This dimension of the way we form moral judgements is largely excluded from Kant's argument. Silber's reflections at least point towards it, even though he fails to clarify how this dimension of moral judgement could be deployed in the context of a Kantian perspective. Silber fails to recognize that Kant himself systematically trivialized this problem.

The trivialization can be seen at work, for example, in the case of the non-generalizable maxims we referred to earlier. It would certainly be possible to say that the act of asserting that these maxims are non-generalizable also presupposes a kind of hypothetical shift of perspective, in the sense that it must *in principle* be possible for me to put myself in the place of a helpless person in order for me to arrive at the judgement that I cannot will the maxim of refusing help as a universal law. This is a matter of anthropological common factors of a fundamental kind, which were so self-evident to Kant that he would have envisaged the shift of perspective called for by Silber as having already been effected in the very act of somebody in a particular situation recognizing the predicament of somebody else *as* that of someone in need of help. Kant thus presupposes that the shift of perspec-

tive which is necessary in each instance has already been effected *before* the question of the generalizability of maxims for action presents itself. And in the context of the 'elementary instruction' in morals to which his ethics is ideally 'suited', this makes good sense. Things are very different in the non-elementary domain of moral reasoning, where we are concerned with an adequate apprehension of complex situations or the historically variable ways in which people see themselves and the world in which they live. At this non-elementary level of morality, not only the *knowledge*, but also the adequate *apprehension* of the needs and values of others becomes a *problem*, and with it my own understanding of myself and the world. Together, moreover, these things constitute a problem which has to be solved before there can be any possibility of forming a correct moral judgement.

Silber tries to interpret Kant's moral principle in such a way that it incorporates *this* dimension of the formation of moral judgements. In the nature of Kant's ethics, there is certainly some justification for 'opening the subject up' in this way. But Silber obscures the difficulties that lie in the way of such an 'opening up' of ethics from a Kantian point of view. That is why it also remains unclear at which point the *problem* of a hypothetical shift of perspective presents itself within the framework of a Kantian ethics.

Neither Silber's reflections nor the thoughts we have developed from them have so far enabled us to escape from the magic circle of a 'monologic' moral principle. But at least it has become clearer that the 'problem areas' of Kant's ethics, to which we referred earlier, also represent a dialogic dimension of morals which Silber's reflections have at least pointed us towards. For the moment, however, our attempt to develop Silber's interpretation of the categorical imperative and find a basis in Kant himself for a dialogic ethics has led us back towards the *second* of the three alternative ways of reconstructing Kant's universalism which we outlined earlier.

In the sections that follow I should like to discuss the third alternative – the one that makes the stronger claim – in the form in which it has been elaborated by Habermas and Apel. This is the alternative in which a dialogue principle is *substituted* for the moral principle.

II A Critique of Discourse Ethics

VI

In this part of my essay I shall concern myself primarily with that form of discourse ethics which has been developed by Habermas. Only when I come to discuss the way in which discourse ethics is based on premises drawn from consensus theory, and also when I discuss the problems of fundamental grounding, shall I deal explicitly with Apel's version of discourse ethics (and with W. Kuhlmann's more closely argued presentation of the question of fundamental grounding). This manner of proceeding obviously imposes its own limitations; I consider it to be justified nevertheless, because my purpose is the *exemplary* clarification of certain matters of principle with reference to a conveniently concise and particularly rich text. I believe that my objections to Habermas's formulations of discourse ethics apply equally to Apel's more recent reflections on the subject⁴⁰, but this is not the place for me to demonstrate the point.

Habermas has compared the historical (phylogenetic) transition to a universalistic moral consciousness with the (ontogenetic) emergence of a post-conventional moral consciousness in adolescents. In either case, the emergence of a post-conventional moral consciousness is a response to the fact that previously unquestioned norms come to be seen as questionable and *in need of justification*. Where this happens, argument itself – as the 'reflexive' form of communicative action⁴¹ – becomes the only possible authority for redeeming normative validity-claims. The transition to a post-conventional moral consciousness also means the transition to a new way of *understanding* normative validity-claims, in which the intersubjective validity of such claims is understood as the expression of a potential accord among those affected by a given norm, freely achieved by means of argument. A procedural criterion – the redeeming of normative validity-claims by means of argument – is substituted for material criteria such as are characteristic of the conventional form of moral consciousness. The following quotation from Habermas describes the ontogenetic development of a post-conventional moral consciousness in a suggestive manner:

If we imagine for the sake of argument the phase of adolescence as concentrated into a single moment of time in which, for the first time, our adolescent adopts a hypothetical attitude towards the normative contexts of his life-world which enables him to see through everything unmercifully, then this illustrates the *nature of the problem* which every individual must come to terms with on passing from the conventional to the post-conventional plane of moral judgement. The social world of legitimately regulated interpersonal relations to which he has been naively accustomed and which he has accepted in an unproblematical fashion is suddenly deracinated, stripped of its natural validity.

If our adolescent is neither able nor willing to return to the traditionalism and unquestioned identity which characterized the world from which he has come, then he must reconstruct the ordered normative relationships which have disintegrated under this penetrating hypothetical gaze from first principles – on pain of total disorientation. Out of the ruins of devalued traditions, which have been revealed as conventional and in need of justification, a new edifice must be constructed that is capable of withstanding the sober critical gaze of one who will henceforth be unable to do other than distinguish between socially accepted norms and valid ones, between norms that are acknowledged as a matter of fact and norms that actually *deserve* to be so acknowledged. Initially it is with reference to principles that the new edifice can be planned and valid norms can be generated; but ultimately there remains only a procedure for the rationally motivated choice between principles which have in turn come to be recognized as standing in need of justification. When measured against the moral actions of everyday life, the change of attitude which discourse ethics demands with respect to the procedure it singles out (i.e. the transition to argument) retains something unnatural; it signifies a break with the naivety of spontaneously held validity-claims upon whose intersubjective recognition the communicative practice of everyday life depends. This unnaturalness is like an echo of that developmental catastrophe which the devaluation of the world of tradition signifies historically – and which is what prompts us to attempt a reconstruction on a higher plane.⁴²

Thus, for Habermas, the transition to a post-conventional moral consciousness is equivalent to the discovery that there exists no possible basis for normative (or cognitive) validity beyond the medium of rational argument. Post-conventional moral consciousness is the result of a reflexive insight into the conditions

for the possibility of normative validity. This thesis marks the starting point for Habermas's reformulation of Kant's moral principle, i.e. for his reformulation of the universalization principle in the terms of discourse ethics.

This reformulation of the universalization principle (U) runs as follows:

Thus every valid norm must satisfy the condition that the consequences and side-effects which result (foreseeably) from its *general* observance with respect to the interest of *every* individual can be accepted (and preferred to the effect of known alternative possibilities for regulating action) by *all* those affected. (DE 75f)

Habermas also calls the universalization principle a 'rule for argument'. As a rule for argument, principle (U) determines the *aim* of arguments in the context of moral disputes; we might say that it determines the sense of the validity of the moral 'ought'. Now Habermas maintains that this rule for argument cannot be applied in a 'monologic' fashion, but that in the meaning of the term it requires a transition to *real* discourse.

The given formulation of the generalization principle does indeed aim at a co-operative execution of the argument in question. On the one hand, it is only by the immediate participation of everyone affected that it is possible to prevent the interests of individuals from being distorted in the interpretation by the perspectives of others. In this pragmatic sense, each individual is himself the ultimate authority for judging what really is in his own interests. On the other hand, however, the description of his own interests that each individual acknowledges to be accurate must remain amenable to the criticism of others. Needs are interpreted in the light of cultural values; and since these are always elements of an intersubjectively shared heritage, the revision of values by which needs are interpreted cannot be something which the individual determines in a monologic fashion. (DE 77f)

In this passage Habermas characterizes fairly precisely the blind spots of Kant's ethics to which Hegel – arguing from different premises, it is true – had also drawn attention. Much as Habermas's elucidation is intuitively persuasive, however, the reformulation of the universalization principle itself is nonetheless problematic. This is what I wish to demonstrate below. I shall begin with Habermas's reformulation of principle (U), and

then go on to discuss the premise for this reformulation, the consensus theory of truth.

At first glance it appears to be a particular strength of principle (U) that it links questions of morally right action directly to questions of whether norms are just or not. In this way, justice and morality are related to each other from the outset through a concept of what is normatively right that is fundamental to both. This strength of principle (U) is revealed on closer inspection, however, to be a weakness. For principle (U) succeeds in binding justice to morality only at the price of assimilating moral problems conceptually to legal ones. In principle (U), a universalistic moral principle is inscrutably 'shuffled in' with a principle of democratic legitimacy in such a way that it ultimately fails to convince *either* as a moral principle *or* as a legitimation principle. I should like to elucidate this thesis in four stages.

(1) If we try to read principle (U) as a legitimation (or justice) principle, as the formulation of it encourages us to do, then the following difficulty emerges: principle (U) leaves open the question of what it means to say that someone (I) 'could accept without coercion' the consequences that the general observance of a norm would have for each individual; consequently it also leaves open the question of what it means to say that a norm is acceptable to *all* in this sense. It is apparent from many of Habermas's formulations that he understands the notion of 'being able to accept without coercion' in the sense of an *impartial* formation of judgement; and this would imply that a norm is valid if and when all those affected can persuade themselves that the general observance of this norm is 'equally in the interests of all those affected' – as Habermas indeed puts it at another point (cf. DE 76). *This*, then, would be what arguments *aim at* if principle (U) were 'applied' as a rule for argument. In an argument about norms, each individual would try to show everyone else that a particular norm is equally in the interests of all. This being so, principle (U) could be reformulated (and abbreviated) as follows:

(U₁) A norm is valid precisely when the general observance of it by all those affected could be judged to be equally in the interest of all those affected.

According to Habermas, then, we can only discover *whether* this

is the case by means of real discourse.

Let us ask next what is meant by the term 'valid' as it is used in principle (U). There are two possible answers to this question. We might try firstly to continue to read principle (U) as a principle concerning the justice of norms. The answer to our question would then emerge from a remark which Habermas makes in connection with his derivation of the universalization principle, where he writes that 'the sense we associate with justified norms is that these regulate social matters in the common interests of those who may be affected' (cf. DE 103). It would seem an obvious step to interpret 'justified' here as identical in meaning with 'valid'; but if precisely those norms are 'valid' of which it can be shown that they regulate social matters in the common interests of those who may be affected, and if we then apply the criterion for the validity of norms as formulated by principle (U), then we are likely to arrive at the following, quasi-circular reformulation of principle (U):

(U₂) A norm is equally in the interests of all those affected precisely when it can be accepted without coercion by all those affected *as being* equally in the interests of all those affected.

I spoke of a *quasi-circular* (and not simply of a circular) formulation of a principle of justice here because in (U₂) we have to distinguish between different levels on which the expression 'equally in the interests of all those affected' is used. On one level it is presupposed that those affected know the goal their argument must aim for in order to show that a norm is justified; on another level (U₂) is saying that only an uncoerced consensus of all those affected can show *whether* a norm really is 'equally in the interests of all affected'. (U₂) does not provide a plausible interpretation of (U), however, because it in fact contains nothing more than the application of a general consensus theory of truth to the specific case of the concept of justice. To this extent (U₂) is not a specific principle of *justice* at all.

Even setting aside the problems of a consensus theory of truth, to which I shall return, our attempt to interpret principle (U) has so far ended in a cul-de-sac. Evidently our first answer to the question of the sense of 'valid' in principle (U) was incorrect. But then Habermas himself suggests another answer, which leads to an interpretation of principle (U) as a *moral* principle.

(2) The error in our reflections so far lies in the fact that we have been reading the expressions 'norm' and 'observance of a norm', as they occur in principle (U), naively, so to speak. This stands in contradiction to Habermas's own elucidation of the 'grammar' of normative validity-claims. Moreover, Habermas interprets moral 'ought' or 'must' as a 'higher degree' predicate comparable to the predicate 'true' (cf. DE 63). In this case the 'deep grammar' of the statement

'In the given circumstances one ought to lie'

would have to be rendered as

'It is right (commanded) to lie in the given circumstances',

where 'is right' should be understood as the normative equivalent, so to speak, of the expression 'is true'. This is how Habermas arrives at a structural parallel between

'It is true (the case) that p'

and

'It is right (commanded) that a.' (DE 63)

In this sense, normative rightness can be understood as a validity-claim analogous to truth. This interpretation of the sense of 'ought' statements would open the possibility of understanding the word 'valid' as it occurs in principle (U) as identical in meaning with the truth-analogous predicate 'right'. We should then have to read principle (U) in the following way:

(U₃) In situations S, it is (morally) correct (commanded) to do a if such a way of acting could be conceived as universal and could be accepted without coercion by all (those affected) as being equally in the interests of all after taking into consideration its foreseeable consequences for each individual.

A further possible reading would be this:

(U₄) In situations S, it is (morally) right (commanded) to do a if all individuals can will (without coercion) that a corresponding

way of acting – taking into account the foreseeable consequences for each individual – become universal.

In this way the apparent norm predicate 'valid' ('justified') would be replaced with the normative predicate 'right'. Translated into ordinary parlance, (U₃) and (U₄) would thus read:

'One must do a in situations S if ... etc.'

Alternatively we could revert without fear of further misunderstanding, to the formulation contained in (U):

'Every valid norm must satisfy the condition that ...'

Principle (U) has emerged as a genuine moral principle. But how do things stand with the parallelization of 'It is true that p' and 'It is right (commanded) that a' which Habermas presupposes? In the first case we have a valid equivalence of the kind:

'It is true that p precisely when p'

whereas in the second case the equivalence could only take the form:

'It is right (commanded) that a precisely when X',

where X stands for the criterion of validity formulated by principle (U). But this means that the *formal* elucidation of the predicate 'true' in the one case would stand against a *material* elucidation of the predicate 'right' in the other. In other words, the term 'true' would denote that which can justifiably be maintained, without a *criterion* for truth being yielded in the process; but the term 'right' would denote that which can justifiably be *demand*ed in the sense of a quite specific *criterion* of rightness. The *sense* of the validity of (moral) 'ought' would thus be determined a priori by a *criterion* of the validity of moral 'ought'.

There is an obvious basis for comparison with Kant here. Kant, too, had in a certain sense determined the rational *sense* of the validity of moral (categorical) 'ought' by means of a *criterion* of the validity of moral 'ought' (the categorical imperative). According to Kant, as rational beings we have always already recognized such a categorical 'ought' as justified; to act in contradiction of

it would mean acting in contradiction of the conditions for the possibility of our self-respect as rational beings. In this sense, the unconditional moral 'ought' as it is expressed through the categorical imperative represents for Kant a 'fact of reason'. In a similar way, Habermas's elucidation of the *sense* of moral validity by means of a *criterion* of moral validity should be understood as pointing to a universal structure of linguistically mediated intersubjectivity; what is expressed in the unconditional character of moral 'ought' is the fact that our possible identity as creatures capable of speech is tied to such a structure of intersubjectivity. I shall return to this idea later. But first I should like to consider the question whether principle (U) is satisfactory when understood as a moral principle – i.e. in the sense of versions (U₃) or (U₄).

(3) Let us recall that principle (U) should be understood as a reformulation of the categorical imperative in the terms of discourse ethics. In this sense Habermas quotes McCarthy with approval:

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 is shifted from what each can will without contradiction to be a general law, to what all can will in agreement to be a universal norm. (cf. DE 77)⁴³

Thus whereas Kant says that 'we [and therefore I – A. W.] must be able to will that a maxim of our action should become a universal law – this is the general canon for all moral judgement of action',⁴⁴ principle (U) is intended to shift the emphasis from 'I must be able to will' to 'we must be able to will'. And the corollary of this would be that we can only find out through real discourse whether *we* are able to will that a maxim should hold as a universal law. Now when Kant speaks of the universalization of maxims he is not at all concerned with the question of the justice of norms. What Kant is postulating is much rather a requirement that I consider whether I would want to live in a world where, as if by a law of nature, everybody acted (and in particular, acted towards me) in the manner suggested by my maxim, and thus whether I could will that the way of acting expressed by my maxim should become universal. When I use

the term 'way of acting' here and in what follows, it should be understood in the sense of a 'way of acting in a kind of situation'. I have various reasons for preferring the term 'way of acting' (when understood in this way) to 'maxim'; the crucial reason here is that I wish to avoid any suggestion that we are already discussing *norms* at this stage of the argument, and thus that we might be *presupposing* precisely that moral 'ought' whose sense and possibility Kant is trying to *explain*. (In other words we must conceive of maxims, in so far as they represent 'subjective' principles of action, as being formulated without a moral 'ought'.) Now I have already established (section II above) that Kant's criterion for morally right action only makes sense if we understand it in a *negatory* sense. Contrary to what Kant himself supposed, those ways of acting that I am able to will as universal cannot themselves be the ones that I morally ought to follow; it is rather the categorical imperative itself that says what I 'ought to' or 'must' do, namely that I *must not* perform an act *p* in a situation *S* if I am unable to will such a way of acting as a universal one. Thus, if I am unable to will that everyone should lie to me if it is to their advantage to do so, then I may not lie simply because it is to *my* advantage to do so. Now from this it would be possible to derive a moral *norm* to the effect that 'one must not lie' – but with norms like this we should always have at the same time to keep in mind those descriptions of situations from which the non-generalizability of a way of acting becomes apparent.

If we understand the categorical imperative in this way, the 'monologic' character of Kant's moral principle is not such a serious problem as it appears to Apel and Habermas. For if I am unable to will that a way of acting should become a universal rule, then *we* cannot will it either (otherwise *I* should be able to do so as well). We might equally express the point like this: in moral judgement I am above all confronted with myself. But the question I always have to answer at that point is clearly of a different kind from the question of whether a social norm is just or not.

The objection is nevertheless valid that Kant was mistaken in assuming that a *serious* moral judgement was *eo ipso* also *intersubjectively valid*, and thus that *my* 'being able to will' or 'not being able to will' must necessarily coincide with that of all other rational beings. Kant was only able to make this assumption

because he immediately went on to develop the fruitful idea of the *Groundwork* in a formalistic manner. But if a monologic application of the categorical imperative does not guarantee the intersubjective validity of moral judgements, then it does indeed seem an obvious step to formulate Kant's *assumption* in the form of a *postulation* of the sort, 'Act in such a way that your way of acting could be willed by *all* as a universal one.' This is how we should also have to understand the reformulation of the categorical imperative by McCarthy which Habermas quotes.

Now principle (U) seems at first glance to be saying the same thing: a way of acting is right if, when understood as a universal one, it would be acceptable to all (those whom it would affect). (U₄) is the version which most nearly corresponds to this reading. We should have to understand the word 'right', however, as it occurs in (U₄), in the sense of 'morally permitted' instead of 'morally commanded'. For if my comments on Kant were correct, it makes no sense to assume that the ways of acting that we are able to will as universal ones are also the ones that are morally *commanded*. There is no need for us to pursue that particular dispute further, however, since the reading of principle (U) that we now have at least comes very close to a reformulation of the categorical imperative such as this:

'Act only according to maxims of which *we* are able to will that they should hold as universal laws.'

I believe that (U₄) is the reading of principle (U) which is least encumbered with premises rooted in consensus theory. It is therefore to this reading that I shall return later on.

(4) It is not (U₄), however, but much rather (U₃), that corresponds to Habermas's idea that in an argument about moral norms each individual should impartially assess a norm with a view to deciding whether the observance of that norm is equally in the interest of all. Let us therefore return to (U) as read in the light of (U₃). If we understand (U) as an elucidation of our provisional conception of moral validity, then this means that our moral convictions and our moral judgements must contain implicit judgements to the effect that the consequences and side-effects that the universal observance of a norm would have for each individual could be accepted without coercion by all. But this, it

seems to me, would make it impossible for us ever to arrive at fully justified moral judgements. Let us take as an example norms like '*Neminem laede*' or 'Thou shalt not lie'. Whereas a simple reflection in the spirit of the categorical imperative yields the result that I cannot will that in the world in which I am living people tell untruths or do injury to each other when it pleases them, a corresponding reflection in the spirit of principle (U) presents us with enormous problems. For the sake of simplicity I shall argue from the premise that, given ideal conditions for discourse, all human beings would agree that the *universal* observance of both the above norms would be equally in the interests of all if *ideal* conditions for achieving agreement could be assumed. But of course this still says very little about how we should act in *real* conditions of communication, i.e. in historical reality as it actually exists. If we try to apply principle (U) as a principle for judging action in *non-ideal* conditions, however, the following difficulties result:

(a) Let us first try to clarify what the consequences and side-effects for each individual would be if the norm 'Thou shalt not lie' were observed *universally*, which, if the words 'norm' and 'universal' have any meaning here, can only mean *without exception*. Kant was able to prohibit lying *universally*, i.e. without exception, because he did not concern himself with the consequences. But if we do concern ourselves with the consequences, and if we assume that for the rest the world remains as it is, then we must suppose that the consequences of universal truthfulness would be harsher for the victims than for their persecutors. To this extent the norm 'Thou shalt not lie' – things being as they are – could not possibly be a valid one. In order to find out what is the right way of acting under the *given circumstances*, we should clearly have to formulate more complicated norms with qualifications and exception clauses, much along the lines that Hare has postulated (although in his conception of it this would be a never-ending task).⁴⁵ But this increases enormously the difficulty of the task of determining the consequences and side-effects of a *universal* observance of norms for *each* individual and, beyond that, of finding out whether *all* would be able to accept without coercion these consequences and side-effects, as they would arise for each individual. Even real discourse cannot ultimately help us here. For as long as we have to conduct our discourse under conditions in which the

victims have to protect themselves from their persecutors by telling lies, it is not possible to imagine any uncoerced consensus existing. Conversely, if it were possible in fact to achieve a general consensus, then the conditions that made exceptions and qualifications necessary would immediately fall away. In any case, it evidently makes no sense to assume that, under non-ideal conditions for reaching agreement, we should be able to solve our real moral problems by trying to achieve *real* consensuses. Where the possibility for reaching agreement ends, the only course that remains open to us is to consider what rational and competent people *would* say, or what those affected by our actions would say *if* they were sufficiently rational, possessed of good will, and competent to judge. And in *this* sense it is of course true that every moral judgement anticipates a possible rational consensus. But *if*, in the course of every moral reflection (which must ultimately always be monologic in nature), we had to reach a decision on the question whether the consequences and side-effects arising for each individual from the universal observance of a norm – and this would require in turn the *formulation* of a universal norm – could be accepted without coercion by all, then we should never be able to arrive at a fully justified moral judgement.

(b) Another kind of difficulty arises if we consider norms such as '*Neminem laede*' of which it may be assumed that it must be possible to achieve an uncoerced consensus to the effect that the universal observance of such norms is equally in the interests of all, and is so, moreover, precisely if we argue *on the basis* of the non-ideal conditions in which we live. It is meaningful in this instance to assume the possibility of achieving such a consensus because the *universal* observance of a norm such as '*Neminem laede*' would enable us to discount those conditions under which exceptions and qualifications are in fact necessary (e.g. in cases of legitimate self-defence, punishment etc.). But this is precisely why principle (U) produces false results here, namely ones which offend against our intuitive moral judgements. For the – hypothetical – assumption of a universal observance of the norm means in this case that questions of moral rightness are answered with reference to *ideal* and not *real* conditions for action. (U) would thus require us to act in the way that, as far as we are able to judge on the basis of our hypothetical assumption, we

really would act if the conditions for acting and achieving agreement were ideal. We might be surprised at this point to find a problem reappearing that also plays a central part in Kant's ethics: Kant's 'practical laws' are in essence norms of action for the members of a potential kingdom of ends. Kant was at least consistent and categorically disputed the possibility of *exceptions* (to the prohibition of lying, for example). But precisely *this* position is not available to discourse ethics; it would contradict the very foundations of its argument.

There is a possibility for escaping the difficulties outlined here, and that is, as it were, to 'de-dramatize' the concept of a norm. We might understand the word 'norm' in the sense of Hare's 'prima facie norms', for example. Principle (U) would then only be concerned with the grounding of *those* norms for which it ought to be possible to achieve an uncoerced consensus to the effect that the universal observance of it *would* be equally in the interests of all under *ideal* conditions of action and communication. Everything else would be a problem of the correct *application* of such norms to a non-ideal reality. But quite apart from the problems which reside in the idealized concepts themselves that we are here presupposing as having been adopted (I shall return to his question in my next section), it seems clear to me that the possibility we are considering is no way out in reality. To give only my decisive reason for saying this, principle (U) would lose its point if the application of it were limited to that area of basic moral judgements for which Kant gives us more or less adequate guidance. After all, principle (U) is *intended* precisely as a principle for judging *such* norms as Kant cannot accommodate *because* he is understanding morally valid norms as norms of action for the members of a kingdom of ends. But if *this is* correct, then the distinction between problems of the *grounding* of norms and problems of the *application* of norms loses its significance here.

The problems and uncertainties we have encountered in our discussion of principle (U) can, I believe, be attributed to *two* problematic assumptions that Habermas makes. The first (a) is that questions of morally correct action can be treated in a similar way to questions of the justice of norms, and the second (b) concerns the way that discourse ethics is rooted in consensus theory.

(a) Habermas's formulation of the moral principle assumes that in our moral thinking we address the same question as in a discussion about the justice of social norms which we are in a position to introduce or refrain from introducing, to set aside or preserve. In a discussion *such as this* it really is a matter of deciding whether all those affected by a norm ought, as impartial judges, to be able to accept the consequences which would arise for each individual from the universal observance of that norm, and thus whether the introduction or preservation of a norm is 'equally in the interests of all'. A paradigm case of such an *introduction* of a norm would be the unanimous resolution of a body of people to deal with matters of common interest according to a particular set of rules. We can see in a case like this that while the hypothetical assumption of a universal observance of the norm does play a part in assessing the *justice* of the norm in question, another step has to be added beyond the assessment of the justice of the norm – and in this case it is a formal *resolution* – in order to constitute a corresponding *obligation* to act. This obligation, resulting from a common resolution, can be understood as a *moral* obligation; but it clearly cannot be grounded in the same way as the assessment of the justice of a norm which can *in principle* be introduced or set aside by means of a formal resolution.

