

Practical Reason

G. H. VON WRIGHT

Philosophical Papers
Volume I

© G.H. von Wright 1983

First published 1983
Basil Blackwell Publisher Limited
108 Cowley Road, Oxford OX4 1JF, England



British Library Cataloguing in Publication Data

Wright, G.H. von
Practical reason.
Vol. 1
1. Deontic logic
I. Title
160 BC145

ISBN 0-631-13315-1

Typeset by MHL Typesetting Limited, Coventry
Printed in Great Britain by T.J. Press Ltd, Padstow

Contents

Introduction	vii
Acknowledgements	x
Practical Inference	1
On So-called Practical Inference	18
Determinism and the Study of Man	35
Explanation and Understanding of Action	53
The Foundation of Norms and Normative Statements	67
On Promises	83
On the Logic of Norms and Actions	100
1. Deontic logic as modal logic—analogies and disanalogies	100
2. On action sentences and the logic of action	107
3. Deontic logic—a new approach	119
Norms, Truth, and Logic	130
1. Deontic logic as a logic of norm-satisfaction	130
2. Conditional norms and hypothetical imperatives	150
3. Norms with temporal quantifiers	159
4. Norms for individual occasions	164
5. Action logic as a basis for deontic logic	169
6. Note on omission and on higher-order action	193
7. Quantifiers in deontic logic: <i>Sein-Sollen</i> and <i>Tun-Sollen</i>	196
Subject Index	210
Index of Persons	214

Introduction

The first six studies published in this collection have their point of departure in my book *The Varieties of Goodness*, the remaining two in *Norm and Action*. Both works were based on my Gifford Lectures in the University of St Andrews and published by Routledge & Kegan Paul, London, in 1963. The first had no ancestor worth mention in my earlier literary output. The second had grown out of my preoccupation, since 1951, with "deontic logic" which was in origin a by-product of my inquiries into the logic of modal concepts.

The two works mentioned, although simultaneously written and published, are different in method and spirit. My contributions to the logic of actions and norms are straightforward "philosophical logic" in the sense that they use methods and techniques of formal logic for the purpose of clarifying the structure of actionist and normative discourse. The discussion of the same kind of discourse in *The Varieties of Goodness* can perhaps be labelled "logical analysis" but it is decidedly not "formal". I always had the feeling that formal methods simply are not applicable to this second approach. I mention this because others who are working in the same field may not share with me the impression of opposition and maybe irreconcilability between the two types of approach.

At the time of writing *The Varieties of Goodness* I had been much influenced by Elizabeth Anscombe's book *Intention* and her revival of the Aristotelian theme of "practical reasoning (inference, syllogism)". Anscombe also drew my attention to Charles Taylor's book *The Explanation of Behavior* (1965). It impressed me deeply. After reading it, my interest in practical inference became an interest in action-explanation—and this interest led me to consider the relationship between the (predominantly) "nomothetic" natural sciences and the (typically) "hermeneutic" sciences of man. The outcome was a book, *Explanation and Understanding* (Cornell University Press, Ithaca, N.Y./Routledge & Kegan Paul, London, 1971) in which I tried to give a new twist to a classic debate in the theory and philosophy of science, the so-called *Erklären—Verstehen* controversy.

The type of practical inference which I have studied can, in rough outline, be described as follows: An action is being related as "conclusion" to an aim or end of a given agent and an opinion of his concerning the means to its attainment as "premisses". When the action is a *fait accompli* the inference schema provides an explanation of the action. The structure of the schema and the nature of the connection between premisses and conclusion in it is, however, very much of a problem.

dissatisfaction with my earlier attempts to deal with practical reason urged me to return to the topic time and time again. The first attempt—simultaneous with the publication of *The Varieties of Goodness*—and three “returns” to the subject are published in this collection. They testify, I hope, to some progress towards a better understanding of the subject-matter—but they also represent essential shifts in the angle under which the problems are being attacked. These changes seemed to me important enough to justify republication of the papers in spite of minor inconsistencies between the author’s successive positions.

In *Explanation and Understanding* (here echoed in the essay “On so-called Practical Inference”) I claimed that the practical inference pattern holds a position in the human and social sciences similar to that of the deductive—nomological inference pattern (“the covering law model”) in the natural sciences. This is an exaggeration. Far from all explanation in the human sciences is explanation of *action*—either individual or collective. And far from all explanation of action conforms to a practical inference pattern. Moreover, though action is, primarily, *intentional* behaviour not all forms of intentional action spring from the *intentions* of agents. The much wider range of the intentional has only gradually become clear to me. Yet the type of teleological schema which I have discussed under the heading “practical inference”, linking intentions with actions, still seems to me to hold a pivotal position in the field of action explanation. Of this I have tried to give a fuller and more balanced picture in the essays “Determinism and the Study of Man” and “Explanation and Understanding of Action”, printed in this volume, and also in a monograph with the title *Freedom and Determination* (North-Holland Publishing Company, Amsterdam, 1980).

The two papers “The Foundation of Norms and Normative Statements” and “On Promises” also belong in the orbit of *The Varieties of Goodness*. They reflect efforts to show how normative requirements may be given a foundation (justification) in the practical necessities under given ends or basic valuations. I think it is of the essence of norms (norm-giving activity) that there is a connection between what might be called the “deontic” and the “technical” aspects of norms—or between genuine (categorical or conditional) norms and that which Kant and others have called “hypothetical imperatives”.