Kant was well aware of the distinction we are making here, and it is therefore possible to elucidate it with reference to the different approach to the question of an uncoerced consensus of rational beings that is respectively implied in the formulation of the moral principle by Habermas and Kant. As we have already established, the content of a fully grounded consensus in the sense implied by principle (U) would be the judgement that the universal observance of a certain norm is equally in the interests of all (those affected). By contrast, however, the content of a 'Kantian' consensus in the case of moral norms would be that *we* (as rational beings) are unable to *will* that a certain way of acting should become universal. Here I am agreeing with Habermas that Kant's 'I' is to be replaced with a 'we', regardless of the problems that may result. On this assumption, the rational consensus that we would expect to find in a moral judgement from a Kantian point of view would take the form of saying, 'Neither I nor any one of "us" can rationally will that *this* way

of acting should become universal.' The word 'rationally' here governs the 'being able to will'; whether we are *in fact* able to will something depends on our interpretations, convictions, and our understanding of ourselves, and *these* things can be more or less 'rational', i.e. appropriate, justified, correct or even truthful. 'Rationally', then, means something like 'if we see ourselves and the world and the situations of others correctly'. In this sense, it is possible to conceive of the argumentative or indeed communicative resolution of disputes, or of learning processes in the medium of argument, without difficulty. But if we understand the discursive dimension of morals in *this* way, then we can get by without a consensus theory of truth; we can construct instead, as I shall show below, a *fallibilistic* interpretation of the rational consensus we might expect to find in moral judgements. If on the other hand we understand the content of the rational consensus which we 'anticipated' in moral judgements to be the common assessment of the justice of a norm, then it is difficult to see how such an idea could be worked out other than in terms of a consensus theory. This brings me to the second of Habermas's problematic assumptions, as mentioned above.

(b) Since I shall be dealing fully with the consensus theory of truth in my next section, all I should like to do here is point once again to the problems that arise as a consequence of Habermas's commitment to consensus theory. I pointed earlier to the paradoxes that result from the opposition of *ideal* and *real* conditions for communication and discourse, which is effectively built into principle (U). This opposition is a direct expression of premises drawn from consensus theory. My critique of consensus theory will therefore also entail a critique of the idealizations upon which it is based. What we have seen so far is that these idealizations of consensus theory also lead to *internal* difficulties in discourse ethics, difficulties which constitute an affinity with Kant that was certainly *never intended* by discourse ethics. It remains to be shown, however, that it really is the premises of consensus theory themselves that lie at the root of the difficulties of discourse ethics, and no mere contingencies of Habermas's formulation of principle (U).

VII

The fundamental thesis of the consensus or discourse theory of truth in Habermas's version of it is that the validity-claims which can be called 'true' or 'valid' are precisely those upon which it would be possible to achieve a discursive consensus under the conditions of an ideal speech situation. In Habermas's view, the structures of an ideal speech situation (which he also claims is actually *assumed* in any serious argument) are characterized by an even distribution of chances to perform a variety of speech acts, as well as by a tolerance towards any change in the level of discourse.⁴⁶ The fundamental thesis of consensus theory thus defines the 'rationality' of consensuses in terms of the formal structural features of an ideal speech situation, and it also defines 'truth' as the content of a rational consensus. By contrast, I should like to show (1) that the rationality of consensuses cannot be characterized in formal terms, (2) that the rationality and the truth of consensuses need not coincide, (3) that the rational consensus therefore cannot be a *criterion* of truth, and finally (4) that if consensus theory is interpreted as not providing a criterion of truth it is rendered, if not vacuous, then at least unsuitable for supporting a universalization principle within discourse ethics.

(1) My thesis is that our assessment of consensuses as rational is dependent upon our assessment of whether our (own or common) reasons are appropriate. This dependence is a *logical* (conceptual) one: the concept of a consensus achieved on the basis of reasons presupposes the concept of a personal conviction achieved on the basis of reasons. It must, of course, be conceded that we cannot consider a consensus rational if we have reason to suppose that some of the participants only appear to agree, or that they do so out of fear or because of some psychological block. To this extent, Habermas's criterion would be correct in a weaker sense: the concept of a rational consensus necessarily implies that it rests on good *reasons* and not on fear etc. But the same is already true of the convictions of individuals: the rationality of these may be judged by whether or not they rest on good reasons.

Now in a trivial sense it is certainly the case that we should consider a conviction that is arrived at in common to be true, and we should hold this opinion on the strength of the reasons

or arguments which we have all found convincing. And in so far as we have really become convinced of something in common, we are able to speak of a *rational* consensus. Thus it can appear as if a *rational* consensus is necessarily also a 'true' one. But this is only the way it looks from the point of view of those who are actually involved in the situation: if I have reasons for agreeing, then this *means* precisely that I consider a validity-claim to be *true*. But the truth does not *follow* here from the rationality of the consensus, it follows from the appropriateness of the reasons which I can advance for a validity-claim, and I need to have convinced myself that these reasons are in fact appropriate *before* I can speak of the rationality of the consensus. Now in principle it is always possible for such reasons to turn out *after the event* to be insufficient. But if that happens it cannot possibly be identical in meaning with the discovery that an earlier consensus was not *rational* in the sense that the conditions of symmetry and free opportunity which characterize an ideal speech situation were not realized on that occasion. If it is supposed to be possible to characterize these conditions in formal terms, then it is precisely *not* the case that we may allow our assessment of whether such conditions obtain to depend on which reasons we consider appropriate in the given situation. Otherwise the significance of consensus theory as providing a criterion would evaporate. But quite apart from this, there are strong reasons for not equating the rationality with the truth of consensuses. Why should the consensus among leading physicists of the nineteenth century about the truth of Newton's theories not have been *rational* (in the sense of the conditions of an ideal speech situation)? The answer cannot be that physics has advanced since their day.

(2) The truth of consensuses cannot follow from *their* rationality any more than their *untruth* can automatically follow from their lack of rationality (unless we revert to a tautological explanation of concepts). It is only from the point of view of those involved in the situation that consensual rationality appears to be identical with truth. But this cannot mean that the rationality of the consensus is an additional *reason* for it being true. To say this would be just as false as if I were to advance the fact that my conviction is well founded as an *additional* reason for its truth alongside the reasons that I have for holding it. The fact that I hold something to be true with good reason cannot be an

additional reason for the truth of the thing I hold to be true, at least not *for me*; and by the same token, the fact that *we* hold something to be true with good reason cannot represent *for us* an additional reason for the truth of the thing we hold to be true. In other words, the fact that a consensus exists, even if it were arrived at under ideal conditions, cannot be a *reason* for the truth of the thing that is held to be true. We are then thrown back, however, upon the reasons or criteria for truth which are always available to us if we understand the meaning of validity-claims. We should only be able to deduce the truth of consensuses from their rationality *if* we were to include among the conditions for an ideal speech situation the requirement that all participants have an adequate *capacity for judgement*. But then it would no longer be possible, for one thing, to characterize the conditions for an ideal speech situation in *formal* terms, and for another thing, the consensus theory of truth would effectively be reduced to the thesis that true validity-claims are precisely those upon which an uncoerced consensus can be achieved among those who are sufficiently capable of judgement. But this thesis would be void of any substantial content. Consensus theory as a substantial theory of truth stands and falls with a *formal* characterization of the rationality of consensuses; but precisely this formal condition of rationality renders it false. If, on the other hand, we take the step that suggests itself in the light of what we have seen here, and try to understand the concept of rationality in non-formal terms, then consensus theory stands revealed as empty of content.

(3) More recently, Habermas has distanced himself from the interpretation of consensus theory as providing a *criterion*.⁴⁷ Moreover he concedes that in a certain sense we must always know in advance what constitutes good reasons in order to be able to argue at all. But whether such 'good' reasons are ultimately *sufficiently* good reasons, he now says, is something which only 'shows' itself under the conditions of an ideal speech situation.⁴⁸ In the light of this new turn in Habermas's thought, I should like to clarify once again the real point of consensus theory as I understand it. If Habermas says that only a consensus achieved under the conditions of an ideal speech situation can 'show' whether our arguments really are sufficiently good arguments, then what he is describing is a specific reassuring function of

consensuses. By achieving a consensus we reassure ourselves collectively that we (each of 'us') really do see things from a public, a general point of view, i.e. that our judgement is not distorted by idiosyncrasies, psychological blocks, emotions, wishful thinking, impaired faculties etc., *and* that our convictions or reasons would stand the test of a renewed discourse among persons of adequate good will and competence to judge. A consensus reassures us that we have not left the territory of a common world or a common language, or that if we did (as in some sense happens again and again in science and philosophy), then we have done so for reasons which make a new and better commonality possible. Now it is possible to interpret this internal relationship between the validity of truth-claims and the commonality of a world in a variety of ways; consensus theory is the attempt to understand it in a *non-relativistic* way. In order to clarify the point I should like to distinguish between two forms of commonality or mutual accord in the medium of language. The first form is the kind of commonality which we must always *presuppose* in language. As far as this sort of commonality goes, we can say, as Wittgenstein does, that what is 'right' or 'wrong' in our use of words or in our judgements is ultimately determined by intersubjective practice. Agreement among adult speakers of a language is thus in a certain sense the criterion of whether a word is being used rightly or wrongly, or whether an assertion is true. As Wittgenstein says, 'If language is to be a means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgements.'⁴⁹ But what is implied here is not a matter of consensuses achieved by means of argument, but a mutual accord inherent *in* the language which makes arguments possible in the first place. It happens from time to time, of course, that such a 'natural' accord *in* the language is challenged with good reason; science, for example, might be understood in certain respects as a continual process of language *criticism* taking place in the medium of argument. The thought therefore suggests itself that 'natural' accords *in* the language might in principle be replaced by an accord achieved by means of discourse *upon* the appropriateness of basic concepts and rules of language, in short, of our interpretation of the world through the medium of language. In his essay 'Wahrheitstheorien', Habermas has in fact envisaged such a discursive form of linguistic change and language criticism as something that is possible and

in a certain sense necessary. As Habermas puts it, we can only speak of true utterances in the full sense where the language in which we formulate such utterances is itself 'appropriate'⁵⁰; but we should only be able to speak of an 'appropriate' language where the development of that language had itself taken place within the medium of argument, which is to say, where the rational consensus about validity-claims contains within it a rational consensus about the appropriateness of the language.⁵¹ In this way, the *prior* accord in the language which Wittgenstein analyses would, as it were, be drawn into (and transformed through) the discursive revision of our convictions. It is only on the basis of such an assumption that the point of consensus theory becomes entirely clear. For if we concede that, at an initial stage, the accord among speakers of a language is a kind of provisional ultimate standard of the truth or falsehood of utterances, and if we concede that a discursive revision of such an accord is in principle possible, then we can easily go on to say that, if not a factual consensus, then at least a rational consensus (i.e. one achieved by means of discourse) is an *ultimate* authority by which we reassure ourselves of the truth of our validity-claims. At the same time it becomes clear why the rationality of such a consensus may only be characterized in formal terms. There appear to be only two possibilities: *either* we say that each language, each life-form contains within itself its own standards of 'true' and 'false', with the implication that it is no longer possible to put meaningful questions about the truth or falsehood of these standards (this is the answer given by Peter Winch on the level of cultural comparisons, and by Thomas Kuhn on the level of comparative theory); *or* we resist this deeply disturbing relativistic thesis and hold fast to the unconditional nature of validity-claims, and thus to the possibility of a standard which transcends each particular language and each particular life-form. A counter-thesis to the relativistic position would thus be that it is not the *factual* agreement of the speakers of a language that provides the ultimate standards of truth and falsehood, but only that agreement which can be interpreted as a rational consensus. And here it must be apparent that what is meant by 'rational' cannot be explained by the substantive standards of rationality of *one* particular culture; rather it must be defined by purely formal characteristics. It is therefore entirely consistent with the anti-relativistic approach of the consensus theory of truth when

it makes the structural characteristics of an ideal speech situation the defining feature of rationality.

(4) Now, what I have said earlier about the function of rational consensus in providing criteria applies equally to that 'showing' or reassuring function which Habermas has recently emphasized. For the fact that we can reassure ourselves that our reasons really are good reasons by achieving a consensus through discourse does not alter the fact that each consensus is only provisional. But if it does not necessarily follow from the fact that reasons show themselves within the context of a finite rational consensus to be sufficiently good reasons that they will prove themselves *in the long term* to be sufficiently good reasons, then the undisputed reassuring function of consensus is not adequate to sustain the heavy burden of a consensus *theory* of truth.

An obvious way out of this difficulty is to transfer the function of consensus as providing criteria or reassurance to an *infinite* rational consensus.⁵² An infinite rational consensus would be one for which it would *never* be possible to present reasons which cast doubt upon it. In this case, therefore, we can discount the problem that results from the fact that every *finite* rational consensus is only provisional and therefore incapable of providing proof positive of 'truth'. As I have shown above, this problem can only be circumvented in Habermas's version of consensus theory by including among the structural characteristics of an ideal speech situation a sufficient competence to judge on the part of all participants. But then it would no longer be possible to characterize ideal speech situations by means of purely formal descriptions of their structure; this is precisely what is necessary, however, if the notion of a 'consensus under the conditions of an ideal speech situation' is to be the substantial explicator for the concept of 'truth'. If, on the other hand, we conceive the consensus notion of truth in terms of an *infinite* consensus, this problem falls away, for the idea of an *infinite* rational consensus contains within itself not only the assumption of rationality, but also the assumption that no new arguments will arise (and of course that none will be suppressed).

Now, an infinite rational consensus is not only incapable of providing criteria, but cannot strictly speaking fulfil a reassuring function either. For it is not an 'object of possible experience',

but an idea that points beyond the boundaries of possible experience. This brings consequential changes for the possible *meaning* of a consensus theory of truth: if the guarantee of truth does not reside in *every* rational consensus, but only in an *infinite* rational consensus, then the theory loses that explicative substance that Habermas would like to give it. This can be seen from Habermas's most recent exposition of the basic idea of consensus theory.⁵³ 'The core of the discourse theory of truth,' Habermas now says, 'can be formulated by means of three basic concepts':

conditions of validity (which are fulfilled when an utterance holds good), *validity-claims* (which speakers raise with their utterances, for their validity), and *redemption* of a validity-claim (in the framework of a discourse which is sufficiently close to the conditions of an ideal speech situation for the consensus aimed at by the participants to be brought about solely through the force of the better argument, and in this sense to be 'rationally motivated').⁵⁴

The crucial point of consensus theory now consists in elucidating what is meant by 'fulfilling conditions of validity' with the help of the other two basic concepts:

An utterance is valid when its conditions of validity are fulfilled. . . the fulfilment or non-fulfilment of conditions of validity can only be ascertained by means of the argumentative redemption of the corresponding validity-claims. The discourse theory of truth, then, explains what it means to redeem a validity-claim by an analysis of the general pragmatic preconditions for the attainment of a rationally motivated consensus. This theory of truth provides only an explication of meaning, it does not provide a criterion; in the end, however, it undermines the clear distinction between meaning and criterion.⁵⁵

If we understand the 'redemption' of validity-claims here to mean the achievement of a consensus by means of argument under the conditions of an ideal speech situation, then the objections to consensus theory expounded above remain in force. If, on the other hand, we transfer the function of guaranteeing truth to an infinite rational consensus, then it is strictly speaking no longer possible to speak of a *redemption* of validity-claims; and as a result, the explicative connection between the three basic concepts that Habermas constructs would disintegrate. But this difficulty cannot be circumvented by, as it were, merging the par-

ticular (empirical) consensus with the infinite consensus. If we say that a rational consensus – *qua* rational consensus – is by definition capable of infinite repetition, then in reality we are making the consensus under the conditions of an ideal speech situation, and not the infinite consensus, the guarantor of truth, and all the objections which I have made against *this* version of consensus theory remain in force. This means that the possibility of an infinite consensus cannot follow simply from the fact that a consensus has been achieved under the (formally characterized) conditions of an ideal speech situation – this was precisely the point of my objections. To assume the possibility of an infinite consensus means something more in reality – or something other – than to assume the rationality of a particular consensus in the sense of the formal characterization of an ideal speech situation. This 'something more' is connected with the fact that, to repeat the point, the *concept* of a consensus achieved with arguments cannot meaningfully be equated with the *concept* of a consensus against which it is not going to be possible to bring forward any apposite arguments at any time in the future. Otherwise we should have to include the condition that all *possible* arguments have been taken into consideration among the conditions of rationality for *finite* consensuses. But this is impossible, unless we were to make the possibility of an infinite consensus the *criterion* that decides whether the conditions of rationality for finite consensuses are fulfilled. But in this case it would no longer be possible to characterize these conditions formally, i.e. in terms of the *procedure* of argument and the *structural characteristics* of an ideal speech situation.

As we can now see, it would be possible to distinguish between a stronger and a weaker version of the consensus theory of truth. The weaker version is the one which sees the guarantee of truth as residing in an infinite rational consensus. It is impossible to resolve all the differences between the two versions of consensus theory because it is impossible to derive from the formal characterization of ideal conditions for discourse any guarantee that a consensus attained under such conditions will stand the test of an *infinite* discursive examination. But might it not be the case that, via the weaker version of consensus theory which, as we have said, can no longer be understood to provide criteria, we could justify the strong background assumptions which underlie the attempt by discourse ethics to reformulate the universaliz-

ation principle? This is the question I wish to pursue in my next section.

VIII

I have so far refrained from expressly linking the two versions of consensus theory which we have differentiated above with the names of Habermas and Apel. The reasons for this are firstly that both authors have to a certain extent identified themselves with *both* versions of consensus theory, and secondly that Apel's version of consensus theory goes beyond what I have here called the 'weaker' version in one essential respect. This weaker version of consensus theory might be understood as the elucidation of the internal connection between the idea of truth and the idea of a possible universal, rationally grounded accord. These two ideas, we might say, *mutually* elucidate each other: the idea of truth necessarily implies that it will not be possible at any time in the future to advance apposite arguments against what we now see to be true, and this in turn implies that it will not in future be possible to find good arguments with which to call into question the way we talk about the world and formulate our problems. On the other hand it is difficult to see in what sense an *infinite* grounded consensus might not also be called *true*; at any rate we could argue that this would only be conceivable if we were to introduce the problematic notion of a truth that could not be recognized or could not be expressed in language.

Now Apel's version of consensus theory differs from this 'weaker' version in that he elucidates the idea of an infinite (grounded) consensus with reference to the idea of an unlimited *ideal communication community*, which fulfils much the same function within Apel's theory as the idea of an ideal speech situation does within that of Habermas; that is, it represents both a constitutive necessary *assumption* for the situations in which arguments take place, and also a future-orientated ideal or regulative idea. In either case, the idealizing assumption (or anticipation) simultaneously characterizes the conditions which guarantee the rationality of possible consensuses. We have already established, however, *firstly* that the fact that conditions for discourse are ideal can give no guarantee of the truth of consensuses as long as these consensuses are particular (i.e. finite and empirical), and

secondly that *in actual fact* we must always judge the rationality of consensuses according to the reasons upon which they are founded. From all this it may be seen that the idealization implied in the notion of an ideal communication community gets us nowhere: it contributes nothing to our understanding of what a rationally grounded consensus – or even an infinite grounded consensus – is. On the other hand the notion of an ideal communication community does intimate the possibility of a future location for final and absolute truth, it intimates the idea of an ultimate language in which not only will science have reached its final end, but humanity too would have become fully transparent to itself. To be sure, these are only regulative ideas for Apel; but as regulative ideas, they constitute for him ideal ultimate goals which humanity has the task of realizing, and which it is within humanity's grasp to realize, if perhaps only in the sense of an infinite approximation.⁵⁶ What started as presuppositions about speaking and arguing have become ideals of reality which we, as speakers and arguers, are inescapably committed to realizing: here we have arrived at the heart of discourse ethics.

That there is something wrong with this construction has so far only been apparent from the curious redundancy of idealizing concepts when it comes to the problem of truth. Now, I believe that it is perfectly possible to concede that these idealizing concepts are *rooted in* unavoidable idealizing presuppositions about speaking and arguing, as Apel and Habermas state; but I suspect that their way of assimilating and interpreting such idealizing presuppositions is misleading. It is easy to concede that the anticipation of an infinite consensus – like the assumption of an 'ideal speech situation' – comes into play in any consensus achieved by discursive means. But it seems to me that unavoidable assumptions of this kind are hypostatized by the consensus theory of truth in much the same way that formal semantics hypostatizes the equally unavoidable assumption that our words and sentences have a definitive intersubjective meaning. In my view, such unavoidable assumptions about speaking and arguing are vested with a quasi-transcendental, dialectical semblance (*Schein*): in adopting them, we *forget* that linguistic meanings and insights which are capable of linguistic formulation are essentially historical in character, as we can assure ourselves *reflexively*. Only in cases where hermeneutic problems and problems of linguistic expression become marginal – as in mathemat-

ical physics, for example – can we understand the unavoidable assumptions about speaking and arguing in a more or less *realistic* sense, as Apel and Habermas do, for it is only in such cases that the idea of an ‘ultimate’ language, as it is contained within the idea of an infinite consensus in an ideal communication community, makes sense, at least as a *regulative* idea.⁵⁷ It is no mere chance that Peirce, whose thinking has inspired Apel’s version of consensus theory in particular, developed ideas of this kind in connection with advances in the field of physical inquiry especially. The interpretation of scientific progress as a process of continuous language criticism governed by the regulative idea of an ultimate, ‘correct’ language (or an infinite consensus) is part and parcel of the pragmatic reformulation of transcendental philosophy which Peirce initiated. But it seems to me that a pragmatically revised transcendental philosophy which tries to ‘generalize’ Peirce’s regulative principle of an infinite consensus within the ‘community of investigators’ into the idea of an ideal communication community, as Apel does, must ultimately remain trapped within an objectivistic notion of knowledge and experience; and it does so because it fails to see through the dialectical semblance (*Schein*) in which idealizing assumptions about linguistic communication are veiled. The deception arises, not from the fact that such assumptions invariably turn out to be false (whenever our utterances turn out to be incomprehensible, or the situations in which communication takes place turn out to be distorted), but from the fact that these assumptions present themselves to us as ideals to be pursued in human communication and praxis, and thus conceal the historical and imperfectible nature of linguistic meaning. Even after its transformation at the hands of language pragmatists, transcendental philosophy has not severed its contacts with the scientific tradition of the European Enlightenment, which found its classical expression, after all, in the transcendental philosophy of Kant. Even in its pragmatic form, transcendental philosophy remains, as it were, locked into figures of thought which may not have been derived from the way physical science has progressed, but which are certainly to some extent tailored to the course of such progress. This is admittedly a strong thesis, which I should like to elucidate with reference to an earlier text by Apel, his important essay on ‘Scientism or Transcendental Hermeneutics?’⁵⁸

In this essay Apel tries to show that Peirce’s interpretation of truth as the ‘ultimate opinion’ of an ‘indefinite community of investigators’⁵⁹ can be generalized so as to serve as the ‘*regulative principle of an unlimited community of interpretation which realizes itself in the long run both theoretically and practically*’.⁶⁰ Apel had interpreted Peirce’s notion of the consensus within an unlimited community of investigators as the ‘highest point’ of a pragmatically revised transcendental philosophy, corresponding more or less to the transcendental ‘consciousness as such’ as the highest point of Kant’s transcendental philosophy.⁶¹ The universal realism of Peirce, which is grounded in semiotics and ‘dynamized’ by the logic of investigation, does not seek the guarantee for the objectivity of knowledge in a categorial synthetic a priori grounded in the transcendental I, but in the logic of a process of inquiry which, through the inventive and self-correcting interaction of abduction, induction and deduction, must in due course eliminate all that is false. What is *true* are those convictions which establish themselves intersubjectively *in the long run*, through this self-correcting process, as tenable; and *reality* is the correlate of such true convictions.

The real . . . is that which, sooner or later, information and reasoning would finally result in, and which is therefore independent of the vagaries of me and you. Thus, the very origin of the conception of reality shows that this conception essentially involves the notion of a *Community*, without definite limits, and capable of a definite increase of knowledge.⁶²

It is with reference to this early formulation of Peirce’s that Apel summarizes the essence of Peirce’s transformation of transcendental logic:

In other words, the ‘highest point’ of Peirce’s transformation of Kant’s transcendental logic is the ‘ultimate opinion’ of the ‘indefinite community of investigators’. At this point one may find a convergence of the semiotical postulate of the transindividual *unity of interpretation* and of the postulate of the logic of inquiry concerning the *validation* of experience *in the long run*. The quasi-transcendental subject of this unity is the *indefinite community of experimentation* which is identical with the *indefinite community of interpretation*.⁶³

What replaces the grounding of synthetic a priori principles

in the dynamized version of transcendental philosophy is the demonstration of the necessary validity of synthetic *modes of inference*, namely abduction and induction, *in the long run*.