This connection does not mean reduction of the deontic “ought” to an anankastic “must”. Nor does it establish a one-to-one correlation between norms and means—end relationships. The connection is reflected in phenomena of a more “global” nature which might be described as an “*aura of normative pressure*” surrounding an efficacious code (order, system) of norms. (See below, pp. 39 and 55.)

Ever since the appearance of my first paper on deontic logic in *Mind* in 1951 I felt that there was some philosophically essential aspect of norms (normative concepts and discourse) which the formal system I had con-

structed either did not capture at all or tried to capture in the wrong way. In the 30 years which have passed I have again and again returned to the topic—often with a new idea which I thought would at last put things essentially right. But always, so far, to be disappointed. When writing a survey paper¹ on the subject for a conference in Rome in 1977 I had again some ideas leading in a hitherto unexplored direction which seemed to me promising. They eventually found embodiment in a longish study “On the Logic of Norms and Actions”, first published in an anthology *New Studies in Deontic Logic* (ed. by Risto Hilpinen; D. Reidel Publishing Co., Dordrecht, Holland, 1981) and republished here with minor revisions. A feature of this new approach is a dualistic conception of the basic deontic notions. Under the one conception these notions are *operators* on some proposition-like entities; under the other they are *predicates* of (actions regarded as) logical individuals. The two conceptions supplement and depend logically on each other. This approach here stands side by side with another, rather different one, which is attempted in the concluding study of the present volume. How the two approaches are related and what may be fruitful about each of them will have to be disclosed by future research.

For my part, I regard my passage through the wilderness of deontic logic as terminated. I hope the feeling I now have will last, that the new essay “Norms, Truth and Logic” has eventually removed the uneasiness I felt about advancing with the instruments of logic beyond the frontiers of truth and falsehood. The approach in the concluding essay is, incidentally, not in a, for me, new direction but is rather a continuation and refinement of the (formal) approach taken in *Norm and Action*. If I am not mistaken this work has had a certain amount of influence on recent action and norm theory; but it has surprised me that what in it was a *new* approach to deontic logic has seemingly gone unnoted by later research.

Georg Henrik von Wright

¹ “Problems and Prospects of Deontic Logic—A Survey”, in *Modern Logic—A Survey*, ed. by E. Agazzi; D. Reidel Publishing Co., Dordrecht, Holland, 1980.

Acknowledgements

Those of the papers in this volume which have been previously published originally appeared as follows:

"Practical Inference", *The Philosophical Review*, 72 (1963), 159–79.

"On so-called Practical Inference", *Acta Sociologica*, 15 (1972), 39–53.

"Determinism and the Study of Man", in *Essays on Explanation and Understanding*, ed. by J. Manninen and R. Tuomela, © 1976; D. Reidel Publishing Company, Dordrecht, Holland, pp. 415–35.

"Explanation and Understanding of Action", *Revue Internationale de Philosophie*, 35 (1981), 127–42.

"The Foundation of Norms and Normative Statements", in *The Foundations of Statements and Decisions*, ed. by K. Ajdukiewicz; Polish Scientific Publishers, Warszawa, 1965, pp. 351–67.

"On Promises", *Theoria*, 28 (1962), 276–97.

"On the Logic of Norms and Actions", in *New Studies in Deontic Logic*, ed. by R. Hilpinen, © 1981; D. Reidel Publishing Company, Dordrecht, Holland, pp. 3–35.

A preliminary version of the first part of the essay "Norms, Truth, and Logic" appeared in *Deontic Logic, Computational Linguistics and Legal Information Systems*, ed. by A.A. Martino, North-Holland Publishing Company, Amsterdam, 1982.

The papers are here republished with the kind permission of the copyright-holders.

Practical Inference

I

In this paper I shall be dealing with a type of logical argument, which I propose to call *practical inference*.

Aristotle distinguished between theoretical and practical syllogisms. His treatment of the latter is very scanty and unsystematic. The examples, of which he gives hints without elaborating them in detail, are a rather mixed bunch. Some of his general remarks on practical syllogisms, however, are of great interest. They show that Aristotle was aware of the peculiar character of a type of reasoning which logicians after him have tended either to ignore or to misrepresent.

The nearest Aristotle comes to giving a full example of a practical syllogism is in the third chapter of the seventh book of the *Ethica Nicomachea*. The two premisses are "All sweet things ought to be tasted" and "That thing is sweet." Then, instead of stating the conclusion in words—"That thing ought to be tasted"—Aristotle goes on to say that you are bound, if able and not prevented, to taste the thing immediately. The practical syllogism thus leads up to or ends in action. "When the two premisses are combined, just as in theoretical reasoning the mind is compelled to affirm the resulting conclusion, so in the case of practical premisses you are forced at once to *do it*", he says.¹ And in another place Aristotle calls the conclusion of a practical syllogism an *action*.²

Aristotle seems throughout to be thinking of practical inference in terms of the subsumption of an individual act under a general rule of action by the intermediary of a particular fact-stating premiss.³ I shall not in the present paper be dealing, directly, with this type of inference.

The type of argument which we shall here be primarily studying is concerned with (necessary) means to an end. Although the notions of means and end are prominent in Aristotle's ethics, Aristotle seems not to have had this type of argument in mind when speaking of practical syllogisms. Yet the chief peculiarity, as I see it, of Aristotle's practical syllogisms—namely, their relation to *action*—is characteristic also of practical inference of the type which is here studied. It is this common feature which justifies us in calling both types of argument "practical" and in contrasting them with various types of "theoretical" reasoning.

¹ *Ethica Nicomachea*, 1147a 26–30. Quoted from the Loeb Classical Library translation by H. Rackham.

² *De Motu Animalium*, 701a 12–14.

³ Cf. *Ethica Nicomachea*, 1147a 25–6.

Consider the following inference:

- (1) One wants to make the hut habitable.
 Unless the hut is heated, it will not become habitable.
 Therefore the hut must be heated.

The first premiss is a want statement. The thing wanted is that the hut be habitable. This I shall call an *end*. The end, moreover, is an end of *action*. This means that we want to attain the end as a result or consequence of something which we do. That the end has this character is indicated in our example by the use of the phrase "one wants to make".

The second premiss may be said to rest upon a causal relationship. This is a relationship between temperature and habitability (of a hut). That there should be this relation is a causal fact about the living conditions of men. Essential to the above inference is the assumption that the temperature does not rise "of itself", that is independently of human interference with "the course of nature". Something has to be done in order to make the temperature rise. "Heating" is a word for (actively) raising the temperature. A rise in temperature may serve (favour, further) the ends of a person, as a favourable wind may further the ends of a sailor. But we do not ordinarily call it a "means" to an end. Raising the temperature, however—that is, producing a rise through action—may quite appropriately be called a means to some end. The action which is mentioned in the second premiss is thus a means to the end which is mentioned in the first premiss.

Of the conclusion I shall say that it expresses a *practical necessity*, namely the practical necessity of using the means mentioned in the second premiss in order to attain the end mentioned in the first premiss. I deliberately use the word "must" in the conclusion and not the word "ought". (Instead of "must" we could also say "has to".) Ordinary usage does not maintain a sharp distinction between the meanings of "must" and "ought". But it may be said to hint at distinctions, which the logician has reason to observe. "Must" is somehow stronger than "ought". To say "I ought to do this, but I am not going to do it" can make sense. To say "I must do this, but I am not going to do it" has the ring of a contradiction.

The inference (1) may be regarded as an instantiation of the following general pattern of inference:

- (2) One wants to attain x .
 Unless y is done, x will not be attained.
 Therefore y must be done.

I shall call an inference of this form a *primary* practical inference. There are a number of related patterns, which will also be called primary; compare

(8), (11), and (24) below. I shall not here attempt a systematic description and study of all such patterns.

Is an inference of the above form *logically conclusive*?

I think that many logicians and philosophers would answer this question in the negative. Some of them would perhaps support their view with the following argument: The two premisses of the inference are *descriptive*. They state what is the case. The conclusion, however, is *normative* or *prescriptive*. And, it is said, one cannot draw a normative conclusion from (only) factual premisses.

The normative "look" of the conclusion is even more striking if one uses "ought" here instead of "must". It is a well-known and widely accepted idea that one cannot deduce an "ought" from an "is". I do not deny that there is an important truth hidden in this idea. But I would deny that it constitutes a counterargument to the logically conclusive character of the pattern of inference which we are here studying.

III

We have so far formulated in only an impersonal way the pattern of inference which we are discussing, using such phrases as "one wants to attain" and "must be done". An end, however, is necessarily somebody's end. And an act is necessarily done by some agent or agents.

The impersonal formulation of the inference may be said to cover a number of personal formulations of it. I shall here distinguish between two such personal formulations. I shall refer to them as the *third-person* and the *first-person* pattern of inference.

An example of an inference in the third person would be:

- (3) *A* wants to make the hut habitable.
 Unless *A* heats the hut, it will not become habitable.
 Therefore *A* must heat the hut.

A here stands for the name of some agent (person). The corresponding inference in the first person would run:

- (4) I want to make the hut habitable.
 Unless I heat the hut, it will not become habitable.
 Therefore I must heat the hut.

The question may be raised whether the person who figures in the premisses and the conclusion must be the same throughout the whole argument. Consider the following pattern:

- A* wants to make the hut (have the hut made) habitable.
 Unless *B* heats the hut, it will not become habitable.
 Therefore *B* must heat the hut.

This, obviously, is not a logically conclusive argument. But what shall we say of the following?

A wants to make the hut habitable.

Unless *B* heats the hut, it will not become habitable.

Therefore *A* must make *B* heat the hut.

This last is analogous to the following inference:

A wants to make the hut habitable.

Unless the temperature rises, the hut will not become habitable.

Therefore *A* must make the temperature rise.

Assume that the temperature rises "of itself," that is independently of anything which *A* does. Then it was not necessary for *A* to *make* the temperature rise. Similarly, it may happen that *B* heats the hut "on his own initiative", that is independently of any steps which *A* takes. Then it was not necessary for *A* to *make* *B* heat the hut (or to see to it that *B* heats the hut). We shall not here stop to discuss the question whether a person can be correctly said to have attained his end of action if the wanted thing comes about independently of anything he does. But we may note that the inference with the conclusion that *A* must make *B* heat the hut is not valid in the form which we gave it. It becomes valid, however, if we amend it as follows:

(5) *A* wants to make the hut habitable.