In a way he has put Kant's *regulative principles* of experience in the place of Kant's *constitutive principles* of experience, on the assumption that the regulative principles in the long run turn out to be constitutive. Thus by shifting the necessary and universal validity of scientific propositions to the end of the (indefinite) process of inquiry it is possible for Peirce to escape Hume's scepticism without insisting with Kant on the necessity and universality of propositions which for the moment are accepted by experts.⁶⁴

In his essay 'Scientism or Transcendental Hermeneutics', Apel attempts to go beyond the scientific limitations, as he sees them⁶⁵, of Peirce's line of inquiry in developing this future-orientated notion of truth. He would like to see the idea of a consensus within an unlimited community of investigators expanded into the idea of 'an absolute truth of understanding in an unlimited community of interpretation and interaction'.⁶⁶ Apel develops this idea initially with reference to the neo-idealistic reinterpretation of Peirce's semiotics by J. Royce, then tries to break free from the idealistic frame of reference and construct a defence against objections from the point of view of hermeneutic philosophy. Against Gadamer, Apel asserts that the regulative idea of an *absolute truth* is fundamental to the activity of understanding meaning, too, and thus to the interpretation of texts, utterances, actions or life-forms.⁶⁷ Now this notion of an absolute truth in the context of interpretation is not something that can be explained in cognitivist terms as the ultimate theoretical conviction of a community of investigators subjecting itself to the methodical discipline of the logical principles of inquiry; rather it is necessary to take account of the moment of *application* in the act of understanding, as described in hermeneutics, and in the last analysis this means taking account of the structure of understanding as something upon which any attempt at a scientific reduction of understanding to a phenomenon within the world of objectifiable facts must founder. It is precisely the point of Apel's pragmatic transformation of transcendental philosophy that it presents the understanding of meaning as a *complementary phenomenon* to the 'scientific knowledge of objective facts'.⁶⁸ If,

then, we are to transfer Peirce's future-orientated concept of truth to the activity of understanding meaning, this is only possible if we replace the regulative idea of an infinite theoretical consensus with the regulative idea of an *ideal communication community*, i.e. with an 'unlimited community of interpretation and interaction' which simultaneously represents an ideal ultimate goal of understanding, the realization of which would be identical in meaning with the 'abolition of all obstacles to communication'.⁶⁹ In the notion of an ideal communication community, theoretical and practical reason converge in the ultimate goal of an ideal situation for communication. The 'absolute truth' of interpretations is only conceivable in connection with the *practical* achievement of such an ideal situation for communication; the practical moment, the moment of application in the act of understanding, necessitates that we relate the truth of interpretation to a living context which has become fully transparent and contains no element of coercion.

It is difficult to resist the fascination of this thought, which occurs in similar form in Habermas: this is where the pragmatic reinterpretation of transcendental philosophy becomes indistinguishable from the pragmatic reinterpretation of Adorno's philosophy of reconciliation. But is the idealization contained in the notion of an ideal communication community a *meaningful* one? Apel formulates the crucial objection himself, albeit in an indirect fashion, when he points to the fact that the infinite consensus of the community of investigators of which Peirce speaks *presupposes* a neutralization of the problem of communication: the 'ultimate' language of physics, the correlate of *ultimate opinion*, is only conceivable as a language which has emancipated itself from the conditions of a hermeneutic mediation of meaning. Peirce's maxim for the clarification of meaning is conceived for precisely this limiting case; it represents the attempt,

to relate all meaning to operations and correlated experiences which every isolated human subject can have at any time, independently of his historical interaction with others, and which – to this extent – are *a priori* intersubjective, and this means they are also objective. One can detect here the basic desire of every progressive empirical analytical science to make the hermeneutic aids of intersubjective communication superfluous for the future by means of a definitive mode of agreement about meaning, and thereby to establish once and for all the preconditions for the possibility and validity of logically and empirically testable theories.

In parenthesis one might add that the ideal of this definitive meta-scientific agreement would be the abrupt replacement of historically constituted everyday language – including the experimentally proven language of science that developed from the latter – by a universal calculus language, which would be both guaranteed to be non-contradictory and experimentally and pragmatically capable of application. This was the original dream of logical empiricism.⁷⁰

Apel here expresses the point very clearly that there is an *internal* connection in Peirce's thinking between the idea of a 'potential unlimited progress'⁷¹ of science governed by the regulative idea of an ultimate consensus among the community of investigators – an ultimate language of physics – and the conception of the community of investigators as a kind of 'singular in the plural'. *Ultimate opinion* will be formulated in a language which will no longer present any problems of clarification of meaning, any problems of communication, *for the precise reason that* all meaning it expresses is related to operations and experiences 'which every isolated human subject can have at any time, independently of his historical interaction with others'. It is only for this reason that the indefinite community of investigators can assume the role of transcendental subject; the progress of science can be understood as the *process of becoming* of this transcendental subject.

What I mean to say is this: Peirce's philosophy can be seen as a transformation of transcendental philosophy precisely because at the 'highest point' of this philosophy reason would have emancipated itself from its dependence on (ordinary) language, and therefore from that condition of 'linguisticity' which transcendental hermeneutics had always insisted upon *against* Kant. To put it in terms that are less open to misunderstanding, at the highest point of this philosophy the language of science would have attained that post-hermeneutic state which was, in Apel's words, 'the original dream of logical empiricism'. The idealization that is at work here concerns not the (pragmatic) structures of communication, but the (ahistorical) intersubjectivity of linguistic meanings. The crucial problem therefore lies in the inconspicuous shift from the idea of an infinite consensus of investigators to the idea of an ideal communication community as the location of an absolute truth of *interpretation*. Are we to think here, too, of an 'ultimate' language in which the truth content of all

philosophical texts would have become *available* as a perfectly transparent one? This would represent the idea of ideal communication in the sense of a *state of perfect understanding and agreement* – a state of communication in which *humanity would* be finally relieved of the trouble of forever having to acquire philosophical or practical truths anew. Or should we think of ideal *conditions* for communication, conditions in which communication and self-expression might still be necessary, but were able to proceed, as it were, without friction? This would represent the idea of ideal communication in the sense of *being willing and able to reach understanding and agreement time and again*, and in so far as the connotation of an infinite consensus is indeed to be preserved, it would also represent the idea of a rational consensus that was forever *renewing* itself. I maintain that the two interpretations we have distinguished here necessarily intersect in Apel's concept of the ideal communication community, so that what is *intended* as a situation of ideal communication stands revealed as a situation that lies beyond the necessity (and beyond the problems) of linguistic communication. This would mean, however, that, within the notion of the ideal communication community, the constitutive plurality of sign-users would be suspended in favour of the singularity of a transcendental subject which has now also attained an understanding of itself in practical, hermeneutic terms – a subject which, having undergone a process of becoming, is now, as it were, *in the truth*.

To clarify my thesis, I should like to try to express more precisely what is meant by 'unlimited mutual understanding' or 'ideal communication'. Apel also speaks of the 'abolition of all obstacles to communication'.⁷² We might start by trying to understand the ideality of communication situations in the sense of Habermas's conditions for an ideal speech situation, an approach which Apel himself occasionally seems to encourage. But we have already seen that the concept of an ideal speech situation, as we have so far understood it, is not adequate to encompass that convergence between successful communication and intersubjective validity which Apel envisages under the notion of an ideal communication community. If the ideal communication community really is intended as a location of absolute truth – albeit merely an anticipated one – then this is only possible if the assumptions that every speaker makes in the act of speaking about the comprehensibility of validity-claims and their capacity

to produce a consensus are constantly *fulfilled* within that communication community. As far as comprehension is concerned, this statement follows directly from the ideality of communication situations as such, and as far as the capacity of validity-claims to produce a consensus is concerned, it follows from the fact that in the limiting case of ideal communication it would no longer be possible for the 'anticipation of perfection' in the act of interpreting texts and utterances to founder on the factual limitations of non-ideal conditions for communication. This is something which Apel indirectly makes clear when he attributes the failure to anticipate the truth 'in the sense of a potential *consensus omnium*'⁷³ when interpreting texts to the non-ideality of factual conditions for communication⁷⁴: hermeneutically speaking, what is untrue is that which cannot be understood.

As Apel says elsewhere:

in a normative sense, the ideal language game of an ideal communication community is anticipated by anyone who follows a rule (implicitly by someone who claims to act in a meaningful way, explicitly by someone who argues) as a *real possibility* of the language game in which he is participating, which is to say that it is presupposed as the *precondition for the possibility and validity* of his action as a meaningful action.⁷⁵

This ideal language game thus betokens the ideal *conditions* for communication just as much as the ideal *result* of a historical process of communication, i.e. an ideal and unlimited state of being in agreement and understanding each other as the ultimate telos of humanity in its *efforts* to communicate at any given time. If this is so – and I cannot see how this conclusion might be avoided – then the idea of an ideal *communication* community must also betoken the idea of an ideal, an ultimate *language* which would enable the assumption of the possibility of intersubjective understanding which we make in any linguistic utterance to be fulfilled at any time. But this is nothing other than 'the original dream of logical empiricism' projected onto the frame of reference of a pragmatic philosophy of language. The ideal communication community would have passed beyond error, dissent, non-understanding and conflict, but only at the price of freezing language, of the extinction of its productive energies, and thus of suspending the very linguistic and historical life-form of humanity.

This is where the profound ambiguity in the notion of an ideal

communication community becomes apparent. To the extent that it represents an attempt to identify the idea of the *absolute* as a 'highest point' within this world, the notion of an ideal communication community still has a strong affinity with that 'original dream of logical empiricism'. By trying to interpret the absolute as the ultimate limit of a potential infinite progress of theoretical, practical and hermeneutic reason, Apel converts it to the image of a reason emancipated from the vicissitudes of its linguistic condition. Adorno was still enough of a theologian to know that such an absolute – which for him, too, represented the condition of the possibility of truth – could only be conceived as an ultimate horizon for the history of reason if it were firmly associated with the idea of a radical *break* with historical continuity: reconciliation, for him, would be the complete obverse of reason as it existed. Apel, on the other hand, having quite rightly built his case against Adorno on the partial nature of reason as it exists and on the possibility of moral progress, cannot resist going one step further, a step which in reality takes him back to a position which Adorno (and Benjamin) had already surpassed. Whereas for Adorno the absolute was 'shrouded in black' (in theological terms it was the kingdom of God), Apel tries to fetch it back into the continuum of history. The messianic perspective is converted back into the perspective of a potential unlimited progress towards the absolute. In the context of the theory of science in Peirce's sense, this perspective does have a certain legitimacy, but when it is transposed onto the historical and moral world as a whole then it is clear that it lacks the redeeming force to which it lays claim. It is not by chance that, from the perspective of a perfected physics, history becomes degraded to prehistory, individuality to contingency, and living language to a mere passing phase; but at least a perfected physics is *conceivable* as a form of knowledge attained by moral beings. The *generalization* of a future-orientated concept of absolute truth, on the other hand, would actually have to delete historical time at the limiting point of the absolute, for a truth to which all eyes are opened is something that must also be shared by those long since dead: the reconciliation of all humanity would have to be something in which the dead, too, participate. But this can only be conceived in theological terms, as Adorno very well knew.⁷⁶ With respect to humanity and its history, the idea of a perfect truth to which *all* eyes are opened is prefigured, not in the idea of a

perfected physics, but in the image of the Day of Judgement. And the image of the Day of Judgement includes the hope of resurrection and salvation. Judgement, salvation and resurrection are categories of a radical break with the historical world; this is precisely what makes them theological categories. To be sure, it would be for philosophy to decipher the power of the images which have crystallized in these categories, but the deciphering offered by the philosophy of the 'ideal communication community' is scarcely more convincing than what Adorno offers with his philosophy of reconciliation.⁷⁷ For regardless of whether the absolute, as the horizon of the history of reason, is conceived of as demanding a *break* with reason as it exists (Adorno), or whether it is conceived as the immanent telos of that history (Apel), in either case it is clearly not recuperable within the boundaries of the historical world.⁷⁸

It is not inadvertently that I compare Apel's philosophy of the ideal communication community with Adorno's philosophy of reconciliation. The common element between the positions of Apel and Adorno is that they both believe that the notion of truth can only be saved if it is conceived from the perspective of a reconciled humanity – an 'ideal communication community'. In both instances the idea of the absolute betokens the precondition for the possibility of truth. For Apel this means that the idea of unlimited progress towards an ideal communication community (as the location of 'absolute truth') offers the only possible alternative to a relativistic dissolution of the notion of truth.⁷⁹ I do not believe that this diagnosis of Apel's is accurate. Rather, I wish to show that the whole problem presents itself in a new light if we put aside the view, which is axiomatic for Apel, that equates the anticipation of an infinite rational consensus with the notion of an ideal communication community.⁸⁰

Apel expressly refers to philosophical statements for which in his view we must necessarily presuppose the anticipation of an ideal communication community in order to understand or raise a corresponding claim to general validity.⁸¹ But validity-claims of the kind that are expressed in philosophical statements are tied to the medium of ordinary language, and to their own context of explication. The argumentative dynamic of philosophizing which is crystallized in philosophical statements, and which is what really lends weight and significance to philosophical theses, is thus something which cannot be 'solidified' once

and for all in the form of philosophical statements or systems of statements. In this sense Adorno was quite correct when he asserted that philosophy cannot be reduced to theses.⁸² But if this is so, then philosophical truths are forever having to be discovered, acquired, thought through and formulated *anew*. Even the great philosophical texts, to which we constantly refer as the paradigm of a written, objectified philosophical truth, only contain a codified form of the truth; and this truth yields itself up to us only through a process of translation in which we, as it were, repeat the original process of their formulation with the means available to us. This is why the interpretation of philosophical texts plays such an important part in philosophy – quite independently of the fact that the interpretation of philosophical texts always implies separating out what is true and false in them, and independently of the fact that it is also possible to have *progress* in philosophy. The crucial point is that any philosophical truth, once uttered, would be lost without the incessant efforts of others to translate it and make it their own. The conservation of philosophical truths is a productive process. Even if it were possible to condense the *entire* truth of philosophy within a single text, we should only be able to preserve it by providing this text with innumerable commentaries; as a mere container of truth, the text would be dead as soon as we ceased to rewrite it.

But if this is the case, then the anticipation of an infinite consensus in this particular instance cannot have the meaning that Apel gives it. Apel's model for the anticipation of an infinite consensus is in the final analysis that of physics, whereby the *ultimate opinion* of the investigators would find expression in an ultimate language and in a stable system of statements. But if every philosophical statement carries an indication of the historical time and place in which it is made, and if the meaning of philosophical statements is a function of the context of which they are a part, then the possibility of an infinite consensus can in this instance really only mean the possibility of an infinite repetition in the sense of reacquiring, reformulating or hermeneutically reconstructing philosophical insights. But in this context the idea of an ultimate limit of *ideal* understanding makes no sense at all, for the 'obstacles to communication' here share the same origins as the preconditions for the possibility of communication itself: both are founded in the dependence of philosophical thought on language. A communication situation could

only be 'ideal' in Apel's sense if linguistic signs had become a completely transparent medium for the communication of intended meanings, so that communication itself had acquired the quality of immediacy. But this would be a state of affairs exceeding the bounds of language.

In the instance of philosophical statements, it is thus not possible to conceive of an 'infinite consensus' as an ultimate and, as it were, 'stable' consensus. Precisely because in this instance there are no rules governing the logic of inquiry which would guarantee truth *in the long run*, it makes no sense to situate the truth at the end of history. It is rather the case that past, present and future are all equally possible 'locations' for philosophical truth. A consensus about philosophical truths among persons of sufficient competence to judge would naturally have to be one that was susceptible to constant renewal, even if mediated through a productive reinterpretation of philosophical texts. But we do not need the idea of an ideal communication community in order to be able to conceive this thought, not even in the form of a regulative principle. There can be no possibility of philosophical insights being 'ultimately' borne out, any more than there can be an 'ultimate' foundation of philosophical truths. This has nothing whatever to do with relativism. The problem of relativism is only produced by the angle of vision from which the philosophy of the absolute perceives the problems of validity. What is needed is a change in the angle of vision, and the problem of relativism would disappear.⁸³

So far I have admittedly only dealt with the problem which Apel emphasizes concerning the possible validity-claim of philosophical statements. It seems to me, however, that it is sufficient to demonstrate with reference to *one* element in the argument that we are not obliged to define our notion of truth in relation to Apel's conception of an ideal communication community. For if it is possible to demonstrate at only *one* point that the immanent criticism of idealizing concepts in Apel's transcendental pragmatics does not necessarily lead to the 'relativistic' or 'historicist' dissolution of the concept of truth which Apel fears, then it is legitimate for us to conclude from this that the problem of relativism had been wrongly posed. We might easily suspect, moreover, that the problem of relativism is merely the abiding shadow of an absolutism which would like to anchor the truth in some Archimedean point lying outside the world of our actual

discourse. Relativism, in this connection, would be the reminder that there can be no such Archimedean point. But if it is true that we can hold fast to the idea of truth *without the aid* of such an Archimedean point, then at the same time as we take our leave of absolutism, we can also bid farewell to its shadow, relativism.

It ought by now to be clear that the 'weaker' version of consensus theory, which I earlier distinguished from Habermas's stronger version (cf. Section VII, above), is not sufficient to justify the strong underlying assumptions which lie at the foundations of the reformulation of the universalization principle proposed in discourse ethics. For if the idea of an infinite rational consensus can be explained independently of the idea of an ideal communication community, then this shows that the idealizing concepts which lie at the foundations of the consensus theories of both Habermas and Apel cannot represent a compelling conceptual reconstruction of inescapable presuppositions involved in speech and argument.

At this point it also becomes clear to what extent discourse ethics has remained *too Kantian*. Just as Kant had to resort to the idea of a kingdom of ends in order to explain the idea of practical reason, so too must Apel and Habermas resort to an ideal communication situation in order to explain the connection between rationality and truth, and with it the concept of practical reason. In either case, the problem lies in the idealizing concepts themselves, or in the fact that they are taken as 'ideals of reality' itself. For if this is how we understand them, then their inherently illusory quality ensures that they immediately become elusive. Just as the kingdom of ends not only betokens a state of affairs in which moral conflicts can no longer exist, but also one where an unbroken unity and communication among subjects would be realized (a state of affairs in reality, then, in which a plurality of subjects would no longer even be conceivable), so too do the formal structures of the ideal speech situation or the conditions of an ideal communication community (if we take them as an ideal telos of a linguistic reality) betoken not only an ideal condition of rational communication, but in reality also an ideal state of mutual understanding and intersubjective agreement – and thus a state of affairs in which, again, the darkness that lies between subjects and within them has been finally dispelled. But without this darkness there would also be no language – unless,

of course, it be the ideal language of the constructive semanticists, which would, however, turn day to night.

IX

In the course of the last two sections, I have already implicitly repudiated the claim of discourse ethics to provide a fundamental grounding for practical reason. But in so far as this repudiation is directed at the arguments of Apel and Habermas, it still stands in need of justification itself. My thesis is that a universalistic moral principle cannot be derived from what Habermas calls 'normatively substantial presuppositions' of argument. I shall here ignore the distinction between a 'strong' and a 'weak' version of fundamental grounding (Apel versus Habermas⁸⁴), since this appears to be only of secondary significance for my own considerations. I shall not discuss Habermas's sketched derivation of a universalization principle from presuppositions of argument because it seems to me self-evident that this derivation is false. Habermas introduces a supplementary 'semantic' premise at a crucial stage (the premise that 'the sense we associate with justified norms is that these regulate social matters in the common interests of those who may be affected': cf. DE 103); in this way the central content of the universalization principle is introduced illicitly through a side door. In this section I should like to tackle the problem *directly* by questioning the sense in which presuppositions of argument can have a universalistically understood moral content. My answer will be that this would only be possible if we presuppose a (strong) consensus theory of truth, something which I have already shown to be false.

I proceed on the understanding that Apel and Habermas are correct in their grounding of inescapable presuppositions of argument, and thus that anyone who tries to contest the validity of these presuppositions involves himself in a performative contradiction.⁸⁵ Being involved in a discourse, I cannot consistently contest the proposition that I am obliged to be sincere towards those with whom I am arguing, that only the better argument should prevail, or that none of the participants may be prevented from advancing their own arguments. But I maintain that the general norms for arguing, which we are dealing with here, do not amount to universalistic moral norms or meta-norms for

morality. I wish to justify this thesis in two steps:

1 It is self-evident that the norms for argument which we are discussing here cannot govern the initiating or breaking-off of arguments. But if these norms leave me free to decide whether or not to become involved in particular arguments, and whether or not to break off a dialogue, then it is *prima facie* not plausible to understand them as having any moral content. Apel and Habermas believe it is possible to circumvent this difficulty by pointing to the general orientation of speech or even, in Apel's case, of solitary thought towards validity. We might say that if I have genuinely understood this orientation of speech and thought, then I have also understood that I may not suppress arguments – least of all those which speak against me – quite regardless of who voices them. In a certain sense this is of course correct: we call people irrational if they prove impervious to arguments or experiences which would shatter the opinions they hold, and thus if they 'suppress' arguments or experiences, not because the arguments are in reality bad ones or the experiences irrelevant, but merely as a *defensive reaction*. The concept of a good argument, however, implies that we disregard the question of who is voicing it. It appears, then, that what is demonstrated here, at least as far as controversial validity-claims is concerned, is a kind of implied obligation to *proceed* from speech, action and thought to argument, precisely as if it would be *irrational* in a fundamental sense for us not to involve ourselves in a discourse with any being capable of speech and action whenever they required us to do so. I believe that this is more or less the fundamental intuition which provides the bridge, so to speak, in Apel and Habermas from the presuppositions of argument to universalistic morality. But it is a bridge that will not hold up. The requirement not to suppress any argument, which we have recognized as fundamental to any rational approach to one's own opinions, is by no means identical in meaning with the requirement that we should not refuse to take part in arguments with others – whoever they may be. Such a refusal would only be irrational if, for instance, we refuse because we are afraid of the arguments of others. It may be that such a refusal could be *immoral* because we are denying to someone else a right which, other things being equal, we might very well claim for ourselves. But this moral dimension of arguing cannot be explained in

terms of the fact that speech is orientated towards validity, together with the presuppositions of argument – although it can be explained by means of a universalization principle in Kant's sense. I maintain, therefore, that the obligation not to suppress any argument, which is grounded in the validity-orientation of speech, has no *direct* consequences for the question of when and with whom I am obliged to argue, nor of what I am to argue about.⁸⁶ It is only the assumption of a consensus theory of truth that makes it appear that there is any such consequence, for the precise reason that, in the nature of this assumption, the achievement of consensus by means of argument is *defined* as the basic form of a rational approach towards one's own validity-claims.

2 These deliberations might lead us to suspect that the unavoidable presuppositions of argument do not in themselves constitute *moral* obligations. Let me emphasize that I am not disputing that the practice of arguing is, so to speak, imbued with moral obligations. But this might be explained by the fact that a maxim of refusing to take part in dialogue is not generalizable. What is questionable, however, is whether those norms of argument which we cannot dispute without committing a performative contradiction actually betoken obligations of a moral nature. To put it another way, it is questionable whether the 'must' entailed in the norms of argument can be meaningfully understood as a *moral* 'must'. The moral 'must' certainly comes into play at the 'edges' of argument, at those points where a dialogue is being initiated or continued or declined. But if the norms of argument say nothing about whether I should permit the discussion partner to whom I must accord equal rights of speech actually to *exercise* these rights in the very next instant, as it were, then the 'must' entailed in the norms of argument can scarcely be interpreted as a morally substantial one. We seem here to be dealing rather with a 'must' of the kind that is associated with *constitutive* rules: I am unable to dispute this 'must' in my capacity as participant in an argument because it is constitutive of the practice of arguing.

Now norms of argument are of course not rules of a game in which we can participate or not participate as the mood takes us. They are *inherently* connected with norms of rationality such as that which says that we may not suppress any argument that is relevant to the point for which we are claiming validity,

and in our capacity as speaking and arguing beings we are unable to escape such norms (this much in the intuitions of Apel and Habermas is correct). But precisely the fact that the inescapability of obligations to rationality can be expressed through a 'principle of avoiding performative contradictions' also shows that the most general norms of rationality are not directly capable of having a moral content. Obligations to rationality refer to the acknowledgement of arguments, moral obligations to the acknowledgement of persons. It is a requirement of rationality to acknowledge even the arguments of my enemy if they are good ones; it is a requirement of morality to permit even those people to speak who are not yet capable of arguing well. Overstating the point a little, we might say that obligations to rationality are concerned with arguments regardless of who voices them, whereas moral obligations are concerned with people regardless of their arguments. It is of course undeniable that requirements of rationality and moral obligations are frequently interwoven in complex ways, but only from the imaginary perspective of the 'highest point' of an ideal communication community can it appear as if the two would ultimately coincide.

I should like to clarify my fundamental arguments against the attempt at a fundamental grounding of discourse ethics with reference to the clear and carefully elaborated arguments of Wolfgang Kuhlmann.⁸⁷ Kuhlmann's attempt at a fundamental grounding starts, like that of Apel and Habermas, by considering the 'rules and presuppositions of meaningful argument'⁸⁸, but he then goes on to interpret these as norms of co-operation obtaining within the medium of discourse.⁸⁹ These norms of co-operation are the counterpart of the 'norms of' discourse mentioned by Habermas (cf. DE 99); they oblige us, as Kuhlmann puts it, 'to co-operate as *equal* partners, acknowledging and treating each other in the course of argument as possessing equal rights.' They demand 'that in the course of argument each participant is accorded the same right to contradict, to interrupt, to begin afresh, to demand that the argument continue, to put questions, to insist that points are justified, to adduce fresh viewpoints, etc.'⁹⁰ By way of these norms of co-operation, which are implicit in the presuppositions of argument, Kuhlmann attempts to derive a fundamental norm for communicative ethics. He formulates it as follows: 'Endeavour in all cases where your interests might collide with those of others to achieve a rational practical consen-

sus with them.⁹¹ But Kuhlmann is only able to make the connection between the obligations inherent *within* discourse and a moral principle that *goes beyond* discourse by neutralizing the difference between solitary reflection and genuine discourse from the outset. He uses the word 'argument' in such a way that it includes solitary reflection in so far as this is itself orientated towards the assertion of validity.⁹² Because he interprets solitary reflection from the viewpoint of genuine argument in this way, he feels entitled to *equate* the endeavour to arrive at 'consensual' – in the sense of 'true' – solutions with the endeavour to bring about a rational consensus at crucial stages in his exposition. Thus in fundamental norm N2, for example, which is intended to express 'the impossibility of going behind the will to a rational consensus', he says: 'If we are seriously interested in the solution of a problem, then we must endeavour to find a solution to which all can agree, i.e. a rational consensus.'⁹³ In elucidation of this, Kuhlmann writes:

The thing that we want if we really want to know something, if we really want to have the solution to a problem, is a solution *in favour of* which all good reasons can be adduced, and *against* which no justified objection is made or can be made, a solution to which everyone could justifiably *agree*. What we want is a rational consensus.⁹⁴

If the will to truth is identical in meaning with the will to achieving rational consensuses, then universal norms of a genuine co-operation on equal terms with *all* others are indeed built into the validity-orientation of speech from the outset. If this were the case, then the fundamental norm of communicative ethics would be nothing more than a specification of the most general obligations to rationality defined for the special case of conflicts of interest.