Unless *A* makes *B* heat the hut, it will not become habitable.

Therefore *A* must make *B* heat the hut.

The answer to our above question of identity of persons is that the agent who is in pursuit of the end and the agent upon whom the practical necessity is incumbent must be the same. But, as also seen from the case which we were discussing, it is quite possible that another person beside the one whose end and practical necessity of action are concerned may be "involved" in the inference. We shall presently return to the discussion of some such cases. First, however, we must settle the question which was raised at the end of Section II, namely that of the logical conclusiveness of primary practical inferences.

IV

There is an important difference with regard to the second premiss to be noted between the inference in the third person and the inference in the first person. This difference is relevant to the problem of logical conclusiveness.

A can be ignorant of the causal relationship between his end of action and various necessary means towards its attainment. *A* wants to make the hut habitable. It is a fact that the hut will not become habitable unless *A*

heats it. *A*, however, does not realize this. Perhaps he thinks he can make the house habitable by some different means, for example by repairing a broken window.

It seems obvious that the inference nevertheless is valid *of him* (of his case). Its validity is independent of whether *A* knows or believes anything at all about the causal relationship. Thus its validity is also independent of whether *A* himself acknowledges the practical necessity for him of acting in a certain manner.

But is this true? Is it really the case that, *if A* wants to make the hut habitable and *if* the hut will not become habitable, unless he heats it, *then A must heat the hut*—even though he is unaware of the necessity?

The answer depends upon how we interpret the “must”. If we understand the phrase “*A* must heat the hut” to mean the same as “unless *A* heats the hut, he will fail to attain some end of his action” or to mean the same as “there is something *A* wants but will not get, unless he heats the hut”, then the answer is affirmative. For the following, undeniably, is a logically conclusive argument:

(6) *A* wants to make the hut habitable.

Unless *A* heats the hut, it will not become habitable.

Therefore there is something *A* wants but will not get, unless he heats the hut.

Or, to put it in more general terms, this is logically conclusive:

(7) *A* wants to attain *x*.

Unless *A* does *y*, he will not attain *x*.

Therefore there is something *A* wants but will not get, unless he does *y*.

Is the interpretation plausible which we have given to the “must” in the conclusion of a practical inference in the third person? It seems to me that it is. Suppose we were asked what we mean by saying of a person that he *must* do a certain thing. We should sometimes answer, I think, that we mean that, unless he does this thing, he will not be successful in the pursuit of some end of his. We sometimes dispute whether a person must, or need not, do a certain thing, in this very sense of “must”. (But there are other senses of “must” beside this one.)

On the view which I have here suggested, it is logically or necessarily true that, if *A* wants to attain *x* and will not attain it unless he does *y*, then he *must* do *y*. But it does not follow that he *will* do *y*.

The logical peculiarity, noted by Aristotle, that practical inference leads to action is thus not a characteristic of such inference in the third person. This peculiarity, as we shall show below, belongs to the first-person case. One could mark the fundamental difference between the two cases by saying that only practical inference in the first person is truly

"practical", whereas the argument in the third person is actually "theoretical". One could also use the terms "subjective" and "objective" to stress the difference. I shall do this and say that the conclusion of a practical inference in the third person states an *objective* practical necessity.

V

An inference in the first person is necessarily conducted or performed by the same subject, of whom (of whose case) it is valid. As a consequence of this, the subject will necessarily know or believe the truth of the second premiss.

Is it (logically) possible that a man should *want* to attain a certain end and should *understand* (know or believe) that a certain action on his part is necessary if he is to attain his end, and yet should not act accordingly?

It is of course possible that a person who wants to attain *x* may *revise* his end, when he comes to understand that unless he does *y* he will not attain *x*. He no longer wants to attain *x*. The two premisses fail to *combine*, as it were, to form a practical inference. The inference, one could also say, fails to *materialize*. (Compare Section IX below.) Then the subject does not act according to "the dictate of practical necessity", because there is no such necessity incumbent on him.

It is also possible that a person wants to attain *x* and understands that unless he does *y* he will not attain *x*—and then *tries* to do *y*, but fails or is prevented. Did he then "act accordingly"? The answer depends upon how we interpret the phrase. If to act accordingly means to accomplish the necessary feat, he did not act accordingly. If, however, to act accordingly is to set oneself to do (try to do, go on to do) the necessary act, then he did act accordingly.

I shall here understand "act accordingly" in the second of the two senses. And I shall maintain that, if a man can truly be said both to want to attain a certain end and to combine with this will a knowledge or belief that unless he does a certain act, he will not attain his end, then he will act accordingly.

Instead of saying "he will act" I could also have said "he will necessarily act". This, moreover, is *logical* necessity. For, if action does not follow, we should have to describe the subject's case by saying either that he did not in fact *want*⁴ his professed object of desire or did not, after all, *think it necessary* to do the act in order to get the wanted thing.

⁴ The English word "want" is both ambiguous and vague. (Cf. below p. 9, fn. 7.) One could make a distinction between wanting something *actively* which entails "going after" the thing in question and wanting something *passively*—which means merely wishing for this thing to come. Active wanting, moreover, presupposes an intention or resolve on the part of the agent to do something which he thinks is in his power to accomplish. (1983)

A practical inference in the first person thus necessarily leads up to or terminates in action. Wanting the end and understanding the causal requirements for attaining it put the subject in motion. One could also say that want is what *moves* and understanding (of causal connections) is what *steers* the movement. The two together *determine* the subject's course of action.

I shall call the determination of action through a practical inference in the first person *subjective* practical necessity. It seems to me that Aristotle must have been thinking of this peculiar kind of necessitation when he insisted that the conclusion of a practical syllogism is an act.

Subjective practical necessity shows that there is a sense in which human action can be, at the same time, voluntary and strictly determined.

VI

Sometimes action must follow immediately, or else the subject will not attain his end. Sometimes, however, action need not take place immediately, but perhaps only by tomorrow afternoon or within a week from now, or "sooner or later".

When action need not follow immediately, the subject may revise his end in the meantime—not only under the impression of his insight into the necessary requirements, but also for other reasons. He may come to aspire after different things. It may also happen that he comes to think he was mistaken in believing that he will not get *x* if he does not do *y*. But unless he changes his attitude to either the first or the second premiss of the argument, he is bound to act in accordance with these premisses. In this sense his future action is (pre)determined by his present want and insight.

Would it not, someone may suggest, in view of the fact that the (necessary) action may become delayed, be more to the point to say that a practical inference in the first person terminates in a *decision* or *resolve* to act than to say that it terminates in *action*? It does not seem to me that this would be more to the point.

A decision is normally the outcome of deliberation. There are at least two entirely different senses in which deliberation may be said to concern the means to a given end. A man can deliberate as to what the means to the end *are*. Or he can deliberate which means to the end he should *choose*—assuming that there are several means productive of the end without being (individually) necessary for its attainment. In the first case the outcome of deliberation is not a decision to act. In the second case the outcome of deliberation may be a decision to act (in a certain way). But the course of action which has been decided upon is not a practical necessity, since it is but one of several possibilities. When action is a practical necessity, there is no room for choice.

VII

A few words should be said about the case in which the subject cannot perform the necessary act, the reason being that he has not learned or does not know how to do the thing in question. Then, necessarily, he will not do the thing.

If the subject does not know that he cannot do *y*, he may try to do it and find out that he cannot. Then his practical inference ended in action. This case does not differ in principle from the one which we have been discussing.

If, however, the subject *knows* that he cannot do *y*, then, I would say, he cannot correctly apply to himself the phrase "unless (now) I do *y*".⁵ He will then have to modify his argument as follows:

(8) I want to attain *x*.

Unless I learn (how) to do *y*, I shall not attain *x*.

Therefore I must learn to do *y*.

This is, in fact, how we often infer when we aspire after things which are beyond our present abilities of action. And having thus inferred, we take some steps to acquire the needed ability—unless at once we relinquish the end.

VIII

It is important to see *how* different the inference in the first person is from the inference in the third person. The difference is concealed, not only by the use of the impersonal inference schema (1) to cover the two personal inference patterns (3) and (4), but also by our linguistic presentation of the two last patterns themselves. Let the question be raised: what *are* (is the nature of) the premisses and the conclusion of the two types of inference? They are not sentences which stand printed on paper. In the case of the inference in the third person the correct answer, I should think, is this: the premisses are the *propositions* that a certain person pursues a certain end of action and that a certain thing is a necessary means to this end. The conclusion is a third proposition, namely that the person will fail to reach some end of his action unless he does this thing. In the case of the inference in the first person the correct answer seems to be this: the premisses are a person's *want* and his *state of knowing or believing* a certain condition to be necessary for the fulfilment of that want. The conclusion

⁵ A thing which is not *now* possible cannot *now* be a necessary condition of anything. To say "Unless I do *y* (which, however, I cannot do), I shall not attain *x*" is not to state a necessary condition of my attaining *x*. But to say "Unless in future I do *y* (which at present I cannot do), I shall never attain *x*" is to state a necessary condition of something—even if, in fact, I never learn to do *y*.

is an *act*, something that this person does. Wants, states of knowing or believing, and acts are not only mutually rather different from each other. They are all of them entities of a radically different sort from propositions. It is of the essence of propositions that they are expressed by sentences. (Some philosophers would say that a proposition is the meaning or the sense of a type of sentence.) Wants, states of knowing or believing, and acts have no analogous essential connection with *language*. Therefore the relation to language of a practical inference in the first person is in principle different from the relation to language of a practical inference in the third person.⁶

IX

An end of action is something, which one can be said to *want to attain*. Kant, in a celebrated passage, expressed the opinion that whoever wants to attain an end is bound also to want to use the necessary means towards its attainment.⁷ Kant thought, moreover, that this was an analytic principle of the will. Properly interpreted the principle is, I think, analytic. But then it is important to see what its "proper interpretation" is.

It is plausible to think that, if one wants to attain an end and realizes the necessity of doing a certain thing in order to attain that end, then one also wants to do that thing *for the sake of* (attaining) *the end*. But normally one does not want to do this thing *for its own sake*. It may, moreover, happen that this thing is something which one shuns doing or which one does reluctantly or which one would rather leave undone, if it were not necessary to do it for the sake of something else. When this is the case, our will to attain the end can be said to fight against our unwillingness to use the necessary means.