My objections are not actually directed at the interpretation of solitary reflection as a virtual dialogue. On the contrary, if in our reflection we take various points of view into consideration, raise objections to our own arguments etc., then this process can hardly be understood other than as an internal dialogue. In the same way, we could understand the endeavour to achieve 'correct' solutions as an endeavour to arrive at an agreement with ourselves which might substitute for an agreement achieved in public dialogue, the voices of others being allowed to speak on

the stage of the solitary subject. It is for this reason that genuine 'public' dialogues invariably also have the function of a test which is to show whether we have really represented the *possible* arguments, viewpoints or objections of others accurately in our solitary reflections. But then these others whose voices are heard in our solitary reflections are always 'representative' others; their claim to be heard is the claim that their arguments make on our consideration. The obligation to enter into genuine discourse can therefore only stretch as far as the obligation not to suppress any relevant argument, or not to evade possible objections. But this obligation is not *identical in meaning* with the obligation to achieve a genuine and universal rational consensus, and it is therefore also not identical in meaning with universalistically conceived obligations to co-operation. It is rather the case that the requirement not to suppress any argument leaves open the question of which persons I am obliged to argue with, and what about, and on what occasion; it therefore also leaves open the question of the specific instances in which I am obliged to pursue a genuine consensus. It is only if we *presuppose* a strong version of consensus theory as providing a criterion of truth that we can interpret elementary obligations to rationality *directly* as an obligation to direct our efforts towards the achievement of a rational consensus on controversial issues. If we do not make this presupposition, on the other hand, then general obligations to rationality, or even general presuppositions of argument, are manifestly too weak to serve as vehicles for a universalistic moral principle *on their own*.⁹⁵

These last considerations suggest a new possibility for interpreting the notion of an 'ideal communication community'. We might say that the real communication community is present as an ideal within the process of solitary reflection, which is to say that it is present in the form of all the possible arguments that might be articulated by the members of an unlimited communication community. But this virtually present communication community is 'ideal' in a double sense: it is ideal *firstly* because it is only present in the form of possible *arguments* which *might* be articulated by real persons and from a multiplicity of perspectives; thus it is present as a community of arguers in which the only form of compulsion that obtains is that of superior argument. And *secondly* this community is ideal because the unlimited communication community we are referring to is envisaged ide-

ally as assembled in one place and at one time. Now in this sense it is easy to concede that the assumption of an ideal communication community is constitutive even of real situations in which arguments take place. This idealization genuinely does elucidate a precondition for what we call 'rational argument' or 'rational reflection'. It is an idealizing abstraction of the empirical persons who articulate arguments – considering arguments *qua* arguments entails thinking of them as detached, as it were, from the persons who voiced them or who might voice them. When it is understood in this way, the assumption of an ideal communication community may well be *necessary*; but we should be deceiving ourselves about the *possible sense* of this assumption if we were to understand it as the anticipation of an *ideal state of affairs* to be realized by a real communication community, just as we are deceiving ourselves about the sense of the assumption of intersubjectively shared meanings if we understand them as the anticipation of an ultimate, an ideal language. That is to say, we deceive ourselves about the *sense* of the *necessity* of these assumptions if we hypostatize them as ideals attainable in reality, even if, as I indicated earlier, the roots of this deception are perhaps to be sought in language itself. It is of course possible to understand the presence of the ideal communication community within the real one as an expression of the inescapable validity-orientation of human speech, as Apel would have us do, but the stuff of which this ideal is made is not suited to designing an ideal life-form. The idealizing assumptions of argument contain neither an ultimate foundation for morality nor the promissory glimpse of an ultimate reconciliation.⁹⁶

III Towards a Mediation between Kantian and Discourse Ethics

X

In section VI, above, I tried to show that Habermas's formulation of the universalization principle contains an unfortunate conflation of a universalistic moral principle with a procedural legitimation principle. This conflation is grounded in a consensus theory of truth which cannot be defended as a substantial theory in the sense that Apel and Habermas intend it. But as far as the conflation of moral principle and legitimacy principle in

Habermas is concerned, I have so far presumed a distinction between questions of moral rightness and questions of the justice of norms rather than elucidating the matter. For the sake of simplicity I shall begin here by elucidating the distinction between moral and legal *norms*. As far as so-called moral norms are concerned, the thing we always have to bear in mind is that these are either unspecific like the moral principle itself ('Human dignity is inviolate') or like ethical obligations of a 'broadly' binding nature ('help the needy'), or else they are so constructed as to admit of exceptions.

This last point is connected with the fact that moral judgements are primarily concerned with ways of acting in particular situations (in Kant's terms, with maxims); only in a derived, albeit *psychologically* important sense are they concerned with universal *norms*. If we may take this reservation as read, I should now like to proceed to a discussion of three characteristic distinctions between moral and legal norms.

(1) Legal norms are distinguished from moral norms firstly by the fact that they are brought into force or set aside, and that they hold, if they are in force, for a particular group of people who are affected by them. Legal obligations are a function of the legal norms that are currently in force. Moral norms and moral obligations, on the other hand, hold, if they hold at all, independently of any act that might bring them into force. The essential distinction here can be easily illustrated with reference to article 1.1 of the Basic Law of the Federal Republic of Germany. The principle of the inviolacy of human dignity holds, as a moral command, independently of the fact that it is enshrined in our constitution. The point of incorporating this moral command into our constitution as a legal norm was, of course, to bind the legislative and the judiciary powers, in the light of the experiences of German history, by means of a corresponding *legal* obligation. – Naturally, the analytical distinction between moral and legal norms is not applicable to the concrete ethical life of traditional societies. However, the transition to a post-conventional morality also means a *conventionalization* of law: to a certain extent the validity of laws becomes a matter of decision, subject of course to moral limitations. Some of these moral limitations have, with good reason, been incorporated *as* legal norms into the constitution of the Federal Republic.

The concept of moral obligation is connected with the grounding of a normative validity-claim; the concept of legal obligation is connected with the concept of the *social* (and thus in a certain sense factual) validity of a norm. Even if the validity of laws is scarcely conceivable without a moment of (social) recognition, it is never entirely subsumed by such recognition. It possesses a moment of pure facticity, if only in the form of a common resolution voluntarily made. It is only because moral and legal validity are not analytically identical that we can even raise the question of to what extent we are morally obliged to follow legal norms that are factually valid (in force). And even if we presuppose a morally grounded distinction between just and unjust norms, it nevertheless remains *meaningful* to ask to what extent we are morally obliged to respect even unjust norms, or whether under certain circumstances we might have the moral right or even the moral duty to offend against a just norm. It would be sheer nonsense, on the other hand, if someone were to ask whether we are morally obliged to follow moral norms which are not valid.

(2) Legal norms are secondly distinguished from moral norms by the fact that they are as a rule *constitutive* of a particular praxis: it is impossible to conceive of legal systems without a large proportion of constitutive rules. Legal norms are constitutive rules in so far as they do not simply lay down rights and duties, powers and sanctions, but also 'constitute' practices (such as 'parliamentary elections'), institutions (such as 'parliament' and, in the case of the Federal Republic, the 'Constitutional Court'), or offices (such as the 'Federal Chancellor'). Parliamentary elections, governmental decisions, proclamations of law, and even taxes owed, would not exist without the law which, whatever else it is, is *also* a system of mutually supporting definitions and constitutive rules. It is of course possible for such a system of institutions and practices to have evolved, as is the case in Britain, by a natural process, as it were, which is to say, historically; equally it is possible for new institutions and practices to develop spontaneously in revolutionary situations, in the form of a system of soviets, for example. But the constitutive character of rules is not crucially affected by the question of whether they are explicitly codified or only grounded in a general acceptance. As in the case of games, a praxis may persist without the written

codification of the constitutive rules for it (e.g. what counts as a 'goal', or as 'check', or what is a correct move in chess etc.). It is sufficient if in cases of disagreement such codifications can be effected as the need arises, whether they are intended to apply once and for all, or on an ad hoc basis.

It is inherent within the constitutive aspect of legal norms that they manifest themselves in the form of *systems*, which again makes them comparable with the rules of games: it is not possible to threaten imprisonment as the punishment for manslaughter without laying down what is to count as manslaughter, without laying down the rules of judicial procedure, and without rules for carrying out the sentence. It is not possible to lay down voting procedures in parliament without also determining what things parliament may decide, how it is to be elected, who is to watch over the implementation of laws that have been passed, etc. Moral norms do not have this systematic character for the reason that they concern the question of correct action in a *given* world, whereby what is given in reality includes legal norms amongst other things. From a moral standpoint, the social validity of legal norms is to start with one fact amongst others, so that, for example, I know that my vote does not 'count' if I fill out the voting slip incorrectly, or that I must expect to be punished if I break the traffic regulations or the tax laws. This brings me to my third point, the problem of sanctions.

(3) Legal norms are as a rule associated with the threat of external sanctions. In so far as constitutive rules are concerned, the sanctions consist quite simply in the fact that the non-observance of the rules renders the corresponding action legally invalid or ineffective. Voting or judicial verdicts, for instance, are invalid if the rules of procedure have been breached – just as a goal is no goal if it was scored from an offside position. In other cases the sanctions consist in legally established punishments such as imprisonment, fines, loss of civil rights, etc. We might even say that basic moral norms such as '*Neminem laede*', 'Thou shalt not kill', 'Thou shalt not lie', etc. find their way into law in the form of penalties: anyone who does such and such will be punished by imprisonment for not less than . . . years. The penal code links particular states of affairs with punitive measures; it is, so to speak, the point of penal law that it introduces a graded system of sanctions for actions whose moral reprehensibility is as a rule (and not always with justice) simply presupposed. The necessity

of an analytical distinction between moral norms and corresponding legal norms becomes clear particularly when we bear in mind that the question of whether an action is morally reprehensible is quite different from the question of whether it ought to be punished. I might consider it morally repugnant for someone to claim that Auschwitz was all lies and yet be opposed to imposing a general punitive threat on such action.

In contrast to legal norms, moral norms are not linked in any *essential* sense with external sanctions; morally good action is not enforceable, by contrast with legally correct action. In the case of morality the essential sanctions are *internal* in nature:⁹⁷ feelings of guilt, remorse, self-reproach, self-contempt. Moral 'ought' can therefore not have the same sense as legal 'ought' or 'should'. The respective sense of 'ought' or 'should' cannot be independent of the answer to the question, 'And what happens if I do not do what I ought to do?' In the case of moral 'ought', the answer can only be of the kind, 'I shall not be at one with myself, shall not be able to look myself in the face.' In the case of legal 'ought', the typical answer consists in the threat of an external sanction.

The moment of facticity in the concept of the validity of laws, to which I drew attention above under point (1), naturally has something to do with the graded system of external sanctions which the law *also* represents, amongst other things. Legal systems cannot, of course, survive in the long term through sheer coercive authority: the social validity of the law implies, in addition to everything else, that at least a significant part of the legal system is acknowledged by those affected as legitimate ('just') and therefore as associated with moral obligations. But 'the law in force' does not *mean* the same as 'law that is acknowledged as valid (just)'.⁹⁸ It is rather the case that in the concept of the validity of laws, the moments of acknowledgement and enforceability are interconnected in a complex fashion. The moment of a facticity that is associated with the expectation of external sanctions cannot be eliminated from the concept of the validity of laws, any more than can the moment of acknowledgement. If this were not so, then the question of whether and when and to what extent I am *morally* obliged to obey (or to apply) the existing laws would have no sense whatsoever. When the legitimacy of the law is linked with the idea of the freely given approval of all those affected, of course, (and thus ultimately with democratic procedures) then a state of law becomes *conceivable*

in which physical sanctions would no longer be necessary because the form in which conflicts took place would be non-violent. A society without prisons is conceivable. But it seems to me uncertain whether it would be meaningful to assume the possibility of a legal system without external sanctions, for the 'externalization' of morality in the form of positive law and its external sanctions also contains an element of emancipation from internalized normative compulsion.

I have tried to clarify the distinction between morality and law with reference to three characteristic aspects of law. I should now like to elucidate more precisely why and in what sense a universalistic moral principle should be distinguished from a democratic principle of legitimacy. In *either* case the way of distinguishing between 'right' and 'wrong' involves an appeal to the idea of a common will formed in the absence of coercion, whether it is the will of rational beings or of those affected. But the nature of this appeal to a common will should be understood differently in each case. In the case of moral judgement we are concerned with finding a way of acting in concrete situations which we should be able, in B. Gert's terminology, to 'publicly advocate' as a generalizable one. I shall show below what function argument fulfils in this connection. The question that arises in each specific instance is whether we – as rational beings – are able to will that a particular way of acting should become universal. And only the *negative* answer to this question constitutes a moral 'ought'. Norms therefore fulfil a derived function in morality, important as they may be from a psychological point of view.

In law, by contrast, we really are dealing with norms and rules. As I pointed out earlier, the 'de-conventionalization' of morality in the course of a transition to post-traditional society has also meant the *conventionalization* of law. With law and morality developing in contrary directions like this, the law has also become subject to the demands of morality: morality becomes an authority existing beyond and 'above' the law. Connected with this process is the development of a procedural, a democratic concept of legitimacy, according to which a legal system is legitimate if it can be understood as an expression of the common will of those subjected to it. The modern tradition of natural law, right down to Kant, has tried to develop a corresponding concept of the legitimacy of law. But the appeal to a common will of those

affected by a legal system means something structurally different here from what is implied in the case of morality. For here we are dealing with the *positive* common will of those affected to subject their lives to certain rules – which necessarily also means subjecting them to these rules and not to others – and to the sanctions associated with them. The common will is to be conceived here in action, as it were, in the form of a resolution or formal agreement; the act of bringing-into-force or setting-aside is analytically a part of positive law. This concept of the legitimacy of law also has a *counterfactual* application; it is in this sense, for example, that Kant says that the legislator may only make laws which the people would have been *capable* of imposing on themselves. Admittedly, it lies within the logic of the modern concept of legitimacy that the common nature of any decision-making process must as far as possible be realized *in actual fact* – that is, in so far as all those affected are ultimately to be accorded an equal right to participate in the collective processes by which the common will is formed: this is the idea of democracy. But if legitimate laws are to be such that all those affected would have been capable of passing them collectively, and if all those affected are – in principle – to have an equal right to participate in the collective decision-making process, then it goes without saying that the settling of normative questions by means of public argument must play a central part in any attempt to realize the possibility of legitimate law in the sense of the modern concept of legitimacy and to ensure that the law is acknowledged as legitimate. To argue in favour of a legal norm – or a system of legal norms – means in this case the attempt to provide reasons which convince all other affected persons why all people of good will and discernment should necessarily be able to deem it to be equally in the interests of all that this norm or these norms should prevail in society. When it comes down to it, Habermas, as we have seen, makes this *particular* case of the connection between normative validity and real argument into a model case for normative validity in general. But in this way his universalization principle falls behind the differentiation of moral and legal questions which is already clearly articulated in Kant (even if Kant does not clarify the issue satisfactorily). In particular this means that Habermas, because in structural terms he starts out on the level of the justice of norms, necessarily fails to come to grips with the problem of *moral* validity. It is no mere coincidence

– on the contrary, it lies in the very nature of the subject – that the contract theorists of modern natural law from Hobbes to Kant have treated questions of *moral* validity either in the context of their *preliminary* discussion or else as the very *foundation* of questions of the legitimacy of laws. The aim of discourse ethics to ‘sublate’ Kant’s ethics and modern natural law by binding the law to a universalistic morality – *against* the trend of the moral counter-enlightenment – is an entirely justified one; but it can only be realized if we do not fall behind the degree of differentiation that has been achieved in earlier discussions of the problems.

XI

The thoughts expressed in preceding sections already contain the most important elements of a fallibilistic reconstruction of discourse ethics. The task now is to assemble these elements into a whole. I shall do this in an indirect way, by showing in what way the discursive clarification of moral validity-claims can be put into effect within the quasi-Kantian perspective which I have so far adopted. I call this perspective ‘quasi-Kantian’ because I have tried from the outset to separate the fruitful basic idea that Kant has from the formalistic husk in which he has concealed it. This selective reading of Kant rests on a critique of him which is entirely analogous with my critique of discourse ethics. In both cases my criticism is aimed at a philosophical architectonics which depends on an *ideal* as its keystone – the kingdom of ends in Kant’s case, an ideal situation for communication in the case of Apel and Habermas. But just as the arch and the keystone can only maintain their position by virtue of mutual interdependence, so it is also the case here that the criticism of idealizing concepts necessarily also has implications for the overall construction. What this means in Kant’s case I have only indicated so far, and not yet fully explained in its implications. My thesis is that the formalism and rigorism of Kant’s ethics are directly connected with the attempt to ground ethics *sub specie aeternitatis*, i.e. from the point of view of a kingdom of ends. Kant’s moral norms are maxims of action for the members of a kingdom of ends. That is why there can be no exceptions, disagreements, insoluble conflicts or irresolvable issues for Kant, and it is also why the faculty

of judgement cannot play any significant part in Kant's ethics. For the kingdom of ends, the 'form of universality' is sufficient, and this is something that tolerates no muddying of the picture. But then it is only with the problem of *mediating* between the particular and the universal that the real problems of morality begin; to this extent at least, Hegel was correct. Now it is true that discourse ethics addresses precisely this problem, but it is unable to solve it because in one central respect it clings to Kantian architectonics: discourse ethics, too, describes morality *sub specie aeternitatis*.

The interpretation I should like to set against one that is guided by the idea of the perfection of moral sense, is one that rests on the idea of eliminating nonsense. My thesis is that the elimination of nonsense is conceivable even if we do not relate it to the idea of perfected sense, of ultimate reconciliation, of a final truth. I believe, moreover, that Kant's basic idea can in this sense be interpreted fallibilistically (and at the same time dialogically).

I have chosen an indirect way, as I say, showing the precise significance that arguments and ways of arguing have, in the context of forming moral judgements and of moral learning processes, if they are understood in a 'Kantian' sense as I have explained it above. As soon as it is made clear in what sense argumentative and communicative clarifications of moral questions are *possible*, incidentally, it will not be difficult to ground a norm for dialogue in Kantian terms. For in so far as dialogic clarifications are *possible* at all, and perhaps also *important* for those affected, it is easy to see that a maxim of *refusing* to take part in dialogue is not generalizable. The fact that a corresponding *norm for dialogue* must largely remain unspecific and can, so to speak, only assume a specific content in the context of specific interpretations of situations (which themselves remain subject to revision) is something that I consider rather to be an advantage by comparison with the quasi-transcendental norms for argument offered by discourse ethics, which of necessity promise more than they can deliver.

I shall argue here initially on the basis of a simplifying assumption, which I shall then retract in the second stage of my argument. My assumption is that the logic of moral arguments is already determined by a universalistically understood moral principle. This should not be understood in the sense of an empirical assumption about all members of our society, but in the

(methodical) sense of bracketing out arguments and convictions in which competing sources of normative validity such as the will of God, the natural order of things, or the authority of tradition are presupposed. We are thus limiting ourselves to moral arguments in which the generalizability of ways of acting is presupposed as a criterion of what is morally correct or as a measure of moral value. My thesis is that, given this presupposition, moral argument is concerned almost exclusively with the interpretation of the situations attendant upon actions and needs, as well as with the way that those who act or suffer the consequences of actions see themselves – with the result that if we have reached agreement about the interpretations of situations and the way people see themselves, moral controversies will as a rule dissolve. This means that the question whether *we* are (rationally) able to will that *my* maxim should become a universal law becomes more or less identical in meaning with the question whether *my* understanding of situations, the way I see *myself*, *my* interpretations, are appropriate, accurate or truthful. The 'we' that so disquiets discourse ethics resides, so to speak, in the validity of my descriptions of situations, my understanding of reality, and the way I see myself. For the same reason, this is also the point at which to begin any critique or argumentative clarification.

This thesis could be elucidated by means of examples on two distinct *levels*: *firstly* that of the collective matrix of interpretation, *secondly* that of moral judgement in complex situations. As far as the level of the collective matrix of interpretation is concerned, relevant examples could be drawn from changes in traditional views of homosexuality, the role of women, education, abortion or the rights of children. Of course, the proponents of a universalistic morality (and it is with them that we are concerned here) never did believe that morality stops when it comes to homosexuals, women or children. They believed rather that homosexuality was corrupting, that women were not capable of rational self-determination, or that children must learn obedience above all in order to become decent human beings. In proportion as such views become questionable, which is to say that it is no longer possible to advance good reasons in their defence, so, too, do the *moral* views change that are associated with them. Physical chastisement of children becomes morally questionable if we recognize that it represents a senseless injury and not a necessary

pedagogic measure. Legal prosecution and social discrimination against homosexuals becomes morally questionable if we recognize that the condemnation of homosexuals is unfounded. It becomes morally questionable to prevent women from realizing their personal potential if we recognize that the traditional views on the nature of women are untenable. In other words, socially prevailing moral orientations determining behaviour, for example, towards homosexuals, women or children, are anchored in collective matrices of interpretation; collective moral learning processes take place when reasons are brought forward which call such matrices of interpretation into question and prompt their revision, although it should immediately be added that such revision does not as a rule take place only in the medium of argument, but under the *pressure* of a struggle for recognition and under the *influence* of new experiences. Such learning processes result in a new way of talking about and behaving towards homosexuals, women and children (keeping to the examples we have already mentioned). They also result in those affected seeing themselves and behaving towards themselves in a new way. But from a *moral* viewpoint, what we are dealing with here is the *elimination* of inequalities and unequal ways of treating people which have had the floor pulled from under them, as it were, once the dogmatism of traditional views has been shown to be without foundation. Looked at in this way, collective moral learning processes consist in the extension of relationships of mutual recognition through the critical undermining of socially inherited attitudes and matrices of interpretation. What is involved here are *specific negations* rather than advances towards some ideal, as can be seen from the fact that the false or ideological inequalities of treatment conform, as it were, to an archetypal image – by which I mean those cases of *grounded* inequality in which human beings are not *yet*, or no *longer* accorded an equal possibility of *factual* self-determination. Small children, cases of severe mental handicap, and criminals are three examples. I do not want to be misunderstood: precisely children, the mentally handicapped and criminals are also examples of the idea of self-determination remaining in force well beyond the boundaries of traditional interpretations of it. But this only means that the demand that we behave towards every human being in the light of their *possible* self-determination must become more radical in meaning in proportion as false views about the socialization of

children or the nature of psychic illnesses or the causes of crime are dissipated. Kant's insight that freedom can only be learned by training for it has acquired a whole new field of application today, for example, in the sphere of democratic psychiatry. But we are not even able to conceive an ideal ultimate goal for such changes: the principle of moral progress is not the perfection of sense, but the elimination of nonsense.

The second level on which I should like to illustrate my thesis is that of moral judgement in complex situations. I should like to start by distinguishing between three different forms of morally relevant complexity. I call those situations *morally complex* in which different moral demands, as it were, collide with each other and there is no easy or unambiguous possibility of choosing between them. I call situations *morally inscrutable* if the moral *significance* of actions in those situations is unclear, either because the agents are mistaken about their motives or because the communication situation as a whole is distorted. And finally I call situations *practically inscrutable* if the *consequences* of our actions in those situations are unclear. A situation would be morally *complex* if it prompts the question, 'Should (or may) I (really) help him?' A situation would be morally *inscrutable* if it prompts me to ask, 'Do I really *want* to help him?' A situation would be *practically inscrutable*, finally, if it prompts me to ask, 'Can I help him *in this way*?' All three forms of morally relevant complexity clearly have to be considered if we are asking about the logic of moral arguments concerning the correct way of acting in concrete situations. Now we could, at least for the purposes of crude orientation, classify *morally inscrutable* situations under the validity-dimension of *truthfulness* and *practically inscrutable* situations under the validity-dimension of empirical *truth*. And we could then label the corresponding dimensions of moral discourse, along the lines that Habermas proposes, 'therapeutic' and 'empirical-theoretical' discourse respectively. I use these classifications here merely in order to separate out that dimension of *specifically normative* argumentation which the moral principle of discourse ethics has in mind and which occupies a distinct *category* of normative discourse alongside therapeutic and empirical-theoretical discourse. Beyond this, we must make one further qualification. We have already dealt with *one* important aspect of moral discourse, namely the one concerned with general moral orientations and ultimately with socially prevailing modes

of interpreting reality and needs. What remains as the essential core of moral discourse, after we have made all these qualifications, ought to enable us to make clear the logic of moral argumentation.