There are several possible outcomes of such a struggle. One is a "compromise". One sets oneself to do the thing, but half-heartedly. As a

⁶ I am indebted to Mr Lars Fröström of Lund University for his having drawn my attention to this difference between the two patterns of inference.

⁷ *Grundlegung zur Metaphysik der Sitten* (2nd ed.), pp. 44–5: "Wer den Zweck will, will (sofern die Vernunft auf seine Handlungen entscheidenden Einfluss hat,) auch das dazu unentbehrlich notwendige Mittel, das in seiner Gewalt ist." In the translation by H. J. Paton (*The Moral Law*, Hutchinson, London, 1956, pp. 84–5): "Who wills the end, wills (so far as reason has decisive influence on his actions) also the means which are indispensably necessary and in his power." The peculiarities of meaning of the words "want" and "will" in English make adequate translation difficult. To say that who *wills* an end *wills* also the necessary means to it sounds to me somewhat artificial and calls for a "translation" into more colloquial English. To say that who *wants* an end also *wants* the necessary means to it suggests a meaning which makes the statement appear plainly false. But to say that whoever *wants to attain* an end *wants also to use* the necessary means to its attainment seems unobjectionable from the point of view of language and not wholly implausible from the point of view of truth.

consequence of this, one perhaps fails to accomplish the act. One could then say that the will which, in accordance with Kant's principle, has been "transferred" from the end to the necessary means is weakened by our ("natural") impulse against the use of these means. This is one aspect, among many, of the phenomenon known in moral philosophy under the name of *akrasia* or weakness of will.

Another possible outcome of the struggle is that the unwillingness to use the means gains the upper hand, conquers the will to attain the end. Then, I think, can we no longer be said to *want* to attain it. The practical inference "collapses". We may still wish that the thing which we wanted to attain as a result or consequence of our action will happen, "of itself" or thanks to the action of some other agent. We should welcome it if it were to happen. Perhaps we can still be said to want it to happen. But we can no longer be said to want to *make* it happen.

A third possibility is that the will to attain the end conquers the impulse against the use of the means without "compromising" with it. Of such cases we sometimes say that the subject *forces* or *compels* himself to action. We can also say that the subject then *makes himself do* the action.

The case in which a subject makes himself do a certain thing by thwarting an impulse to the contrary exhibits an interesting analogy to the case in which a person commands or orders another person to do something. To command others is "heteronomous" commanding. To make oneself do something may, by analogy, be called "autonomous" commanding. Autonomous commanding does not reflect the practical necessity, as such, of action. It reflects the practical necessity of making oneself act.

I think we are here in the neighbourhood of the Kantian notion of an autonomous command (norm, rule). Perhaps one could say that we have arrived at the Kantian concept, stripped of its moral connotations. But then it should be remembered that, on Kant's view, a conceptual separation of the notion of autonomy from that of moral duty is not possible.⁸

X

Consider once again the inference mentioned at the end of Section III:

⁸ It seems to me that there are (at least) two entirely different notions of "autonomy" involved in Kant's discussion of the moral law. The first notion is that of a man commanding (giving laws unto) *himself*. This notion corresponds fairly closely to our notion here of autonomous commanding. The second notion is that of a man's will being affected by any (subjective) *ends* of his action. On Kant's view that which I here call practical necessity of action is not an instance of autonomy at all, but of heteronomy (although not of heteronomous commanding). The two notions of autonomy should be kept apart. The second is closely related to the Kantian idea of "duty for duty's sake".

(5) *A* wants to make the hut habitable.

Unless *A* makes *B* heat the hut, it will not become habitable.

Therefore *A* must make *B* heat the hut.

Assume that *A* himself conducts this argument in the first person. The conclusion will then be that he sets himself to make *B* heat the hut. How does *A* do this? There are many ways of making or trying to make people do things: asking them, telling them, requesting them, commanding them, persuading them; threatening them with some evil, if they do not comply with our wish; holding forth some reward or promising a service in return if they comply.

The above inference rests on the assumption that *B* does not heat the hut on his own initiative, but has to be *moved* to do this. *A* has to *make* him heat the hut, *see to it* that he does this.

Assume next that *A* can heat the hut himself and that there is in fact nobody who could do it in his place. Then it is necessary for *A* to do the heating himself. Assume, however, that he does not do the job willingly, but has to conquer his laziness or reluctance. He can then, as it were, "identify" himself with *B* in the above example and conduct the following argument about his own case:

(9) I want to make the hut habitable.

Unless I make myself heat the hut, it will not become habitable.

Therefore I must make myself heat the hut.

How does one make oneself do a thing? And how does making oneself do a thing differ from simply doing it? (One does not very often use the phrase "make oneself do".) Literally, I would say, to make oneself do a thing is to set oneself to do it in spite of an impulse to the contrary, for example, under the impact of a practical necessity. One cannot literally (in a primary sense) command or persuade or promise or threaten oneself to do any of the acts whereby one makes or tries to make others do things. But in an analogical or secondary sense, making oneself do a thing (by conquering a contrary impulse) may sometimes aptly be called commanding oneself (and sometimes perhaps rather persuading or threatening oneself). It is aptly called commanding when we think of the movement to action and the conquest of the contrary impulse as a victory of reason over blind passion or of our rational self over our animal nature.