It might at first seem improper to limit discussion to morally complex situations. But I am proceeding on the understanding that, as far as elementary moral doctrine is concerned (e.g. cases of wilful lying, injury, killing, or even leaving others in the lurch), this causes no problems for the quasi-Kantian perspective I am propounding here. That is to say, I proceed on the understanding that – rationally – we cannot will that corresponding ways of acting should become universal. That is how we arrive at prima facie norms like ‘*Neminem laede*’ or a prohibition on telling untruths. The problem of morally complex situations concerns the question how, in the case of a *conflict* of norms, for example, we are to understand the justification of *exceptions*. When formulated in this way, however, our statement already contains a misleading element. If norms, when looked at logically, do not have primary status in morality, but a derivative one, then the grounding of a moral judgement in morally complex situations does not mean the grounding of an exception, but ultimately, once again, the grounding of the generalizability – or non-generalizability – of a way of acting. I refer back here to the thoughts contained in section III. There I showed that the so-called exceptional moral situations cannot be made to conform to rules in a strict (Kantian) sense. Let us remind ourselves of this by looking at two examples, choosing the following two maxims for this purpose:

‘If need be I shall try to preserve an unjustly persecuted (or accused) person from arrest (or sentence) by telling a lie,’ and
‘I shall assist a terminally ill patient to die if that is their wish.’

In the case of either maxim it seems clear that, as they stand, they *cannot* be called generalizable. In either case I have no difficulty in imagining situations in which I should consider it disastrous or at least wrong for these maxims to be acted upon. The person I assisted to die might only *believe* themselves to be terminally ill; or I might equally well want to get rid of them before they made a will, and be using this convenient moment etc. And where the innocent victim is concerned, I might be

mistaken about their innocence, it might be that my lie would put another innocent person at risk, etc. Notice that we are not here dealing with the question of whether what I am able to will to be a universal law can also be willed by everyone else as a universal law. Rather it is possible that, if I think the matter through carefully, I am not myself able to will the corresponding maxim to be a universal law, even if I might perhaps consider a corresponding *action* to be correct in a specific situation. It thus transpires that so-called exceptional moral situations, as opposed to morally elementary situations, really cannot be made to conform to rules. If we wanted to formulate corresponding ‘norms of permission’ they would have to take the form, ‘In situations that are sufficiently similar to *this* one, one may . . .’ (or perhaps even ‘one must’).

This is where we encounter once more the curious asymmetry between morally elementary and morally complex situations. In the case of the first example, there is a norm which results from the non-generalizability of ways of acting such as might bring about a wilful injury to others, and that norm runs as follows: ‘*Neminem laede*, unless you have a good reason, one that can be “publicly advocated”.’ But our analysis of the second example shows that reasons which can be publicly advocated may only be formulated in the form of *norms* for exceptions if we think of them *either* as carrying an indexical element *or* as qualified by some unspecific limitation clause such as ‘Under certain circumstances it is morally correct to . . .’ All this applies, as I have said, quite independently of the question of any possible congruence between *my* ‘being able to will’ and that of everyone else. But on the other hand, if my analysis is correct, since the problem is a purely conceptual one, it poses itself in the same way for each individual making a moral judgement.

This brings me to the final step in my reflections. We have seen that the moral judgement that is made in morally complex situations cannot be expressed as a judgement about the generalizability of a maxim (in the strict sense of the term). But this means that a judgement about the generalizability or non-generalizability of ways of acting in particular situations – and I should like to stick to this interpretation of moral judgement – can ultimately only be grounded by means of the analysis of concrete situations. In other words, moral arguments are concerned in this case above all with the *appropriateness* and the

relative *completeness* of descriptions of situations, including the various alternative ways of acting that are available in a given situation. This becomes even clearer if we remind ourselves of the 'negatory origins' of moral 'ought' or 'should' to which I pointed in section II, above. For it follows from these negatory origins of moral 'ought' or 'should' that the primary element in the formation of moral judgements and in moral arguments is not the generalizability, but the *non-generalizability* of ways of acting. Ways of acting are generalizable (permitted, legitimate) if they are not non-generalizable. There is no tautology here because we are dealing with a conceptual or cognitive primacy of negation – a cognitive primacy because it is, as it were, the elementary operation of moral judgement to establish the *non-generalizability* of ways of acting in a given situation. Now it seems to me clear that any judgement to the effect that a way of acting is *non-generalizable* is a function of understanding that way of acting *as* a way of acting in a given situation. Whether I judge the act of handing a fugitive over to the police to be non-generalizable or not depends entirely on whether I understand such a way of acting as an act of co-operation with legitimate state authority or as leaving a helpless and innocent victim of persecution in the lurch (or indeed as an act of complicity with a system of terror). But in a given situation, only one of these two interpretations can be correct. As soon as the question of a correct understanding of the situation is clarified, however, it will as a rule be the case that the question of the generalizability of specific ways of acting is also resolved. In this way we could understand moral judgement as the ability to grasp those aspects of situations upon which the non-generalizability (or the generalizability) of ways of acting depends. Moral discourse, however, would above all be a discourse about the correct way of understanding reality from a moral point of view.

My thesis is, therefore, that as a rule moral controversies are dissolved when agreement is achieved in those various dimensions of moral discourse which I have so far discussed – general interpretations, the way those affected see themselves, descriptions of situations and the understanding of the consequences of actions as well as the alternative ways of acting that can be discerned within a situation. In this sense we might say that the question of whether we are – rationally – able to will that a way of acting should become universal is above all a question about

whether we have adequately understood the concrete situation in which action takes place. This also explains how the question of what we – as rational beings – are collectively able to will is reduced in practice to the question of how we – the persons affected – can achieve an adequate understanding of the situations in which we act. But as far as this question is concerned, a consensus among a few individuals who are sufficiently close to the concrete situation to be able to judge it is often more important for the process of assuring ourselves that a moral decision is correct than any actual general accord on the subject.

The preceding reflections force us to reconsider the distinction we previously made between 'therapeutic', 'empirical-theoretical' and (in the narrow sense) 'normative' aspects of moral discourse. It ought to be clear by now that we cannot bracket questions of truthfulness and empirical truth (in the broadest sense) out of moral discourse without robbing it of its substance. It is clearly not the case that what remains once we have bracketed out such questions is a problem of the grounding of moral norms honed down, so to speak, by analytical means to a fine point. What I have called the 'essential core of moral discourse', i.e. what remains after bracketing out those subsidiary aspects, seems rather to represent that aspect of moral judgement which is *either* self-evident (in the sense in which, according to Kant, it is self-evident that in the light of the categorical imperative I may not tell lies in order to secure my own advantage) *or else* no longer admits of any intersubjectively binding decision. This is less paradoxical than it sounds; we only have to abandon the premise that moral *judgements* can only be grounded with recourse to norms in order to see that moral *arguments* do not have to be normative in character. 'You promised him that you would' is a (simple) moral argument. But the proposition that, other things being equal, promises should be kept, is not really the premise of the conclusion 'Therefore you must do it' – i.e. it is not the kind of 'premise' about which we would begin to argue on a higher level of discussion if the case is controversial. It is much rather the case that this 'premise' really only expresses our understanding of corresponding situations for action in the form of a *prima facie* norm.

I do not, of course, wish to dispute that moral judgements carry an index of normative generality; in *this* sense it is easy to concede that moral arguments are always concerned with the

grounding of norms. But the crucial point is how we understand this connection between the grounding of norms and the evaluation of ways of acting. Habermas understands this connection as a form of derivation: the moral command to act in a particular way is seen as following from the fact that this way of acting corresponds to a valid norm. By contrast, the view that I am propounding here is that the validity of moral norms only stretches as far as the validity of the moral judgements that can be – not grounded, but – expressed through these norms. The norms themselves carry, so to speak, a situational index which binds them to the situations in which they have their origins. This is the only reason why there is a problem in *applying* moral norms – and this is the only sense in which it can be understood. In other words, it is not possible in the case of moral norms to separate discourses concerning the justification of norms and discourses concerning the application of norms categorically from each other. Only if we bear this in mind can we interpret the problem of moral judgement in concrete situations meaningfully as a problem of the ‘application’ of moral norms.

I believe that the interpretation of moral argumentation that I am proposing here tends to strengthen the plausibility of the basic idea of discourse ethics. Moral dogmatism and moral self-deception barricade themselves as a rule behind interpretations of situations (including interpretations of needs and interest) which are kept out of the discussion itself. When reality is warded off like this, there is always a potential danger that some injury may be done to individuals. The command to seek communicative or discursive clarification of interpretations of situations, and of the way we see ourselves, therefore has the status not only of an obligation to be rational, but of a moral norm – at least in so far as it is a matter of letting those affected speak for themselves. Of course, what I said earlier about other moral norms goes for this one, too; for this reason alone, it cannot provide a foundation for everything else.

Excursus Contrary to the position I have expounded here, Habermas insists on a sharp analytical distinction between the problems of justification and the problems of application.⁹⁹ In the distinctions drawn between these two sets of problems, Habermas even sees a new level of differentiation which Kant was the first to attain, and below which ‘we must not allow ourselves

to fall’.¹⁰⁰ I have argued, against this, that Kant systematically *neglects* the problem of application for reasons connected with his rigorism on the point of ‘laws’. What Kant is doing in reality is differentiating the problems of grounding norms *at the expense of* the problems of application. Moreover, Habermas’s differentiation thesis seems to me in itself unclear. As far as the grounding of moral norms is concerned, we have already seen that the norms we are talking about here can only be ‘prima facie’ norms (such as ‘Thou shalt not lie’). But if that is the case, then the problems of application largely coincide with the problems of exceptional situations or situations of conflict (which means much the same as morally complex situations). But if, as I have also shown, morally complex situations cannot be reduced to rules in the same sense as morally elementary situations, and if what is actually being grounded is the generalizability or non-generalizability of ways of acting in situations of a particular sort, then it is no longer possible to separate the problems of grounding from the problems of application in the sense that Habermas means it. Not to put too fine a point on it, we might say that the problem we are dealing with in the process of moral grounding *is* a problem of application; what is being ‘applied’ is the moral principle itself. I have already shown what this means in the case of morally complex situations, but it would also be possible to understand the grounding of general moral orientations in the same way. Returning to the examples we used before, we were looking there at principles like ‘Human dignity is inviolate’ or ‘Every person has an equal right to the free development of his or her personality’ – principles which are, as it were, not very far removed from the ‘ends formula’ of the categorical imperative – and asking what they *mean* in connection with behaviour towards women or children or homosexuals. In contrast to Habermas, then, I am of the opinion that in the case of morality, the problem of grounding has the character of a problem of application; what moral discourse is concerned with is the ‘application’ of the *moral point of view*, whether to concrete social problem areas or to the situations in which individuals act.

If Habermas’s response to this is to say that ‘no norm [contains] the rules of its own application’,¹⁰¹ then this may be true, but it does not justify the separation of the problems of grounding from the problems of application in this case. It seems to me

rather that he is conflating two different problems of application. One of these is the kind of problem that presents itself when prescribed rules, codes of behaviour or norms – such as the norms of penal law – are to be applied to concrete cases, in which case the *grounding* of norms and the *application* of norms are two different things: the grounding of a norm (or at least the ‘promulgation’ of it) precedes its application. But precisely because moral consciousness becomes emancipated from the dogmatism of *substantial* prescribed norms in the process of differentiation between justice and morality and in the transition to a post-conventional sense of morality, the problem of application that emerges for morality is a problem of a *different* kind. This *second* kind of problem of application is concerned with the question of the correct way to bring the ‘moral point of view’ to bear in the case at issue. *This* is the question with which moral discourse is concerned; only in a derived sense is it concerned with the grounding of norms. In an essential sense it is therefore a *discourse of application*. Moral discourse and moral judgement thus do not differ in their *object*; practical reason expresses itself as moral judgement. (This seems to me to be the real point of Hannah Arendt’s reflections, too, in her essay ‘Thinking and Moral Considerations’.¹⁰² But then again, Hannah Arendt only deals with an aspect of the matter which complements Habermas’s approach. Whereas Habermas marginalizes the problem of application by comparison with the problem of grounding, Arendt does not make it clear what moral judgement has to do with the possibility of moral *discourse*.)

In the interview to which I referred earlier, Habermas gives a further peculiar reason for separating the problem of grounding off from the problem of application. He says there that moral theories which follow Kant are ‘typically restricted to the question of the *justification* of norms and actions’, and that they ‘have no answer to the question of how justified norms can be *applied* to specific situations and how moral norms can be *realized*’. But he accounts for this position by saying that ‘one should not place excessive demands on moral theory, but leave something over for social theory, and the major part for the participants themselves – whether it be their moral discourses or their good sense.’¹⁰³ This is a curious way to account for the ‘differentiation thesis’ because the question of ‘leaving the major part for the participants themselves’ is not at issue. Habermas himself argues that the *grounding*

of norms is not part of the business of moral theory, but is a matter for moral discourse among the ‘participants’. What is at issue, then, is not the drawing of correct boundaries for moral theory, but the correct understanding of what should be left for the participants themselves, namely moral discourse.

I have so far proceeded on the assumption that the logic of moral argument is determined by a universalistic moral principle. As I indicated earlier, I wish to drop this assumption in the second stage of my argument. If Kant asserts that the categorical imperative is a universal and inescapable ‘fact of reason’, then this thesis clearly makes little sense if we understand the categorical imperative as a universalistic moral principle in itself. We could, however, also understand it to be saying, in a weaker sense, ‘Act according to your normative convictions,’ which is to say, ‘Make no exceptions for yourself,’ or ‘Do what you (believe that you) ought to do.’ When understood in this way, the categorical imperative *is* a fact of reason, for it is merely formulating an elementary condition of consistency for human action. When we understand it like this, of course, the categorical imperative is compatible with the most diverse systems of norms, which might be particularist or feudal in nature, or based in religion. I believe, however, that even when its meaning is thus restricted the categorical imperative contains no *trivial* requirement – at least, not if we may assume that the tendency to moral self-deception and to making exceptions for oneself is a feature of all known human societies.

The ‘minimal’ interpretation of the categorical imperative which we are considering here rests, of course, on the assumption that a dimension of moral judgement and self-judgement is constitutive of all forms of human community.¹⁰⁴ That is to say that a categorical ‘ought’ is built into the structure of reciprocity that characterizes human social relations, and that the commands of this ‘ought’ can only be violated at the price of moral condemnation and self-condemnation (feelings of guilt). We are unable to withdraw from this dimension of moral judgement *as such*, and this implies that we are unable to withdraw from the conditions of living in mutual recognition of each other. Nevertheless, the fact that a categorical ‘ought’ might (perhaps) obtain universally does not, of course, make it as such a fact of *reason*. It is rather the case that particularist, traditionalist or religious ways of viewing and grounding this categorical ‘ought’ must first

have disintegrated before it is even possible to ask about the possible *rational* sense of it. Now, I think that the Kantian moral principle gives us an answer, if not to the question about the *rational sense* of categorical 'ought', then at least to the question about its *rationalizable core*. The rationalizable core of categorical 'ought' – which is in itself less a fact of reason than a fact of human natural history – is the thing that ought to be done as the negation of what *we* are unable to will as a universal way of acting. *In retrospect* this also holds for traditional societies or even for the particularist moralities of tribes, although moral commands were not understood in this way at the time, but rather as divine commandment, for example, or as the expression of a natural order of things. The rationalizable core of categorical 'ought' is thus anchored in the structure of reciprocity as such. The development of a universalistic morality can then be understood as the successive elimination of the foundations of a particularist understanding of such structures of reciprocity. *In retrospect* we can identify that common feature which constitutes a universal core of morality in all structures of reciprocity – in terms of content it can be expressed in commands like 'Do not lie', 'Do not kill', 'Do no wilful injury', etc. But on the other hand it is only through the discovery that the traditional categorical 'ought' is without foundation that this 'ought' becomes accessible to reason, that it becomes 'rationalizable'. Universalistic morality itself owes its being to an elimination of what is false, without the possibility that it might be able to completely recover its foundations by means of some fundamental grounding. It retains a moment of mere facticity which is connected with the fact that we are not able to become the persons we are, nor to live, outside the structures of reciprocal recognition. But this fact, which is not a fact of reason, but something fundamental to all possibility of reason, can *retrospectively* be included among the preconditions of reason. In *this* sense it is true that reason recovers its own foundations in universalistic morality. But the impossibility of a fundamental grounding of morality in terms of a pragmatic philosophy of language is connected with the fact that the impossibility of leading a morally good life if we cannot look ourselves straight in the eye is something for which we cannot in the final analysis provide a grounding, but which we simply have to accept. It is impossible to conceive of any process of successful individuation which does not involve a confrontation

with others as, in Kantian terms, 'ends in themselves', or which is not, in Hegelian terms, bound into structures of reciprocal recognition. The medium of such mutual recognition is language. It is in language that mutual recognition is represented in the form of normative validity-claims, and as linguistic utterances, such validity-claims are always implicitly orientated towards the possibility of concurrence among all beings capable of speech. This is precisely where the attempt to look for universalistic morality in the foundations of speech finds its legitimation. But it would be inconceivable that we could work away in the medium of language at eliminating the original particularism of forms of mutual recognition if there were not some pre-existing basis for an affectively anchored moral 'ought' connected with the conditions of the possibility of our being ourselves. In the power of this moral 'ought' it is still possible to discern weak traces of a real power which accompanied the process of individuation in the form of a threat. In universalistic morality, this real power is sublated within a state of affairs in which the only form of compulsion that obtains is that of superior argument. But it is sublated only if the mere awareness of a 'necessitation' by the moral law is replaced by an awareness of the price that has to be paid if structures of reciprocity are infringed by a self that owes its existence to the internalization of such structures of reciprocity. Then, and only then, is the categorical 'ought' of moral validity-claims sublated within a practical knowledge of the preconditions of a good life. Moral validity-claims are claims in two senses: they contain an expectation of general concurrence, and they *demand* a certain way of behaving. Kant's categorical 'ought' is the expression of this character of morality as a demand. From Schopenhauer to MacIntyre, the rational sense of this categorical ought has been repeatedly challenged;¹⁰⁵ but for Kant it was quite simply an expression of the state of tension prevailing in finite rational beings *between* reason and the sensual. Only of a 'perfectly good will' would it be possible to say, as Kant does, that it 'could not . . . be conceived as *necessitated* to act in conformity with law,'

since of itself, in accordance with its subjective constitution, it can be determined only by the concept of the good. Hence for the *divine* will, and in general for a *holy* will, there are no imperatives:

'I ought' is here out of place, because 'I will' is already of itself necessarily in harmony with the law.¹⁰⁶

Kant conceives the sublation of 'I ought' within 'I will' as the ultimate goal of a potential moral progress. But since the 'perfectly good will' can really only be imagined in Kantian terms as the will of a totally disembodied subject, and thus not as a will at all, his sublation formula remains aporetic. It could be converted – contrary to Kant's intention – into worldly terms, not by secularizing the idea of a perfectly good will, but by identifying the worldly aspect of the categorical ought itself by which it still belongs *outside* the sphere of reason, namely external compulsion which has become internalized. The sublation of 'I ought' within 'I will' – which would also represent the sublation of the opposition between deontological and teleological ethics – could be conceived as the form of a moral consciousness for which there no longer existed any opposition between self-love and solidarity for others, between self-assertion and the acknowledgement of others. *This* scheme of sublation does not force us to assume a 'perfectly' good will (of which it would not even be possible to say how it would have to be constituted), rather it expresses a potential enlightenment of moral consciousness about itself, the sublation of (mere) virtue in (practical) knowledge.

It is in *this* sense that a universalistic morality is cognitive. But at the same time, a 'lack of moral sense' is not a cognitive deficiency, but rather an expression of the fact that the person concerned has not been adequately trained in reciprocal recognition, and this is something against which mere arguments are powerless. If it were the case, however, that a moral consciousness had already developed, then the development of a *universalistic* moral consciousness is the only alternative, under the conditions of enlightenment, to withdrawing from the language game of morality altogether. To withdraw in this way, and thus to rupture the bonds of solidarity with others rather than extending them, would, moreover, involve the individual in doing injury to himself, in extreme cases it would imply his self-destruction.¹⁰⁷ It was an insight of this kind, I believe, that Kant was expressing when he called the necessitating of the will through a moral law a 'fact of reason'. It would be less misleading to speak of the fact of a life lived under *conditions* of reason. This is a fact of which we can *remind* ourselves and others, but to remind ourselves in

this way is not identical in meaning with demonstrating the inescapability of the obligation to rationality. This act of reminding, which certainly cannot take the form of a fundamental grounding, is perhaps the only possible foundation upon which morality can be grounded.

XII

In section VI, above, I mentioned Habermas's attempt to interpret moral ought as a predicate of a higher order (analogous to the predicate 'is true'). This is how Habermas tries to resolve the problem of moral ought in cognitivist terms by interpreting ought as one of three types of universal validity-claim. As I have tried to show, the difficulties which arise from *this* attempt when the discourse-ethical approach is put into practice are virtually irresolvable. Now, I believe that these difficulties are apparent in the very grammatical reconstruction that Habermas proposes, precisely because he relates it from the outset exclusively to *moral* validity-claims. The point of the reconstruction he proposes is, as we have just noted, to make it possible to explain moral demands as one of exactly three types of universal validity-claim (truth, truthfulness, normative rightness), of which Habermas asserts that they are present in *every* linguistic utterance, direct or indirect. If it were possible to explain moral ought in this way, this would mean that it was so deeply anchored in universal linguistic structures that it would be unnecessary to ask about its possible rational sense. But this is also the reason why so much depends on the successful fundamental grounding of a moral principle, for only such a fundamental grounding could bridge the gap between the *general* grammar of normative validity-claims and the *particular* demands of a universalistic morality. Now it seems to me that we already run into problems if we load the *general* concept of a normative validity-claim from the outset with the particular sense of *moral* validity-claims. The moral use of words like 'ought', 'must', 'may', 'is commanded', 'right', 'good', etc. is a very special one. But since the general (non-moral) use of these words is also linked with validity-claims, it might have seemed an obvious step not to relate the grammatical reconstruction of this basic normative vocabulary from the outset to the special case of moral validity-claims. In

other words, if the reconstruction proposed by Habermas is correct, then it ought to be equally applicable to hypothetical imperatives in Kant's sense, to grammatical prescriptions ('here one must use the infinitive'), and even to aesthetic 'must' statements ('here there must be a sudden *forte*'). In all these cases we are dealing with validity-claims which can be grounded and criticized every bit as much as moral statements, and in a general sense they are certainly also normative validity-claims (in contradistinction to claims to truth or truthfulness). Normative validity-claims in the general sense of the word are distinguishable from moral validity-claims, however, in that they only yield *prima facie* reasons for doing something without expressing an *unconditional* (categorical) obligation in the way that moral validity-claims do. The obligatory character is clearly connected with the kind of reasons that one can put forward, from case to case, for normative validity-claims of a particular type. But this means that the categorical sense of moral ought can only be elucidated in connection with the reasons that one can adduce for moral validity-claims. It is for this reason that, in Kant, categorical ought appears in the moral principle itself. But since the concept that appears in Habermas's principle (U) is only that of a 'valid' (just) norm, we should have to understand the connection between his grammatical reconstruction of 'ought' statements and his formulation of principle (U) as meaning that the words 'right' or 'commanded' only assume the sense of 'morally right' or 'morally commanded' in connection with principle (U). What he is saying, therefore, is something like this: 'In circumstances S, it is morally (*unconditionally*) commanded (right) to do p if p corresponds to a valid norm.' But here we can see that the grammatical reconstruction of normative validity-claims has not disposed of the problem of moral ought; rather it remains a *particular* problem which consists in the fact that there is an *unconditional* command in this particular case to do, not what is right, but what is in a *certain* sense right, so that there can be no possibility in this case of finding good reasons for refusing to do what is in *one* sense right by appealing to alternative criteria of rightness. (I leave aside the question of whether this is a totally adequate picture of the priority of moral validity-claims over other normative validity-claims; at least it is a picture that Habermas shares with Kant.) Because the priority of moral validity-claims over other normative validity-claims is not made clear in Habermas's recon-

struction, I think that his use of the insights of linguistic pragmatics has not in reality neutralized the problem of moral ought, but merely off-loaded it, so to speak, into a grey area which lies between his grammatical reconstruction of normative validity-claims and his formulation of principle (U).

The most that could be said against this is that moral ought is so deeply anchored in the way that communicative action is orientated towards the pursuit of validity that even the truth-claim of assertoric utterances could be elucidated in categories of a claim to rightness analogous to that of moral ones. If we understand 'truth' as 'warranted assertibility', then we are interpreting the act of assertion itself as an act of both claiming a right and entering into an obligation: the claim that is raised in the assertoric utterance is one that could be redeemed by means of argument, and if I assert something, then I commit myself to redeeming such a claim by means of argument, as the need arises. We might be tempted, therefore, to invert the priority of propositional truth over other modes of validity that is characteristic of the philosophical tradition, and give priority to normative rightness instead. If it were possible to ground a primacy of practical reason in this sense, then the problem of moral ought would dissolve into nothingness, being, as it were, an expression of the 'logocentric' prejudices of Western thought.¹⁰⁸ The acknowledgement of moral or analogous obligations would turn out to be a precondition of the possibility of participating in any kind of communicatively orientated action, and thus also a precondition of the possibility of assertoric utterances.

The consensus theory of truth is indeed itself an expression of such a radical inversion of traditional priorities. Habermas has also tried to justify this inversion 'genealogically' in his *Theory of Communicative Action*. Following Durkheim, he sees the validity of moral ought as having origins in a pre-rational, symbolically structured sacred sphere – as it were, the sphere of a primordial normative consensus that is not yet conscious of itself.¹⁰⁹ In this way, a still pre-rational understanding of the validity of norms is constituted, which performs a crucial mediating function in the grammatical differentiation of human speech. This process of differentiation takes the form of a 'linguistic-tification of the sacred'.¹¹⁰ The reason why the sphere of the sacred is able to take on this mediating function is that it occupies a special position among the 'three roots of communicative

action'.¹¹¹ These three roots are the pre-linguistic roots of linguistically articulated cognitions, obligations and expressions,¹¹² which should be seen as the basic building-bricks, so to speak, of every grammatically differentiated linguistic utterance. Grammatically differentiated speech is nothing other than the integration of these moments into a whole made up of propositional, illocutionary and expressive components.¹¹³ Now, only the propositional and the expressive components of speech can be derived from a pre-linguistic order that is not itself already symbolically structured. The pre-linguistic correlate of the propositional component of linguistic utterances is to be found in perceptions, mental images, and adaptive behaviour; that of the expressive component in bodily gestures.¹¹⁴ The illocutionary components of speech, on the other hand, which are what really lend assertoric and expressive statements their power 'to motivate a hearer to accept a speech-act offer',¹¹⁵ point towards a pre-linguistic root of a different kind, namely that sphere of the sacred which, though pre-linguistic, is nevertheless not natural, but already symbolically structured.

What is puzzling about this root is that it is from the very beginning symbolic in nature. Cognitive dealings with perceptible and manipulable objects, and expressions of subjective experiences, are in contact with external or internal nature through stimulation of our senses or through our needs and desires. They are in touch with a reality that not only transcends language but is also free of symbolic structures. Human cognitions and expressions, however shaped by language they may be, can also be traced back to the natural history of intelligent performances and expressive gestures in animals. Norm consciousness, on the other hand, has no equally trivial extralinguistic reference; for obligations there are no unambiguous natural-historical correlates, as there are for sense impressions and needs. Nevertheless, collective consciousness, the paleo-symbolically supported normative consensus, and the collective identity supported by it secure for experiences of obligation contact with a reality that is, if not free of symbols, at least prelinguistic [in the strict sense of propositionally differentiated language] – they are 'older' than interaction mediated by grammatical speech.¹¹⁶

The 'binding effects' of the illocutionary component of linguistic utterances are due to the fact that its pre-linguistic root is itself already a symbolically, or rather 'paleo-symbolically' structured

normative consensus. If this were the end of the matter, then, as Habermas states, 'constative and expressive speech actions could not achieve binding effects on their own but only in virtue of their normative content. The illocutionary component of such speech acts would then have no motivating force; the burden of coordinating action would have to be borne instead by the prior consensus supporting the normative context.'¹¹⁷ Now, at this point Habermas reasserts his point about the parallel between

(1) 'It is right that a in S'

and

(2) 'It is the case (is true) that p',¹¹⁸

and expresses the following assumption: let us assume that the assertion of truth-claims with the help of constative utterances of type (2) only became possible by virtue of the fact that an already available concept of norm-validity migrated, as it were, into the illocutionary mode of assertion, and that it did so in such a way that a type of claiming of right was thus constituted in which, by contrast with the case of genuine norm-validity, the grounding of such claims coincided from the outset with their redemption.

It may be the case that the claim to propositional truth originally borrowed the structure of a validity claim that can be *justifiably* redeemed from the kind of claim that rests on valid norms, but it had at once to appear in a radicalized version geared to the giving of reasons in its support. This suggests that the concept of a criticizable validity claim derives from an assimilation of the truth of statements to the validity of norms (which was, to begin with, not criticizable).¹¹⁹

As Habermas explains, following Durkheim, the concept of norm-validity is already linked with the assumption of an idealized agreement among all members of a society, which is how the primordial, symbolically structured normative consensus can become the point of departure for a grammatical differentiation of speech, which is to say that it becomes the model for *all* concepts of validity, and for the concept of truth-validity in particular.

The normative consensus that is expounded in the semantics of the sacred is present to members in the form of an idealized agreement transcending spatiotemporal changes. This furnishes the model for all concepts of validity, especially for the idea of truth.¹²⁰

The normative consensus, articulating itself as the sphere of the sacred, is the prototype of an idealized agreement, 'of an intersubjectivity related to an ideal communication community',¹²¹ and thus the prototype of potential intersubjective validity. Through the linguistification of the sacred, this prototype of all validity is released from its paleo-symbolic shell and becomes available as a foundation for potential rational validity. And finally, the concept of norm-validity itself loses the privileged status it originally possessed in the context of the sacred, and is transformed by way of the differentiation of the modes of validity into the form of a discursively redeemable validity-claim analogous to the validity of truth.¹²² In this way we end up with 'the binding force of moral agreement grounded in the sacred' replaced by 'moral agreement that expresses in rational form what was always intended in the symbolism of the holy: the generality of the underlying interest'.¹²³

With this fascinating sequence of thoughts, Habermas does indeed appear to have succeeded in anchoring moral ought so deeply in the general structures of linguistic communication that the question about its rational sense becomes redundant. If it turned out that an awareness of moral obligation represented, as it were, the core of all possible rationality of linguistic communication, then the problem of moral ought in the form which troubled Kant could finally be seen to be a pseudo-problem. Now, Habermas's 'genealogical' reconstruction of a grammatically differentiated concept of validity draws on theoretical assumptions which need to be substantiated as the reconstruction proceeds. Habermas uses the concepts of an 'idealized agreement' or an 'ideal communication community' in the sense of a presupposed consensus theory of truth. Only if the reconstruction could be clarified at crucial points independently of such presuppositions would it be possible to derive from it an *independent* argument for the theoretical premises of discourse ethics. But in its crucial aspects Habermas's reconstruction seems to me far from clear. If Habermas is interpreting the symbolism of the

sacred as the expression of a *moral* agreement, then this might perhaps be justified in the sense of a functionalist way of looking at things; but it does not seem to me justified as a thesis about the priority of moral validity over the other modes of validity in terms of linguistic pragmatics. A more obvious course, it seems to me, would be to assign the concept of the sacred, as Habermas uses it, to a way of thinking for which the modes of validity that later became differentiated are not yet clearly distinguishable from one another, so that the boundaries between moral validity and truth-validity, for example, were still fluid. If we adopt a perspective such as this, then the problem of differentiation presents itself in a different light than it does in the context of Habermas's development of the thoughts of Durkheim. It presents itself, in fact, rather more in the shape that Habermas himself gives it in volume 1 of his *Theory of Communicative Action*, in the section entitled 'Some Characteristics of the Mythical and the Modern Ways of Understanding the World'.¹²⁴ Habermas there establishes a connection between the 'closedness' of mythical ways of thinking and the absence of fundamental differentiations of the kind that are characteristic of modern ways of thinking. Moreover, he does not only mean differentiation between various modes of validity, but also differentiation between causal and symbolic connections, between culture and nature, between language and world. The absence of such differentiations makes it quite impossible to envisage the sphere of symbolic validity as a sphere of criticizable validity-claims. The mythical way of thinking is, so to speak, still encapsulated in itself because it has not yet developed the linguistic resources that would allow it to look upon itself reflexively.

Evidently there is not yet any precise concept for the nonempirical validity that we ascribe to symbolic expressions. Validity is confounded with empirical efficacy. I am not referring here to special validity claims – in mythical thought diverse validity claims, such as propositional truth, normative rightness, and expressive sincerity are not yet differentiated. But even the diffuse concept of validity in general is still not freed from empirical admixtures. Concepts of validity such as morality and truth are amalgamated with empirical ordering concepts, such as causality and health. Thus a linguistically constituted worldview can be identified with the worldorder itself to such an extent that it cannot be perceived as an interpretation of the world that is subject to error and open

to criticism. In this respect the confusion of nature and culture takes on the significance of a reification of worldview.¹²⁵

If we present the problem of differentiation like this, then for a start it becomes understandable why the sphere of the sacred has been interpreted in the course of the history of anthropology not only as a sphere of primordial *norm-validity* (an antecedent form of morality), but also as a sphere of primordial explanation of the world (an antecedent form of science), as a sphere of mimetic-expressive actions (an antecedent form of art), or even as a sphere of still crude attempts to master the world (magic as an antecedent form of technology).¹²⁶

In reality it seems impossible to consign the interplay of symbolism and ritual in the sphere of the sacred definitively to *one* of these functions.¹²⁷ Habermas himself points to the connection that Durkheim emphasizes between the moral binding function of the sacred and its function as a mirror and external fixation of a collective identity. The collective identity of the group is the consciousness of a 'we' that becomes capable of experiencing itself in sacred symbols and rites, and which simultaneously discharges and regenerates itself in these symbols and rites.¹²⁸ From the functional point of view of keeping the group together, this means 'that the motivational makeup of the associated individuals is taken hold of symbolically and structured through the *same* semantic contents.'¹²⁹ But the sacred is capable of fulfilling this normative binding *function* precisely because cognitive, mimetic-expressive and moral contents are not separated from each other within its *semantics*. Even if we accept Habermas's thesis that it is possible to understand those forms of affective ambivalence that surround the sacred (the close involvement of a sense of salvation with a sense of terror, of respect with horror, of attraction with repugnance) as primordial forms of the emotional ambivalence associated with moral obligations,¹³⁰ we could not really speak of *moral feelings* because a concept of moral obligation, however rudimentary, would still be missing. And such a concept seems to *presuppose* that differentiation of spheres of validity which Habermas wants to use it to explain.

Against this, of course, it would be possible to say that there is a concept of norm-validity already *available* in the sphere of the sacred that is sufficient to bear the weight of Habermas's argument (consider for example prescribed rituals and taboos).

If the authority of the sacred means that *every* prescription, *every* rule is surrounded, as it were, with the aura of an unconditional 'must' and charged with the affective forces that this implies, then we might conclude from this that the norm-consciousness assigned to the sacred is, by virtue of its structure, moral.¹³¹ This would mean that a norm-consciousness was only able to constitute itself *as* a moral consciousness, even if its most important contents – ritual and taboo – were perhaps not moral in our sense. Seductive as this idea is, it seems to me psychologically and conceptually implausible. For it is surely possible to see from the fragmented remnants of ritual and taboo-dominated practices that extend into our own culture that the unconditional ought that is connected with the prescription of rituals and taboos not only does not need to possess any moral *content*, but can also express constraints of a quite different nature from those of a moral ought – and needs of a quite different nature from the need for personal recognition or self-respect. I can only attempt to substantiate these thoughts conceptually here, not anthropologically or psychologically. When a child, for example, insists on a particular order of things or actions, such as a ritual of storytelling or reading aloud that is precisely fixed right down to the last word, then there is certainly an unconditional sense of 'right' and 'wrong' implied in this insistence, and thus an unconditional 'ought'; but with this 'ought', this sense of 'right' and 'wrong', the child's ego is defending the particular order of a world in which it feels at home. An infringement of this order is a threat to the ego. When the child says 'You must,' what is really meant is 'This is the way *it* must be.' This does, of course, constitute a genuine moral claim on others, but it is a moral claim which can only be recognized by someone who recognizes the non-moral character of the 'right' and 'wrong' in question. The *moral* claim is the claim for the needs of the child to be respected; but to express the point paradoxically, what the child is demanding is not that its moral claim be respected, but that the right order of things be respected. This right order of things is not a moral order, it is much rather an order of the world without which not only the child, but – in an expanded sense – ultimately nobody can feel at home with themselves.

What I want to show with this example is the possibility of a categorical 'ought', of a norm-consciousness, that cannot be called moral, even though it is affectively highly charged, because it is

not moral from the point of view of its *function*. This becomes clear if we were to try – retrospectively, as it were – to provide it with a moral *grounding*. It would not then turn out to be the case that the demand was morally unfounded; it would rather turn out to be the case that the demand was not a moral one. But if this is correct, we could not simply ascribe the existence of an affectively highly charged norm-consciousness in archaic societies to morality. We should have rather to assume that this norm-consciousness is of a ‘mixed’ nature. We cannot distinguish the moral from the non-moral elements of these norms by considering whether the norms are well founded; we can only do so by considering what sort of grounding would be conceivable once the ‘grounding game’ had been introduced. Thoughts along these lines are indeed to be found in the literature of anthropology, where the realm of the taboo is discussed. Robin Horton, for example, has followed Mary Douglas¹³² in connecting the taboo with the ‘protective’ attitude of archaic societies towards a system of classification, which means that they experience any challenge to that system as a threat.¹³³ If the anthropological observations are correct, then here, too, the affectively highly-charged distinction between ‘right’ and ‘wrong’, between ‘good’ and ‘evil’, would be connected with the stabilization of a collective identity – but it would not be possible to call it ‘moral’.

Finally we could also remind ourselves of the ‘quasi-sacred’ character that the rules of games, and even rules of etiquette can assume in our society in certain contexts – and not only among children. This is *shown* by the high affective charge that such rules possess. Now, the claim that we should abide by the rules of the game or of etiquette is, of course, always a moral claim *as well*. But the rules *themselves* are not moral in nature, even if they contain a categorical ‘ought’.¹³⁴ The rules say we ought to (or may not) do a certain thing in certain situations, or that we ought to (or may not) do something in a certain way. This ‘ought’ or ‘may’ is not a moral ‘ought’ or ‘may’, it is rather the ‘ought’ or ‘may’ of rules which are constitutive of the playing of a game, or at least constitutive of a certain way of playing a game. We might suppose that under conditions of scant cognitive and social differentiation the categorical ‘ought’ of *any* rule has the aura of a moral ‘ought’ conferred upon it; but this could only mean that the moral rules only become distinguishable from the non-moral ones as the differentiation grows. And what this implies is not

only that conventional moral norms are replaced by moral principles, but also that conventional norms split up, so to speak, into moral and non-moral rules (which include grammatical, aesthetic, juridical rules, and constitutive rules of all kinds). What I mean to say is this: there are conceptual as well as empirical reasons for supposing that the normative consensus of archaic societies cannot be *equated* with a moral consensus. I believe that the only reason why we easily lose sight of this fact is because the concept of ‘conventional’ moral consciousness, as it has come to be commonly used in the wake of Kohlberg, carries the inherent suggestion that all ‘conventional’ norms are precursors of moral norms or norms that can be morally grounded, as if they had the same *point* (or the same function) as moral norms. But even if we argue on the premise that the common norms of an archaic society *express* the ‘generality’ of an ‘underlying interest’, it is not possible to infer from this that they always carry the *intention* of asserting a common – as opposed to an individual – interest, as Habermas says that they do.

In other words, even if the concept of norm-validity that is already available in the sphere of the sacred can be characterized as having an affectively highly-charged unconditional ought, it does not follow from this that this concept of norm-validity may be equated with a primordial concept of *moral* validity. Rather it is to be expected that the concept of moral validity is encapsulated within this primordial concept of norm-validity in precisely the same way that the scientific explanation of the world is encapsulated within the mythical interpretation of the world, namely as *one* of several things it might signify. Which potential *rational* sense is expressed by the unconditional ought of ‘conventional’ norms would then depend on which kinds of justification become available when the ‘conventional’ way of understanding the world opens itself up to reflection. But if it is not the case that every (‘conventional’) categorical ‘ought’ is, in the meaning of the term, a moral ‘ought’, then the problem of differentiation becomes transferred onto that very primordial concept of norm-validity that Habermas presupposes. This would mean that it might well be permissible to interpret the validity of ought as a universal type of validity, but that precisely *as such* it could not be equated with moral validity. And this exactly corresponds with what I had to say about Habermas’s grammatical reconstruction of normative validity-claims at the beginning of this section.

I am very well aware of the tentative character of these reflections, but I hope that I have shown that Habermas's suggestive and imaginative development of Durkheim's interpretation of the sacred throws up conceptual problems which are hardly fewer in number than those it is intended to solve. This is why I cannot at any rate see in Habermas's reconstruction any *additional* argument for the consensus-theoretical premises of discourse ethics which I earlier subjected to a purely immanent critique. But if, as I suspect, the concept of moral obligation is the result of a differentiation of spheres of validity (including precisely *normative* spheres of validity), then this would in turn suggest that (universal) obligations of rationality should be distinguished from (specific) moral obligations, as I have distinguished them above. As I see it, the real point of this distinction for a theory of rationality is that it alone enables us to think in terms of a 'plural' and open concept of rationality which neither depends on fundamental groundings nor looks for ultimate reconciliations. It seems to me that the consensus-theoretical premises of discourse ethics, which are closely associated with the interpretation of moral ought as a pragmatic universal in linguistic terms, is an obstacle to the development of such a plural and open – but in no way relativistic – concept of rationality. Moreover, I believe that this is also the concept of rationality that Habermas himself always has in mind when he 'translates' the meta-theoretical premises of universal pragmatics into an analysis of the normative content of modernity. The thing that particularly becomes clear whenever he does this is that there is absolutely no need for strong consensus-theoretical premises in order for us to be able to conceive of 'the rationalization of the life-world' (in general) or a dialogic 'opening-up' of ethics (in particular). In his latest book, *The Philosophical Discourse of Modernity*,¹³⁵ Habermas defines the historical goals or 'vanishing points' of a potential rationalization of the life-world as follows: 'for culture, a condition of the constant revision of traditions that have been unthawed, that is, that have become reflective; for society, a condition of the dependence of legitimate orders upon formal and ultimately discursive procedures for establishing and grounding norms; for personality, a condition of the risk-filled self-direction of a highly abstract ego-identity'. For there arise through the rationalization of the life-world 'structural pressures toward the critical dissolution of guaranteed knowledge, the establishment

of generalized values and norms, and self-directed individuation (since abstract ego-identities point toward self-realization in autonomous life projects).¹³⁶ The 'vanishing points' of a rationalization of the life-world to which Habermas refers are not the structures of an ideal communication community, but the 'structural pressures' of a life-world imbued with a commonality of universalistic values and the consciousness of universal obligations to rationality. What he is characterizing is not an *ideal condition* for society, but a set of *problems and possibilities as they actually exist* in modern societies, which cannot rationally be circumvented. The 'vanishing points' of the rationalization of the life-world are actually vanishing points of an *understanding* of rationality from which we can only retreat at the cost of regression, suppression or terror. It is only on the basis of this understanding of rationality that appropriate ways can be found to process the substantial problems of the social order and the good life, and that the 'potential for negation inherent in the process of reaching agreement in language'¹³⁷ can be developed in such a way as to leave open the possibilities of a good life, the possibilities of critical revision, and the possibilities of innovative change.

Rationalization of the lifeworld means differentiation and condensation at once – a thickening of the floating web of intersubjective threads that simultaneously holds together the ever more sharply differentiated components of culture, society, and person. The reproductive mode of the lifeworld does not change linearly in the direction indicated by the catchwords 'reflexivity', 'abstract universalism', and 'individuation'. Rather, the rationalized lifeworld secures the continuity of its contexts of meaning with the discontinuous tools of critique; it preserves the context of social integration by the risky means of an individualistically isolating universalism; and it sublimates the overwhelming power of the genealogical nexus into a fragile and vulnerable universality by means of an extremely individualized socialization.¹³⁸

But if this suggestive image of a rationalized life-world cannot signify a potential ideal state of affairs, if it is rather the case that it contains a description of structural changes which are to some extent going on before our very eyes, then it also becomes clear that the concept of a rationalization of the life-world is too unspecific a term with which to describe the *particular* problems

of *specific* societies and the ways in which they are deficient in rationality. The rationalization of the life-world is after all not a process at the end of which it would even be possible to *conceive* of a *perfectly* rational life-world (which would indeed be an idea without any clear sense); it is much rather a process in which the consciousness that there are no secure foundations for potential validity is put into effect socially, and with it the consciousness that the only means by which it is possible to secure a network of underlying common orientations and values and to ensure that this network is repeatedly established anew are those of communicative and argumentative praxis. This process is *directed* in the sense that 'the development of the potential for negation inherent in the process of reaching agreement in language' can only be conceived as a process of learning and innovation. But the point of reference for this process is not an ideal communication community conceived as situated in the future, but the present with all those pathologies, irrationalities, psychological blocks and inhumanities which may be empirically observed.

I think that the two models of differentiation which I have distinguished above imply two alternative possibilities for conceiving of the unity of reason in conjunction with the differentiation of its separate moments. The first model, that of 'consensus theory', remains tied to a perspective of reconciliation which will be formulated in either romantic-utopian or rationalistic terms, depending on the emphasis adopted. The unity of reason is here conceived from the perspective of an ideal final situation in which understanding has been reached, and in which the separated moments of reason would have arrived at a constellation of definitive reconciliation. The second model, on the other hand, is comparatively conventional, linking *directly* with that sense of problems which predominates in modern European philosophy, namely that, among the modes of validity that have become differentiated in the course of time, the one that is more difficult to understand is not the validity of truth, but that of moral ought. Undoubtedly this is *also* connected with deep-seated 'logo-centric', i.e. scientific preconceptions of modern philosophy. But that is not the whole story. The puzzling aspect of moral ought is rather that it is here that the linguistification of the sacred encounters resistances which have no equivalent in the area of truth-validity. The fear that moral consciousness must lose its firm footing if it is not supported by the authority of the sacred

is, of course, a topos of the counter-enlightenment, **but** the fact that the effectiveness of moral arguments remains dependent on preconditions which are not only cognitive, but also affective in nature shows that this fear does have a foundation in fact. A rational equivalent to a moral agreement supported by sacred or religious authority is only possible in so far as a successful adaptation to conditions of mutual recognition between persons – in both cognitive *and* affective terms – has taken place. To the extent that this is not the case, moral arguments lose their point of purchase, although this need not necessarily also be the case with arguments of an empirical or technical nature. There *is* a lack of moral sense – that is a fact. But we can only interpret this as a deficiency in rationality if we assume conditions which can precisely not be fulfilled wherever this lack of moral sense manifests itself.

In the second of the two models of differentiation which we have distinguished, we find *on the one hand* that the autonomy of the differentiated modes of validity is treated seriously; that is what makes it impossible to conceive of the unity of reason from the perspective of an ideal communication community in which the partiality of the separate moments of reason would finally have been sublated in the unity of a moral ideal. *On the other hand* this second model of differentiation enables us to give sharper contours to the internal *connection* between the differentiated modes of validity. As I tried to show earlier, moral discourse can be understood to a large extent as discourse about 'facts' – in the broadest sense – or about the appropriateness and completeness of interpretations of situations. This is why, in the sphere of morality, the transition from 'is' to 'ought' is always preordained, not by any ultimate normative *premises*, but by the 'moral point of view' itself.¹³⁹ But in the interpretation of the facts that are relevant to moral judgements, aesthetic experiences are always brought to bear – there is a fluid boundary also between moral discourse and aesthetic discourse. But then discourse about facts is not impervious to moral or aesthetic viewpoints either. Not only is the language in which we speak about the human life-world and history impregnated with value-judgements, the facts also present themselves differently in the light of various possible orientations within the life-world – and within these orientations it is always the case that moral attitudes and empirical convictions are already linked together. This

appears to point towards a circular process, and thus ultimately towards relativism. This circle is not a theoretical problem, however, but a practical one, representing factual boundaries to rational discourse which repeatedly become apparent. We can break the circle only from within, through the application of a reason which does not withhold *any* validity-claim from critical scrutiny. The reason why we are not confronted with a circular problem in theoretical terms is that the mesh of practical orientations and empirical convictions is not secured at any point by ultimate premises which would not be amenable in principle to immanent criticism or to criticism in the light of new experience.

In the life-world at least, then, there are always interconnections between moral, practical and technical, and aesthetic perspectives or modes of discourse, as well as those concerned with truth. Rationality manifests itself here both in the ability to *distinguish* between various perspectives, and also in the ability to *connect* them with each other in the right way. But to a greater or lesser degree, much the same is true of the institutionally differentiated 'value-spheres' of science, art and law. In the case of law this seems to me self-evident. Where art is concerned, Martin Seel has shown¹⁴⁰ that it is possible to explain the sense of aesthetic validity with reference to the interrelationship between empirical, moral, and expressive validity-claims in aesthetic discourse. And in the case of science, finally, the problem presents itself in different ways, depending on the type of science in question. Human and social sciences participate by their very nature in the interrelationship between spheres of validity that is characteristic of the life-world, even when they specialize in questions of empirical or theoretical truth. Perhaps the natural sciences in their mathematical aspect are the only paradigm of an empirical science which is only affected at its 'edges' by normative, let alone aesthetic issues – they are affected by moral issues where they are concerned with the aims and applications of research, and by methodological and 'grammatical' issues where the *foundations* of scientific inquiry are concerned. It is precisely the mathematical language of the natural sciences that has become the true paradigm for the differentiation of spheres of validity in all modern philosophy, i.e. the paradigm of pure truth-validity.

This is still true of Habermas and Apel. And if this is the point of reference one takes, then of course the question of the sense

and possibility of moral validity stands out with great clarity, but I doubt whether it is possible to achieve an adequate reconstruction of the internal *connection* between the spheres of validity using this point of reference. It is true that knowledge derived from the natural sciences is playing an ever greater part in moral controversies (the most recent example is the Aids question), but it hardly makes sense to assume that natural science also provides the measure of what is *real*, in the sense of what is or is not an empirical fact, for the purposes of moral argument – as Sellars once argued that it did.¹⁴¹ The difference between 'is' and 'ought' can be easily illustrated in terms of the contrast between 'He told a lie' and 'One ought not to tell lies', or between 'He is innocent' and 'Innocent persons must not be condemned'. Sellars, as an empiricist through and through, would not *ultimately* be able to accept the possible truth of the factual statements we have just cited because they are not recognizable in the terms of natural science as factual statements.¹⁴² But if we measure the concept of an empirical fact by Sellars' yardstick as, curiously enough, Apel and Habermas do, then the sphere of social facts becomes a puzzling entity which has, so to speak, to be 'reconstituted' from the spheres of validity as they have become historically differentiated.¹⁴³ Perhaps this also explains why, in universal pragmatics, the *one* (extreme) concept of truth-validity is only opposed by *one* (extreme) concept of normative validity. It would then be the case, as I suggested earlier, that consensus theory is the complement of a covert scientific residue in the theories of Apel and Habermas.