XI

Compare these two patterns of inference:

(10) *A* wants to attain *x*.

Unless *A* does *y*, he will not attain *x*.

Therefore *A* must do *y*.

and:

- (11) *A* wants to attain *x*.
 Unless *A* makes *B* do *y*, he will not attain *x*.
 Therefore *A* must make *B* do *y*.

The conclusions of these two inferences are *logically independent* of one another. It can be true both that *A* must do *y* (himself) and that he must make *B* do it (as well). Or one of the propositions may be true and the other false. Or they may both be false. (The case when they are both true has peculiar problems of its own, which will not be discussed here.)

(10) is the pattern of a primary practical inference. (Compare Sections II and III above.) (11) I shall also call a pattern of a primary practical inference (though of a more complex form). (11) can be obtained from (10) if we substitute "makes *B* do *y*" for "does *y*" in the second premiss and "make *B* do *y*" for "do *y*" in the conclusion.

Compare next (10) with the schema:

- (12) *A* wants to attain *x*.
 Unless *A* makes himself do *y*, he will not attain *x*.
 Therefore *A* must make himself do *y*.

The conclusions of (10) and (12) are not logically independent. The conclusion of the second argument entails the conclusion of the first. For if it is true that an agent will not attain a certain end of his unless he makes himself do a certain act (which he is reluctant to do), then it is also true that he will not attain this end unless he does that act.

An inference with the conclusion "*A* must make himself do *y*" therefore *presupposes* an inference with the conclusion "*A* must do *y*". An agent can be under a practical necessity to make himself do something (to compel or force himself to do something against inclination) only when he is (already) under a practical necessity to do that same thing. For this reason I shall say that the inference pattern (12) above is secondary to the pattern (10).

XII

It follows from what has been said in Section XI that, if the premisses of a practical inference which ends in the conclusion "*A* must make himself do *y*" are affirmed, then the premisses of an inference of the following form are also affirmed:

- (13) *A* must do *y*.
 Unless *A* makes himself do *y*, he will not do *y*.
 Therefore *A* must make himself do *y*.

This pattern may be regarded as a derivative case of a more general pattern:

- (14) *A* must do *x*.
 Unless *A* does *y*, he cannot do *x*.
 Therefore *A* must do *y*.

(13) is obtained from (14) through substitution and by weakening the second premiss of (14) to "Unless *A* does *y*, he will not do *x*". If it is true that unless *A* does *y* he *cannot* do *x*, then it is also true that unless *A* does *y* he *will not* do *x*.

I shall call (14) the pattern of a secondary practical inference (in the third person). The secondary practical inference has its "starting point" (first premiss) in a practical necessity, and it ends in a further practical necessity. The logical conclusiveness of a secondary practical inference can be seen from the following expansion of its pattern:

- (15) There is something *A* wants but will not get unless he does *x*.
 Unless *A* does *y*, he cannot do *x*.
 Therefore there is something *A* wants but will not get unless he does *y*.

XIII

I shall next discuss some patterns of practical inference with "ought". Consider the following example:

- (16) The hut ought to be made habitable.
 Unless the hut is heated, it cannot be made habitable.
 Therefore the hut ought to be heated.

This impersonal formulation of the inference may be said to cover a third-person and a first-person variant of it:

- (17) *A* ought to make the hut habitable.
 Unless *A* heats the hut, he cannot make it habitable.
 Therefore *A* ought to heat the hut.

and:

- (18) I ought to make the hut habitable.
 Unless I heat the hut, I cannot make it habitable.
 Therefore I ought to heat the hut.

Are these arguments logically conclusive? Someone may think that the question of their conclusiveness is less problematic than the question of the conclusiveness of primary practical inferences. For the above inference schemas contain an ought-sentence among the premisses. Thus the idea that one cannot draw normative conclusions from factual premisses cannot be adduced against the logical conclusiveness of the inferences. I do not know how much weight should be attached to this argumentation. It

seems to me that the question of logical conclusiveness is more, rather than less, problematic in the case of the inference patterns (16) through (18) than it is in the case of inferences from ends and causal connections to practical necessities of action.

The question of conclusiveness essentially depends upon how we understand the crucial word "ought" in the context. "Ought" is commonly used when we talk about duties or obligations (legal or moral, as the case may be). Duties and obligations are not, by themselves, practical necessities of action. Sometimes, however, doing one's duty or fulfilling one's obligation may become a practical necessity.

Let "ought" be translated by "it is duty", and consider the following inference in the third person:

- (19) It is *A*'s duty to make the hut habitable.
 Unless *A* heats the hut, he cannot make it habitable.
 Therefore it is *A*'s duty to heat the hut.

The duty mentioned in the conclusion I shall call a *derived duty* or a *derived obligation*. And a derived obligation I propose to define (interpret, understand) in the following way: a derived obligation is something which *must* be done in order that we may be able to fulfil another (derived or primary) obligation. A derived obligation, one could also say, is something which *must* be done for the sake of doing one's duty. The "must" here is the "must" of practical necessity. The derived obligation to do a certain thing is a practical necessity of action, *relative* to some other obligation.