Against this I would argue that while 'is' is fundamentally different from 'ought', there are various criteria of 'ought', just as there are of 'is', depending on the sense of the statement in either case, and that there are therefore various possible forms in which 'is' or 'ought' statements can be grounded or criticized, as well as various possible relationships between the two. Since it is in any case not possible to allocate aesthetic validity to *one* validity-claim in Habermas's sense,¹⁴⁴ it seems logical not to distinguish between spheres of validity along the lines of a typology of dimensions of validity grounded in speech-act theory and thus to allocate them to 'theoretical', 'practical' or 'aesthetic' discourse, but rather to distinguish between various types of validity-claim, and of forms of arguing to be allocated to them, *within* theoretical and practical discourse. Theoretical discourse is

concerned with the validity of statements and theories, practical discourse with the correctness of actions. Theoretical discourse might, for example, be concerned with the truth-claims (assertions, theories, explanations, interpretations, reconstructions) of mathematics, physics, historiography, literary hermeneutics or moral philosophy, to which quite different forms of argument, criteria of validity, or testing procedures may correspond. 'Science as such' is a conglomerate of individual sciences, their common feature consisting solely in a specialization in the pursuit of truth independent of practical action. Such a pursuit of truth need not be independent of questions of moral or aesthetic validity, but this does not imply that theoretical discourse necessarily has to turn into practical discourse or the discourse of art criticism where such questions arise. We are concerned here, moreover, not in the first instance with the truth of *individual* statements, but with the validity of 'concatenations' of statements (theories, explanations, reconstructions, interpretations, etc) which may be *internally* articulated in a complex way; the validity of these concatenations (which allows for a certain latitude, a 'more' or 'less') cannot be equated either with the truth of individual statements or with the adequacy of language systems.

This is why 'propositional truth' is not an adequate term for what theoretical discourse is about. Theoretical discourse is concerned above all with the validity of propositional structures of a higher kind (which can themselves, under certain circumstances, take the form of interconnections between arguments), and in this connection, of course, it is also concerned with the truth of individual statements. Practical discourse, on the other hand, is concerned with the grounding and evaluation of *actions*, i.e. with questions of whether actions are politically, juridically, economically, technically, aesthetically or morally correct, with different forms of argument and criteria for validity again corresponding to various perspectives on correctness. Whereas in theoretical discourse standards of rationality are provided by the meaning of the validity-claims which are being discussed in each particular instance, or by the internal connection between validity-claims and their presuppositions, with practical discourse we encounter the additional problem that competing standards of rationality have to be related to each other and relativized with respect to each other. Practical reason expresses itself not least as the ability to relate the various dimensions of

rationality implied in action – such as technical, economic, moral or aesthetic rationality – to each other in an appropriate fashion, and to relativize them with respect to each other. As Seel puts it, it expresses itself as an 'interrational faculty of judgement'.¹⁴⁵ The term 'faculty of judgement' carries the implication that the correct (i.e. justified) solutions which it is possible to find for the 'mediation of the moments of reason'¹⁴⁶ are only ever valid here and now; they are not universal or ultimate solutions. 'Unreason' should be understood in this connection as a partial insensitivity towards whole realms of experience and dimensions of validity, and thus as an inability to relate the various dimensions of experience and validity to each other in an appropriate fashion.¹⁴⁷ *Aesthetic* discourse, finally, is concerned neither with the validity of statements, nor with the correctness of actions, but with the meaning of aesthetic objects and whether they are successful or not, i.e. with the (aesthetic) 'validity-claims' of these objects. As with theoretical and practical discourse, interpretations, empirical assertions, and claims to moral correctness are interlinked in aesthetic discourse, but they are not the *themes*, but rather the *arguments* of aesthetic discourse, just as expressive validity-claims are. What is grounded through aesthetic discourse are aesthetic value-judgements; but these point beyond themselves to the validity-claim of the aesthetic objects on which the judgements are made, and that is something which can only be redeemed in aesthetic experience.

Theoretical, practical and aesthetic discourse are interrelated in many ways, but each is concerned with something different. Theoretical discourse aims at valid statements, explanations and interpretations; practical discourse at correct actions, attitudes and decisions; aesthetic discourse at appropriate ways of perceiving aesthetic objects. But within each of these forms of discourse, too, the various forms of argument are always – potentially at least – interlinked in many ways, because the sense of particular arguments is derived from the presence of perspectives and premises which can make it necessary *in a case of doubt* for the discussion to move to a different form of argument. But precisely these internal links between different forms of argument are not capable of explanation by means of a typology of validity-claims (propositional truth, moral correctness, [expressive] truthfulness) grounded in universal pragmatics. To put it another way, distinctions based on speech-act theory are not in themselves sufficient

to render understandable either the *difference* between 'spheres of validity' or the internal *connection* between them. The reconstruction of the unity of reason with recourse to universal pragmatics and consensus theory simultaneously pitches its theoretical argument too low and too high; that is why *on the one hand* it remains committed to foundationalist figures of thought and ones that relate to the philosophy of reconciliation, while *on the other hand* it remains peculiarly encumbered with distinctions of a scientific nature. It is distinctions of this kind which ultimately obscure the very thing which ought to be made clear, namely that the partial moments of reason communicate with each other even after they have become separated from each other.

The unity of reason can now be seen as a network of connecting lines and interchanges between theoretical, technical, moral and aesthetic issues and ways of arguing. Wherever these connections and interchanges are blocked or severed, quite specific pathologies and one-sided usages of reason result. If a form of behaviour offends against elementary requirements of consistency, or if this consistency can only be maintained at the price of a rejection of arguments and experiences, then we can call such behaviour 'irrational'. If, on the other hand, rational behaviour is reduced in such a way that *one* dimension of rationality is treated as absolute at the expense of the others, then we can call such forms of behaviour 'unreasonable', as Seel has suggested.¹⁴⁸ The term 'reasonable' might then be accorded the position currently occupied in Habermas's theory by the term 'communicative competence' which, after all, also means an *integration* of the moments of reason – except that it will no longer be possible to elucidate what is meant by 'reasonable' with reference to an ideal structural model that can be characterized in formal procedural terms. The 'steadfast pursuit of the tortuous routes along which science, morality, and art communicate with one another', as Habermas puts it,¹⁴⁹ requires discernment, imagination *and* good will; these are elements of 'reasonableness' for which there is no ideal state of affairs to be realized, but which aim rather to keep open and to extend latitude for freedom and possibilities for living. The unity of reason is realized in the interaction between partial moments of reason, and this interaction is something for which there can be no ultimate foundations or ultimate yardsticks, nor even ultimate reconciliations. Of course, reason does have a foundation – that foundation is the existence of a 'culture

of reason'.¹⁵⁰ Where such a foundation once exists, the postulate of the freedom of *all* must become a postulate of (practical) reason. This is the indispensable (practical) chiasm of reason,¹⁵¹ which Apel and Habermas, following Kant, rightly wish to preserve. But this postulate only acquires its precise sense against the background of a *lack* of freedom as it exists and as it may be experienced in concrete situations. It does not mean the attainment of any ultimate reconciliation or ideal understanding. If it were ever so that there was no longer any reason to engage in a political struggle for freedom, then freedom would still have to be preserved, transmitted, and acquired anew. But this would not even be *conceivable* within a mode of seeking to attain an ideal understanding, for any such ideal understanding would be disrupted by each new generation in turn. On the other hand, without the element of being able to begin anew there could be no freedom.¹⁵²

- 3 Ibid., pp. 57, 59.
- 4 Joan Campbell, *The German Werkbund*, Princeton 1978; Kurt Jung-hans, *Der deutsche Werkbund. Sein erstes Jahrzehnt*, (East) Berlin 1982; Lucius Burckhardt, *Der Werkbund in Deutschland, Österreich und der Schweiz*, Stuttgart 1978. There is a fine documentation of the history of the Werkbund in the book that accompanied the Werkbund Exhibition in the Staatliches Museum für angewandte Kunst in Munich, 1975: see Wend Fischer (ed.), *Zwischen Kunst und Industrie. Der deutsche Werkbund*, Munich 1975.
- 5 I am leaving out of account those chauvinistic undertones which were not absent from the early phase of the Deutscher Werkbund. (For references, see note 4, above.)
- 6 Adolf Loos, 'Kulturentartung', in *Sämtliche Schriften*, vol. 1, Vienna and Munich 1962, p. 274.
- 7 Julius Posener, 'Le Corbusier', in *Aufsätze und Vorträge 1931-1980*, Braunschweig and Wiesbaden 1981, p. 188.
- 8 W. J. Siedler, E. Niggemeyer, G. Angress, *Die gemordete Stadt*, Berlin, Munich and Vienna 1964, p. 13.
- 9 Jane Jacobs, *The Death and Life of Great American Cities*, New York 1961, chap. 22.
- 10 Theodor W. Adorno, 'Funktionalismus heute', *Ges. Schriften*, vol. 10.1, Frankfurt 1977, p. 389.
- 11 Ibid., p. 387.
- 12 Ibid., p. 388.
- 13 For what follows, cf. Jürgen Habermas, 'Moderne und postmoderne Architektur', in *Die andere Tradition* (Catalogue to exhibition no. 3 in the series 'Erkundungen'), Munich 1981, which I was able to read only after the present text was completed. The affinities with my own argument are readily apparent and by no means coincidental.
- 14 Charles Jencks, *The Language of Post-Modern Architecture*, New York 1977.
- 15 Jürgen Habermas, *The Theory of Communicative Action*, 2 vols, Cambridge, Mass., 1984 and 1987.
- 16 Cf. K. Frampton, *Modern Architecture*, Oxford and New York 1980, p. 293.
- 17 Ludwig Wittgenstein, *Philosophical Investigations*, translated by G. E. M. Anscombe, Oxford 1958, p. 46°.
- 18 Jencks, *Language of Post-Modern Architecture*, pp. 128ff.
- 19 Ibid., p. 14.
- 20 Adorno, *Aesthetic Theory*, translated by C. Lenhardt, London 1984, p. 167.
- 21 Alexander Schwab, 'Zur Abteilung Städtebau und Landesplanung',

- Die Form*, 1930, Heft 3, quoted in F. Schwarz and F. Gloor, 'Die Form'. *Stimme des Deutschen Werkbundes 1925-1934*, Gütersloh 1969, p. 157.
- 22 Ernst Bloch, *The Principle of Hope*, translated by Neville Plaice, Stephen Plaice and Paul Knight, 3 vols, Oxford 1986, p. 735.
- 23 Ibid., p. 737.
- 24 Ibid., p. 735. Cf. also A. M. Vogt, 'Entwurf zu einer Architekturge-schichte 1940-1980', in Vogt, Jehle and Reichlin (eds), *Architektur 1940-1980*, Berlin 1980, p. 12.
- 25 Lucius Burckhardt, 'Design ist unsichtbar', in *Design ist unsichtbar*, Vienna 1981.

4 Ethics and Dialogue: Elements of Moral Judgement in Kant and Discourse Ethics

A note on the references to Kant. For passages quoted from the works of Kant, those translations have been selected which yielded the greatest conceptual clarity in the English. They are the translations by Abbott (for the *Critique of Practical Reason*), Meredith (for the *Critique of Judgement*), and Paton (for the *Groundwork of the Metaphysics of Morals*). Alternative references to other English translations are given where these conveniently combine various relevant works of Kant in a single volume. The references to the collected works of Kant in German (*Werke in sechs Bänden*, ed. W. Weischedel, Darmstadt 1956-64, hereafter cited as *WSB*) are retained for readers who may wish to refer to the original wording of Kant's arguments.

- 1 On this subject, cf. Peter Sloterdijk, *Critique of Cynical Reason*, London 1988.
- 2 Jürgen Habermas, 'Diskursethik - Notizen zu einem Begründungs-programm', in Habermas, *Moralbewusstsein und Kommunikatives Handeln*, Frankfurt 1983.
- 3 Cf. Marcus G. Singer, *Generalization in Ethics*, New York 1971, pp. 37ff.
- 4 Cf. Richard M. Hare, *Moral Thinking*, Oxford 1981, pp. 8ff.
- 5 Singer, *Generalization in Ethics*, p. 38.
- 6 This is where I see the decisive weakness in what is in certain respects a thoroughly convincing reconstruction of Kant's ethics, or of a 'Kantian' ethics, by Singer. Cf. Singer, *Generalization in Ethics*, pp. 63ff.
- 7 H. J. Paton, *The Moral Law. Kant's Groundwork of the Metaphysics of Morals. A New Translation with Analysis and Notes*, London 1961, p. 91; Immanuel Kant, *Ethical Philosophy*, translated by James W.

- Ellington, Indianapolis 1983, p. 32. (WSB, vol. IV, 1956, p. 54 (BA 57).)
- 8 Bernard Gert, *The Moral Rules*, New York 1973.
 - 9 Ibid., pp. 60ff; cf. also Georg Henrik von Wright, *The Varieties of Goodness*, London 1963, pp. 197ff.
 - 10 Cf. Julius Ebbinghaus, 'Die Formeln des kategorischen Imperativs und die Ableitung inhaltlich bestimmter Pflichten', in Ebbinghaus, *Gesammelte Aufsätze, Vorträge, Reden*, Hildesheim 1968, vol. I, section 7, pp. 140–60.
 - 11 Singer, *Generalization in Ethics*, p. 240: 'If the maxim of an action cannot be willed to be a universal law, then it is wrong to act on it, we have the duty or obligation not to, and it can be said that we ought not to. However, if a maxim can be willed to be a universal law, it does not follow that it is obligatory to act on it or that it would be wrong not to. What follows is that it is permissible to do so, or not wrong (and thus right in the *permissive* sense), and hence that it cannot be said that we ought not to – which is not the same as saying that we ought to.' This is also the approach taken by Joachim Aul: 'Aspekte des Universalisierungspostulats in Kants Ethik', *Neue Hefte für Philosophie*, Heft 22, 1983, esp. pp. 85ff. Such an interpretation is not entirely foreign to Kant himself, as the following passage from notes of one of his lectures shows: 'In all moral judgements we ask ourselves: how is this action constituted if it is taken as universal? If the intention of the act is in harmony with itself when it is made into a universal rule, then the act is morally possible; if the intention of the act is not in harmony with itself when it is taken as universal, then the act is not morally possible.' Cf. Kant, *Gesammelte Schriften*, ed. Akademie der Wissenschaften der DDR, vol. XXVII (Kants Vorlesungen, vol. IV: Vorlesungen über Moralphilosophie), Berlin 1979, pp. 1276f. I am obliged to Henry Gerlach for drawing my attention to this passage.
 - 12 Cf. also William K. Frankena, *Analytische Ethik*, Munich 1972, p. 52.
 - 13 Cf. Paton, *The Moral Law*, pp. 89f; Ellington, p. 31. (WSB, vol. IV, p. 53 (BA 54).)
 - 14 See for example Kant's *Critique of Practical Reason and Other Works on the Theory of Ethics*, translated by Thomas Kingsmill Abbott, London 1927, pp. 114f. (WSB, vol. IV, p. 136 (A 49).)
 - 15 Cf. Paton, *The Moral Law*, p. 91; Ellington, p. 32. (WSB vol. IV, pp. 54–5 (BA 56–7).)
 - 16 Cf. *ibid.*, p. 92; Ellington, p. 33. (WSB, vol. IV, p. 55 (BA 58).)
 - 17 This is Schopenhauer's 'principle of justice'. Cf. Arthur Schopenhauer, *On the Basis of Morality*, translated by E. F. J. Payne, Indianapolis 1965, p. 149. I shall not be discussing Schopenhauer's critique of Kant here, but I should like to mention Schopenhauer's thesis

- that 'the concept of *ought*, the *imperative form* of ethics, applies solely to theological morality, and that outside this it loses all sense and meaning' (*ibid.*, p. 130). A similar 'sense-criticism' (in Apel's terminology) of the concept of an unconditional moral ought also occurs in recent discussions on ethics; see, for example, G. E. M. Anscombe, 'Modern Moral Philosophy', in *Philosophy* 33 (1958), and A. MacIntyre, *After Virtue*, Notre Dame, Indiana, 1981 (p. 57); cf. also P. Foot, 'Morality as a System of Hypothetical Imperatives', in *Virtues and Vices*, Berkeley 1978, pp. 163ff, and U. Wolf, *Das Problem des moralischen Sollens*, Berlin 1984, pp. 3ff. I think that we cannot ignore 'Schopenhauer's problem', as I should like to call it, even if the critique of Kant within which Schopenhauer's thesis has its context is not itself convincing. I shall touch on this problem indirectly in section XI.
- 18 Kant, *Ethical Philosophy*, translated by James W. Ellington, Indianapolis 1983, p. 48 (WSB, vol. IV, p. 520 (A 20).)
 - 19 Cf. Gert, *The Moral Rules*, pp. 128ff.
 - 20 Kant, *Ethical Philosophy*, translated by James W. Ellington, Indianapolis 1983, p. 52. (WSB, vol. IV, p. 524 (A 27).)
 - 21 On the term 'prima facie principles', cf. Hare, *Moral Thinking*, p. 38.
 - 22 *Ibid.*, pp. 25ff.
 - 23 *Ibid.*, p. 41.
 - 24 Cf. *ibid.*, p. 46.
 - 25 *Ibid.*, p. 33.
 - 26 Kant's *Critique of Practical Reason*, translated by T. K. Abbott, pp. 119, 114f. (WSB, vol. IV, pp. 140, 136 (A 54, 49).)
 - 27 *Ibid.*, pp. 114f (p. 136 (A 49)).
 - 28 Cf. Richard M. Hare, *The Language of Morals*, Oxford 1952, pp. 68f.
 - 29 Cf. Gert, *The Moral Rules*, chap. 2, esp. p. 37.
 - 30 Cf. von Wright, *The Varieties of Goodness*.
 - 31 John Rawls, *A Theory of Justice*, Cambridge, Mass., 1971.
 - 32 Cf. the thoughts of Gert in *The Moral Rules*, chap. 10: 'Why Should One Be Moral?'
 - 33 Cf. *ibid.*, pp. 204ff.
 - 34 A representative selection can be found in Friedrich Kambartel (ed.), *Praktische Philosophie und konstruktive Wissenschaftstheorie*, Frankfurt 1974. See also Oswald Schwemmer, *Philosophie der Praxis*, Frankfurt 1971; Paul Lorenzen and Oswald Schwemmer, *Konstruktive Logik, Ethik und Wissenschaftstheorie*, Mannheim 1973.
 - 35 John R. Silber, 'Procedural Formalism in Kant's Ethics', *Review of Metaphysics*, vol. XXVIII, no. 2 (1974).
 - 36 Kant, *The Critique of Judgement*, translated by J. C. Meredith, Oxford 1952, p. 152. (WSB, vol. V, 1957, p. 390 (B 158).) The maxims in question are: '(1) to think for oneself; (2) to think from the stand-

- point of everyone else; (3) always to think consistently.'
- 37 Silber, 'Procedural Formalism', p. 216.
- 38 Ibid., p. 199.
- 39 Ibid., p. 221. Cf. Immanuel Kant, *Religion within the Limits of Reason Alone*, translated by T. M. Greene and H. H. Hudson, La Salle, Illinois, 1960, p. 62. (*Werke in sechs Bänden*, vol. V, p. 722.)
- 40 See for instance K.-O. Apel, D. Böhler and G. Kadelbach (eds), *Funkkolleg Praktische Philosophie/Ethik: Dialoge 2*, Frankfurt 1984, esp. units 18–20. K.-O. Apel, 'Ist die Ethik der idealen Kommunikationsgemeinschaft eine Utopie?', in W. Vosskamp (ed.), *Utopieforschung*, vol. 1, Stuttgart 1982, and 'Kant, Hegel und das aktuelle Problem der normativen Grundlagen von Moral und Recht', in Arno Werner (ed.), *Filosofi och Kultur*, Lund 1982. On the question of fundamental grounding, see esp. K.-O. Apel, 'The Problem of Philosophical Fundamental Grounding in the Light of a Transcendental Pragmatic of Language', in *Man and World* 18 (1975), pp. 239–75, and 'Sprechakttheorie und transzendente Sprachpragmatik zur Frage ethischer Normen', in K.-O. Apel (ed.), *Sprachpragmatik und Philosophie*, Frankfurt 1976; also 'The a priori of the communication community and the foundations of ethics', in *Towards a Transformation of Philosophy*, translated by G. Adey and D. Frisby, London 1980.
- 41 Habermas, *Moralbewusstsein und Kommunikatives Handeln*, Frankfurt 1983, p. 136.
- 42 Ibid., pp. 136f.
- 43 Thomas McCarthy, *The Critical Discourse of Jürgen Habermas*, Cambridge, Mass., 1978, p. 326.
- 44 Paton, *The Moral Law*, p. 91; Ellington, p. 32, (WSB, vol. IV, p. 54 (BA 57).)
- 45 Cf. above, section III (Excursus), pp. 131–5.
- 46 This description first appears in Jürgen Habermas, 'Wahrheitstheorien', in Helmut Fahrenbach (ed.), *Wirklichkeit und Reflexion. Festschrift für Walter Schulz*, Pfullingen 1973, esp. pp. 252ff. The critique that follows has some elements in common with the comprehensive and trenchant critique of consensus theory by R. Zimmermann in *Utopie – Rationalität – Politik*, Munich 1985, pp. 303ff.
- 47 See for example 'A Philosophico-Political Profile', in Jürgen Habermas, *Autonomy and Solidarity*, London 1986, pp. 162–3. Habermas here adds the qualification, however, that the consensus or discourse theory of truth simultaneously 'undermines the clear distinction between meaning and criterion' (p. 163).
- 48 Personal letter.
- 49 Ludwig Wittgenstein, *Philosophical Investigations*, Oxford 1963, p. 88^e (para. 242).

- 50 Habermas, 'Wahrheitstheorien', p. 244.
- 51 Ibid., p. 249.
- 52 This comes close to the approach taken by Apel. For full formulations of it cf. Karl-Otto Apel, 'Scientism or Transcendental Hermeneutics', in *Towards a Transformation of Philosophy*, translated by G. Adey and D. Frisby, London 1980, pp. 105, 115f. The approaches of Apel and Habermas differ only in their initial point of departure and their emphasis; the difference in the results they achieve is not always easy to determine. Apel appeals to the authority of Habermas, for example, when he postulates the necessity of assuming an 'ideal speech situation' as a precondition for the possibility of argument. (Cf. K.-O. Apel, 'Sprechakttheorie und transzendente Sprachpragmatik zur Frage ethischer Normen', in Apel (ed.), *Sprachpragmatik und Philosophie*, Frankfurt 1976, p. 121.) On the other hand, Habermas views the *rational* consensus (i.e. a consensus achieved under the conditions of an ideal speech situation) as *eo ipso* a possible *infinite* consensus. (Cf. Habermas, 'Wahrheitstheorien', p. 239: '... the meaning of truth is not the circumstance that a consensus has been reached, but that whenever we enter into a discourse, at whatever time and in whatever place, a consensus can be achieved under conditions which distinguish it as a fully grounded consensus.') There is a simple reason why I have not included what is for Habermas the self-evident precondition of the infinite repeatability of rational consensus in my considerations from the outset, and that is that as long as the presence of the formal conditions of an ideal speech situation is understood as a criterion of truth (cf. Habermas, 'Wahrheitstheorien', pp. 239f), then the possibility of an infinite consensus is merely a *consequence* of the *rationality* of consensus as defined by formal conditions. The explicatum of the truth concept is not the infinite consensus, but the rational one. This was what prompted my initial objections: I tried to show that the – formally characterized – structural features of an ideal speech situation cannot represent an appropriate criterion for truth; *either* the criterion is false, *or* it is inherently vacuous and thus no criterion at all. If we now take into consideration the connection that Habermas assumes between the rationality and the infinite repeatability of consensus, then it becomes clear that the ideal speech situation is conceived from the very beginning rather in terms of a vacuous criterion. For if the infinite repeatability of consensus *follows analytically* from their rationality, then by the same token it follows that a consensus which subsequently turns out to be false and unable to stand up to criticism *cannot* have come about under the conditions of an ideal speech situation (cf. 'Wahrheitstheorien', p. 257f). But then the

- permanence of consensuses would in reality be the criterion of their rationality (their truth). This is the second variant of consensus theory, which corresponds rather to the fundamental intuitions of Apel.
- 53 Habermas, *Autonomy and Solidarity*, pp. 162ff.
- 54 *Ibid.*, p. 162.
- 55 *Ibid.*, pp. 162f. [Translator's note: I have rendered Habermas's term 'Voraussetzungen' as 'preconditions'; in *Autonomy and Solidarity* it is translated – wrongly for its context – as 'presuppositions'].
- 56 Apel does, however, say that this is a regulative idea which can never be 'fully realized'. Cf. K.-O. Apel, D. Böhler and G. Kadelbach (eds), *Funkkolleg Praktische Philosophie/Ethik: Dialoge 2*, Frankfurt 1984, p. 136; also 'Kant, Hegel und das aktuelle Problem der normativen Grundlagen von Moral und Recht', in Arno Werner (ed.), *Filosofi och Kultur*, Lund 1982, p. 85.
- 57 The reflections of C. F. von Weizsäcker on the possibility of achieving a unity, and thus a final perfection, of physics are of relevance here: see C. F. von Weizsäcker, *Die Einheit der Natur*, Munich 1971, esp. pp. 207ff. Von Weizsäcker here advances the ambitious hypothesis that it must ultimately be possible to derive all the fundamental principles of a perfected physics from an analysis of the preconditions for the possibility of experience (*ibid.*, p. 217). The idea of a 'final' (in the sense of totally adequate) language of physics occurs in a different form nowadays in the tradition of American pragmatism, where it finds its subtlest elaboration in Wilfrid Sellars' philosophy of 'scientific realism'. For Sellars, scientific progress represents a process of continual language criticism, much as it did for Peirce; according to this conception of things, 'reality' would be the correlate of those physical theories which had ultimately been found to be true. Cf. Wilfrid Sellars, *Science, Perception and Reality*, London 1963, esp. pp. 119, 126; 'Scientific Realism or Irenic Instrumentalism. Comments on J. J. C. Smart', in R. S. Cohen and M. W. Wartofsky (eds), *Boston Studies in the Philosophy of Science*, vol. II, New York 1965, esp. p. 204; 'Counterfactuals, Dispositions, and the Causal Modalities', in H. Feigl, M. Scriven and G. Maxwell (eds), *Minnesota Studies in the Philosophy of Science*, vol. II, Minneapolis 1958, esp. p. 263; also 'Theoretical Explanation', in Sellars, *Philosophical Perspectives*, Springfield, Ill., 1967.
- 58 Karl-Otto Apel, *Towards a Transformation of Philosophy*, pp. 93ff.
- 59 Cf. Apel, 'From Kant to Peirce', *ibid.*, p. 87.
- 60 Apel, 'Scientism or Transcendental Hermeneutics', *ibid.*, p. 125.
- 61 See for example 'From Kant to Peirce', *ibid.*, p. 87.

- 62 Charles Sanders Peirce, *Collected Papers*, 5.311 (quoted by Apel, *ibid.*, p. 87).
- 63 *Ibid.*, pp. 87f.
- 64 *Ibid.*, p. 88.
- 65 Cf. 'Scientism or Transcendental Hermeneutics', *ibid.*, p. 113.
- 66 *Ibid.*, pp. 125f.
- 67 Cf. *ibid.*, p. 123.
- 68 *Ibid.*, p. 112.
- 69 *Ibid.*, p. 125.
- 70 *Ibid.*, pp. 120f.
- 71 Cf. *ibid.*, p. 123.
- 72 *Ibid.*, p. 125.
- 73 *Ibid.*
- 74 *Ibid.*
- 75 K.-O. Apel, *Transformation der Philosophie*, vol. II, Frankfurt 1973, p. 348 ('Der transzendentalhermeneutische Begriff der Sprache').
- 76 Adorno's 'Meditations on Metaphysics' in part three of *Negative Dialectics* are a sustained attempt to rescue the theological motif which in Kant's case entered into the construction of the connection between the concept of the intelligible and the postulates of pure practical reason. It is true that Adorno attempts to release this theological motif – in materialistic fashion – from the rigid opposition of immanence and transcendence; but by taking it literally, i.e. as an expectation of the resurrection of the body, he also precludes for himself the possibility of merely levelling out the difference. He sees the ambiguity and the aporetic quality of Kant's construction as ultimately justified by the fact that the absolute is *for us* veiled in black, as he puts it elsewhere. 'That no reforms within the world sufficed to do justice to the dead, that none of them touched upon the wrong of death – this is what moves Kantian reason to hope against reason. The secret of his philosophy is the unthinkability of despair. Constrained by the convergence of all thoughts in something absolute, he did not leave it at the absolute line between absoluteness and existence; but he was no less constrained to draw that line. He held on to the metaphysical ideas, and yet he forbade jumping from thoughts of the absolute which might one day be realized, like eternal peace, to the conclusion that therefore the absolute exists. His philosophy – as probably every other, by the way – circles about the ontological argument for God's existence; but his own position remained open, in a grandiose ambiguity. There is the motif of "Muss ein ewiger Vater wohnen – must live an eternal father," which Beethoven's composition of Schiller's Kantian Hymn to Joy accentuated in true Kantian spirit,

on the word "must". And there are the passages in which Kant – as close to Schopenhauer here as Schopenhauer later claimed – spurned the metaphysical ideas, particularly that of immortality, as imprisoned in our views of space and time and thus restricted on their part. He disdained the passage to affirmation.' (Theodor W. Adorno, *Negative Dialectics*, translated by E. B. Ashton, New York 1973, p. 385.)

77 Cf. Albrecht Wellmer, 'Adorno, Anwalt des Nicht-Identischen', in Wellmer, *Zur Dialektik von Moderne und Postmoderne*, Frankfurt 1985, pp. 160f.

78 Kant, too, considers the idea of an infinite approximation to a condition of moral perfection and thus to the kingdom of God as a practically necessary idea. (Cf. *Religion within the Limits of Reason Alone*, pp. 29f, 42, 54, 60f, 113; (WSB, vol. V, pp. 682f, 697, 713, 720f, 786f.) But he considers it precisely as a *practically* necessary idea; it is really the idea of a potentially endless progress 'from a deficient to a better good' (ibid., p. 60). As far as the ideal 'ultimate goals' of moral perfection or of the 'ethical state' (the 'kingdom of virtue') are concerned, Kant's thoughts remain extraordinarily ambiguous; for it is impossible to overlook Kant's remarks to the effect that a realization of these ultimate goals is something which a finite reason under finite conditions is scarcely capable of conceiving adequately (cf. ibid., pp. 58 (footnote), 60, 126). The theological motif of which I spoke earlier (note 76) asserts itself precisely at those points where Kant tries to look beyond the duty to moral progress and imagine its ultimate goals (moral perfection or the kingdom of God) as having been realized by creatures of finite reason. At any rate, Kant was aware of the difficulty of *conceiving* a kingdom of ends, which belongs to the sphere of the intelligible, as something empirically realized. Apel tries to avoid this difficulty by calling into question, as Peirce does, the Kantian distinction between *noumena* and *phainomena* and also that between regulative principles and moral postulates (cf. Apel, 'From Kant to Peirce', p. 90). But in this way the idea of an ideal communication community is accorded, in addition to its regulative function, a *constitutive* function not only for empirical cognition, but also for moral judgement. This means that Kant's difficulties with the sphere of the intelligible are carried, with all ambiguities eliminated, right into the centre of epistemology and moral philosophy. At heart, these difficulties arise from the fact that a subject in the singular is the 'highest point' of (Kant's) transcendental philosophy. My objection to Apel is that the ideal communication community still occupies the position of a subject in the singular – a subject, admittedly, which is now conceived as something *in the process of becoming* within this world. (Apel speaks

explicitly of a *single* transcendental subject which 'on the one hand must always be anticipated, and on the other hand is always yet to be realized': cf. 'Sprechakttheorie und transzendente Sprachpragmatik', p. 127.)

79 Cf. Apel, 'Scientism or Transcendental Hermeneutics', p. 124.

80 It would of course be possible to *equate* the anticipation of an infinite consensus with the idea of an ideal communication community. In fact this would appear to offer *one* possible sense of the concept of an ideal communication community that would not be suspect; it is my belief that Habermas, for example, occasionally uses the concept in this sense (cf. Habermas, 'Moral und Sittlichkeit. Treffen Hegels Einwände gegen Kant auch auf die Diskursethik zu?', *Moralbewusstsein und kommunikatives Handeln*, p. 13). In this instance the ideal communication community is simply the community of *all* beings capable of speech whom we imagine, as it were, ideally assembled at one time. But within this meaning of the concept, it is not possible to speak meaningfully of even an approximate *realization* of the ideal.

81 Apel, 'Scientism or Transcendental Hermeneutics', p. 126.

82 'The crux is what happens in it [philosophy – A. W.], not a thesis or a position – the texture, not the deductive or inductive course of one-track minds. Essentially, therefore, philosophy is not expoundable. If it were, it would be superfluous; the fact that most of it can be expounded speaks against it.' Adorno, *Negative Dialectics*, pp. 33f.

83 This is also, if I understand it correctly, the basic idea of Richard Bernstein in *Beyond Objectivism and Relativism*, Oxford 1983.

84 Cf. also the literature cited in footnote 40, above.

85 Apel has formulated the principle of fundamental grounding for the normative foundations of argument as follows: 'If I cannot dispute something without contradicting myself in the immediate circumstances, and if I can also not justify it deductively without falling into a logical *petitio principii*, then it belongs to those transcendental-pragmatic presuppositions of argument which we must always have acknowledged if the language game of argument is to retain its *sense*. We can therefore call this transcendental-pragmatic mode of argument the *sense-critical form of fundamental grounding*.' (K.-O. Apel, 'The Problem of Philosophical Fundamental Grounding in the Light of a Transcendental Pragmatic of Language', in *Man and World* 18 (1975), pp. 239–75.) Although I am arguing here on the premise that unavoidable presuppositions of argument, in the sense in which Apel and Habermas speak of them, do exist, I have so far been unable to find in either Apel or Habermas a stringently *exposition* of the argument for fundamental grounding. I

believe there is a connection between this and the fact that it has so far not become clear what the unavoidable presuppositions of argument really are. Here are two examples of an allegedly 'performative' or 'pragmatic' self-contradiction which is in reality no such thing. (1) Apel asserts that the following statement contains a pragmatic self-contradiction: 'I hereby assert (= I propose as a statement which can command a universal consensus in the ideal communication community) that not all discursively justified norms – including the pragmatically practical limitations on discourse – are necessarily able to command a universal consensus.' (Apel, 'Lässt sich ethische Vernunft von strategischer Zweckrationalität unterscheiden?', in *Archivo di Filosofia*, 1983, no. 1–3, p. 424.) The assertion we are dealing with states that not all norms that are discursively justified (and thus capable of commanding a consensus) are necessarily capable of commanding a consensus. This seems to me to be an assertion of the kind, 'Not all white elephants are necessarily white.' There may well be a contradiction involved, but it is a contradiction of the simple logical-semantic type. (2) The second example comes from Habermas's work on discourse ethics, where he says:

Similarly it must be possible to demonstrate performative contradictions in the case of statements by a proponent who wished to justify the following proposition:

(3)* After excluding A, B, C, . . . from the discussion (by silencing them or forcing our own interpretations upon them), we were finally able to convince ourselves that N is correct, where the following things are true of A, B, C, . . . : (a) they belong among those who would be *affected* by the implementation of norm N, and (b) they do not differ as *participants in the argument* from the others in any relevant respect. (DE 101)

In what sense, given conditions (a) and (b), could assertion (3)* contain a contradiction? I believe that the answer is once again simple. If those who have been excluded from the discussion do not differ in any relevant respect from the other participants, then this can only mean that their arguments are just as important and worth taking seriously as those of the people who do take part in the discussion. *Suppressing* these arguments therefore means suppressing arguments which might be important for establishing the truth. What assertion (3)* is saying, therefore, is that 'we' have convinced ourselves of something by not taking any notice of some of the possibly relevant arguments. It is thus tantamount to saying

that there are possibly good arguments against the conviction we have formed, but that we shall not take any notice of them. It amounts to saying that our conviction is well-founded, but possibly not well-founded. And this, once again, appears to me to be not a *performative*, but a *logical* contradiction.

I have cited these two examples in order to make clear that everything depends on showing precisely *at which point* the fundamental grounding is really taking hold.

86 This can also be seen from the rules of discourse cited by Habermas (following Alexy), from which principle (U) is supposed to be derived. Rule 3.1 (cf. DE 99) reads as follows: 'Any subject capable of speech and action may participate in discussions.' I need not emphasize that I share the universalistic intuitions which are expressed in this rule. But it cannot be overlooked that the rule, as it is formulated, is either false or else says (relatively) little. Either the rule is saying that I am obliged to enter into a discourse with any being capable of speech and action whenever and upon whatever topic they wish, in which case the rule is quite evidently *false*. Or it is saying that no being capable of speech and action may *in principle* be excluded from discussions, in which case the rule would be far too *weak*.

87 Wolfgang Kuhlmann, *Reflexive Letztbegründung*, Munich 1985.

88 *Ibid.*, pp. 22ff.

89 *Ibid.*, pp. 196ff.

90 *Ibid.*, p. 198.

91 *Ibid.*, p. 208.

92 Cf. the discussion of the 'second objection', *ibid.*, pp. 227ff.

93 *Ibid.*, p. 189.

94 *Ibid.*, p. 190.

95 But see section XI, below.

96 Following these reflections, it is possible to illustrate the error in Apel's idea for fundamental grounding with reference to a single short passage from one of his more recent texts (K.-O. Apel, 'Lässt sich ethische Vernunft von strategischer Zweckrationalität unterscheiden?', pp. 375ff). The passage appears in the context of a critique of Kant's transcendental solipsism which, in Apel's opinion, forced Kant to proclaim the moral law as a 'fact of reason' instead of providing a *grounding* for it. 'This situation is crucially altered,' Apel says, 'if it is shown that intersubjectively valid thought itself, being bound to the medium of speech, has the structure of discourse. Through transcendental self-reflection of the "I think", it is now possible to demonstrate that, together with the structure of discourse, an (in principle unlimited) *community of finite rational beings* and the similarly generalizable *mutuality of*

- claims* (of interests or needs for which arguments may be advanced) and of the competence to evaluate arguments, in short, an ideal communication community is presupposed, which is anticipated within the real communication community. The capacity of an argument to command a consensus within the ideal, unlimited community of argument is thus acknowledged as a regulative idea of the intersubjective validity of arguments, whether these are of theoretical or of practical ethical relevance.' (Ibid., p. 421) It is immediately apparent from this passage that the supposed fundamental grounding of ethics is directly connected with the transformation of a necessary presupposition into a necessary anticipation (a necessary regulative idea), whereby the crucial point is, of course, that the sense of the presupposition itself has been misinterpreted.
- 97 This is something to which Ursula Wolf draws attention in her critique of Ernst Tugendhat: *Das Problem des moralischen Sollens*, Berlin 1984, pp. 23, 35ff. Tugendhat has assimilated this criticism and used it as a basis on which to propose a grounding of morality which comes closer to Kant; I shall take up the principal ideas of Tugendhat's proposal below. Cf. Ernst Tugendhat, *Probleme der Ethik*, Stuttgart 1984, pp. 132ff ('Retraktionen').
- 98 This is where H. L. A. Hart sees the moment of truth as conceived by the positivistic tradition to reside. Hart recognizes morality as an evaluative standard for legal norms, but rejects the reduction of the concept of legal validity to that of moral validity. 'There are thus two dangers, and it will help us to navigate between them if we insist on this distinction [i.e. the distinction between what is and what ought to be the case - A. W.]: on the one hand there is the danger that justice and its authority dissolves into what people think justice ought to be; and on the other hand there is the danger that prevailing notions of justice supplant morality in its function as the ultimate measure of behaviour, and thus elude criticism.' (Hart, *Recht und Moral*, Göttingen 1971, p. 19.)
- 99 Cf. Habermas, *Autonomy and Solidarity*, p. 171, and 'Moral und Sittlichkeit', pp. 21f.
- 100 'Moral und Sittlichkeit', pp. 21f.
- 101 Ibid.
- 102 In *Social Research*, vol. 38, no. 3 (Autumn 1971).
- 103 Habermas, *Autonomy and Solidarity*, p. 171.
- 104 The arguments that begin here follow, somewhat loosely, the thoughts of Ernst Tugendhat in *Probleme der Ethik*, Stuttgart 1984, pp. 132f.
- 105 Cf. note 17, above.
- 106 Paton, *The Moral Law*, p. 81; Ellington, p. 24. (WSB, vol. IV, pp. 42f (BA 39).)

- 107 This is the sense in which I would understand the words of Klaus Heinrich which Habermas quotes in his latest book: 'Keeping the covenant with God is the symbol of fidelity; breaking this covenant is the model of betrayal. To keep faith with God is to keep faith with life-giving Being itself - in oneself and in others. To deny it in any domain of being means breaking the covenant with God and betraying one's own foundation... Thus, betrayal of another is simultaneously betrayal of oneself; and every protest against betrayal is not just protest in one's own name, but in the name of the other at the same time.' (Klaus Heinrich, *Versuch über die Schwierigkeit nein zu sagen*, Frankfurt 1964, p. 20, quoted in Jürgen Habermas, *The Philosophical Discourse of Modernity*, Cambridge 1987, p. 325.) In elucidation of this, Habermas develops an early statement by Hegel, commenting: 'In the restlessness of the real conditions of life, there broods an ambivalence that is due to the dialectic of betrayal and avenging force.' (Ibid.) Any infringement of a communal life is visited with 'avenging force' upon the person responsible for that infringement. But we can only speak of a 'dialectic' here if we can *simultaneously* think of the avenging force as a force sublated within judgements expressible in language, i.e. as condemnation or contempt on the part of others and - in view of the inescapable intersubjectivity of such judgements - as self-condemnation or self-contempt. But we have to think of it precisely as a *force* sublated within a judgement or self-condemnation expressible in language. The fact that such a condemnation or self-condemnation possesses the power to afflict the life of the 'condemned' person shows that moral condemnation and self-condemnation still contain a reflection of real external force. It would not be possible to account for this power of moral condemnation and self-condemnation to afflict a person's life if the 'avenging force' were not merely sublated within moral judgement but had completely disappeared in it.
- 108 Cf. Habermas, *The Philosophical Discourse of Modernity*, p. 310.
- 109 Habermas, *The Theory of Communicative Action*, vol. 2, Cambridge 1987, pp. 43ff, esp. p. 52.
- 110 Ibid., pp. 77ff.
- 111 Cf. *ibid.*, pp. 62ff.
- 112 Ibid., p. 63.
- 113 Cf. *ibid.*, pp. 62ff.
- 114 Ibid., p. 63.
- 115 Ibid., p. 68.
- 116 Ibid., p. 61.
- 117 Ibid., p. 69.
- 118 Cf. *ibid.*

- 119 Ibid., p. 70.
 120 Ibid., p. 71.
 121 Ibid., p. 72.
 122 Ibid.
 123 Ibid., p. 81.
 124 Habermas, *The Theory of Communicative Action*, vol. 1, London 1984, pp. 72ff.
 125 Ibid., p. 50.
 126 For substantiating references see, for example, Bryan R. Wilson (ed.), *Rationality*, Oxford 1974.
 127 Cf. Alasdair MacIntyre, 'Rationality and the Explanation of Action', in his *Against the Self-Images of the Age*, New York 1971, p. 252: 'For when we approach the utterances and activities of an alien culture with a well-established classification of genres in our mind and ask of a given rite or practice, "Is it a piece of applied science? Or a piece of symbolic and dramatic activity? Or a piece of theology?" we may in fact be asking a set of questions to which any answer may be misleading. . . . For the utterances and practice in question may belong, as it were, to all and none of the genres that we have in mind.'
 128 Habermas, *The Theory of Communicative Action*, vol. 2, pp. 52f.
 129 Ibid., p. 55.
 130 Cf. *ibid.*, p. 49.
 131 This is also how Freud interprets the taboo, stating that 'the prohibitions of taboo are to be understood as consequences of an emotional ambivalence' (Sigmund Freud, *Complete Psychological Works*, vol. XIII, London 1955, p. 67). He also interprets 'taboo conscience' as the oldest form of (moral) conscience (*ibid.*).
 132 Mary Douglas, *Purity and Danger*, London 1966. Robin Horton, 'African Traditional Thought and Western Science', in Bryan R. Wilson (ed.), *Rationality*, pp. 131ff.
 133 Horton, 'African Traditional Thought', pp. 164–6. Cf. also Edmund Leach, *Culture and Communication*, Cambridge 1976, pp. 37ff.
 134 Philippa Foot has pointed out that it is not the categorical ought itself, but at best the way in which it is grounded that distinguishes moral rules from the rules of a club or the rules of etiquette. 'It is obvious that the normative character of moral judgement does not guarantee its reason-giving force. Moral judgements are normative, but so are judgements of manners, statements of club rules, and many others.' ('Morality as a System of Hypothetical Imperatives', in Philippa Foot, *Virtues and Vices*, Berkeley 1978, p. 162.)
 135 Habermas, *The Philosophical Discourse of Modernity*, Cambridge 1987.

- 136 Ibid., p. 345.
 137 Ibid., p. 346.
 138 Ibid.
 139 This is similar, as I have subsequently discovered, to the position adopted by William Frankena in his essay, 'Has Morality an Independent Bottom?', *The Monist*, vol. 63, no. 1 (January 1980), pp. 49ff.
 140 Cf. Martin Seel, *Die Kunst der Entzweiung. Zum Begriff der ästhetischen Rationalität*, Frankfurt 1985. Cf. also Albrecht Wellmer, 'Truth, Semblance, Reconciliation: Adorno's Aesthetic Redemption of Modernity', in this volume, pp. 22ff.
 141 Cf. Wilfrid Sellars, 'Empiricism and the Philosophy of Mind', in his *Science, Perception and Reality*, London 1963, p. 173.
 142 Cf. *ibid.*, esp. pp. 32ff. It is true that Sellars goes on to concede that the moral and thus also the social sphere possess a reality of their own. As with every philosopher of importance, it would therefore be possible to produce a critique of Sellars which largely consisted in reading him 'against the grain'. 'Thus the conceptual framework of persons is the framework in which we think of one another as sharing the community intentions which provide the ambience of principles and standards (above all those which make meaningful discourse and rationality itself possible) within which we live our own individual lives. A person can almost be defined as a being that has intentions. Thus the conceptual framework of persons is not something that needs to be reconciled with the scientific image, but rather something to be joined to it. Thus, to complete the scientific image we need to enrich it *not* with more ways of saying what is the case, but with the language of communal and individual intentions, so that by construing the actions we intend to do and the circumstances in which we intend to do them in scientific terms we *directly* relate the world as conceived by scientific theory to our purposes, and make it *our* world and no longer an alien appendage to the world in which we do our living.' (*Ibid.*, p. 40)
 143 Cf. Habermas, *The Theory of Communicative Action*, vol. 2, p. 398: 'In each of these spheres, differentiation processes are accompanied by countermovements that, under the primacy of one dominant aspect of validity, bring back in again the two aspects that were at first excluded. Thus nonobjectivist approaches to research within the human sciences bring viewpoints of moral and aesthetic critique to bear – without threatening the primacy of questions of truth; only in this way is critical social theory made possible.' What I have just said about Sellars (note 142) applies also to Habermas. What I am criticizing – and this is what I am

- aiming at in reading either author 'against the grain' – is the assumption that we have to conceive of social facts *in the first instance* in the same terms as physical facts.
- 144 Cf. Albrecht Wellmer, 'Truth, Semblance, Reconciliation', in this volume, p. 28.
- 145 Martin Seel, 'Die zwei Bedeutungen kommunikativer Rationalität. Bemerkungen zu Habermas' Kritik der pluralen Vernunft', MS (1985), p. 16.
- 146 Cf. Habermas, *The Theory of Communicative Action*, vol. 2, p. 398.
- 146 Here I am following a suggestion made by Seel. For what follows, too, cf. Seel, *Die Kunst der Entzweiung*, pp. 320ff.
- 148 Ibid.
- 149 Habermas, *The Theory of Communicative Action*, vol. 2, p. 398.
- 150 Cf. Friedrich Kambartel, 'Vernunft: Kriterium oder Kultur? Zur Definition des Vernünftigen', in F. Kambartel, *Philosophie der humanen Welt*, Frankfurt 1989.
- 151 An allusion to Kant, who speaks of a 'philosophical millennium, which hopes for a state of perpetual peace based on a league of peoples, a world-republic', as distinct from a theological millennium, 'which tarries for the completed moral improvement of the entire human race'. Kant, of course, also defends this theological millennium as a practical idea. Cf. *Religion within the Limits of Reason Alone*, pp. 29f (WSB, vol. V, pp. 682f.)
- 152 This is the point of view which is emphasized by Hannah Arendt above all; it is what her concept of 'natality' refers to: Arendt, *The Human Condition*, Chicago 1958, pp. 175ff.

Index

- absolute, art as cipher of the 48–9, 50, 85
- absolutism 116, 172–81
- action: problem of morally right 117; prohibition of 124, 127–31; *see also* communicative action
- Adams, Robert A. 54–5
- adolescence, emergence of moral consciousness 145–6
- Adorno, Theodor W., viii, 1–35, 39, 42–3, 50, 52, 71–6, 78–85, 88–90, 108, 111; aesthetic redemption of modernity 1–35; *Aesthetic Theory* 2, 3, 8–11, 12, 14–15, 34–5; *Dissonanzen* 32; 'Fragment on Music and Language' 7; *Minima Moralia* 5, 62; *Negative Dialectics* 3, 4–5, 6, 62, 71; *Philosophie der neuen Musik* 10–11, 46; philosophy of reconciliation 4–7, 11–12, 23–9, 73–4, 85, 87, 173, 177–9, *see also* *Dialectic of Enlightenment*; speech to Werkbund 102–3, 104; traditionalism of 30–4
- aesthetic discourse: and aesthetic validity-claims 229; link with moral discourse 226; structure of 27–9
- aesthetic experience, and self-transcendence of reason 76, 94
- aesthetics, Adorno's redemption of modernity 1–35
- altruism 127–8
- ambiguities 1, 37, 44–5, 107
- anarchism 70, 87, 93, 107
- animism 61
- Anti-Oedipus* (Guattari and Deleuze) 40
- antinomy 8–11
- Apel, Karl-Otto: discourse ethics 114, 116–18, 138, 139–40, 145, 195, 231; 'Scientism or Transcendental Hermeneutics?' 170–4; version of consensus theory 168–81, 188
- aporia 5–7, 63–4, 83, 210
- application, discourse of 204–7
- architecture: eclecticism and historicism in 43–5, 99–100, 106–7; language of 43–5, 103–6; modern functionalism 98–102; postmodern viii, 40, 103–7