Under this interpretation of the duty mentioned in the conclusion of (19), the above schema can also be cast in the following form:

- (20) It is *A*'s duty to make the hut habitable.
 Unless *A* heats the hut, he cannot make it habitable.
 Therefore *A* must heat the hut.

Is this a logically conclusive argument? The answer is that it is conclusive, if we take "*A* must heat the hut" to mean the same as "unless *A* heats the hut, he cannot do (what is) his duty". For the following, indubitably, is conclusive:

- (21) It is *A*'s duty to make the hut habitable.
 Unless *A* heats the hut, he cannot make it habitable.
 Therefore, unless *A* heats the hut, he cannot do his duty.

(The first and the second "cannot", or the second alone, may become weakened into "will not".)

Derived obligations, in the above sense of the term, play a prominent role in the lives of most people. The law prescribes what our legal obligations are. But it leaves to the individual citizen to consider the various measures and steps which he must take in order to fulfil his legal obliga-

tions. This is convenient, since the nature of these necessary steps may vary with the particular circumstances of the individuals concerned. For example: I am under a legal obligation to make an income tax return to the proper authorities. In order to fulfil this obligation I must do various other things, of which the law, however, says nothing. I must keep some records of my earnings, I must procure for myself the needed forms to be filled in, ascertain to whom and before what date they should be returned, and return them. These are "derived obligations", practical necessities of action relative to "primary obligations".

XIV

Consider an inference in the first person which terminates in a derived obligation:

- (22) It is my duty to do *x*.
 Unless I do *y*, I cannot do *x*.
 Therefore I must do *y*.

This sort of argument can be, and very often is, I think, no more than reasoning about one's own case. A man can be aware of the fact that something or other is his duty, realize that, unless he does a certain thing, he cannot do his duty—and yet not go on to do it. Then the inference is on a level with "theoretical" reasoning; that is, its conclusion states an objective practical necessity of action. (Compare Section IV above.)

But cannot the argument also terminate in a subjective practical necessity of action, which will ensure that the subject actually sets himself to do his duty? (Compare Section V above.)

One could make a distinction between *being aware of the fact that* something or other is duty (for example, under the law) and *acknowledging* something as duty. Then the question may be raised: is it (logically) possible to acknowledge something as being one's duty, realize that, unless one does a certain (other) thing, one cannot do what is one's duty—and yet not set oneself to do this other thing? We are here in the neighbourhood of a problem which is familiar from the ethical thought of Plato and Aristotle.

The answer to the question essentially depends upon what it is to acknowledge something as duty, that is upon how one understands the meaning of the phrase "acknowledge something as duty".

It is easy to see that, *if* to acknowledge it as one's duty to do *x* entailed that one wants to do *x* for its own sake, then the logical conclusiveness of the schema:

- (23) I acknowledge it as my duty to do *x*.
 Unless I do *y*, I cannot do *x*.
 Therefore I must do *y*

would be a consequence of the logical conclusiveness of:

- (24) I want to do x .
 Unless I do y , I cannot do x .
 Therefore I must do y .

(For, if p entails q , and if q and r jointly entail s , then p and r jointly entail s , too.)

The last pattern, although slightly different from (2) and the variations of it which we have been discussing earlier in this essay, I shall call a primary practical inference, too. There need not be any doubt that the conclusion of (24) can be a subjective practical necessity, that is, can consist in the subject's setting himself to do y under the impact of his will to do x and his knowledge or belief that it is necessary to do y in order to be(come) able to do x .

There is, however, a grave objection to regarding the subjective practical necessity of fulfilling one's derived obligation as the conclusion of a primary practical inference. The objection is that it is not plausible to think of the primary obligations, which a man acknowledges to be his, as things which he can necessarily be said to *want* to do for their own sake.

Another possibility of making duty-bound action subjective practical necessity would be to let the logical conclusiveness of (23) depend upon the logical conclusiveness of the pattern:

- (25) I must do x .
 Unless I do y , I cannot do x .
 Therefore I must do y .

To let the logical conclusiveness of (23) depend on the logical conclusiveness of (25) is to take the view that derived obligations are subjective practical necessities of action, if and only if acknowledgement of the primary obligations themselves constitutes such necessities. The pattern (25) we recognize as that of a secondary practical inference. (Compare Section XII above.) If a derived obligation emerges as the conclusion of a secondary practical inference, it follows that the practical necessity, which is the first premiss of the inference, will also be the conclusion of another practical inference. This other inference can be secondary or primary. If it is secondary we shall, by tracing the chain of obligations further back, after a finite number of steps reach a primary obligation, which is the conclusion of a primary practical inference from ends and causal connections.

On the suggested view, the things which we acknowledge it to be our duty to do are things which we *must* do—since otherwise some end which we pursue will not be attained. This is, of course, not to suggest that we should call everything "duty" which we are under a practical necessity of doing for the sake of some end. We call it duty only when this end is of a peculiar nature—for example, has some peculiar connection with ideas of

good and evil or of justice. The question of the nature of these ends falls outside the scope of the present essay and has no immediate bearing on the problem of practical inference.

To do one's duty for duty's sake can be a (subjective or objective) practical necessity of action, if this is understood to mean that one has to fulfil one's derived duty for the sake of fulfilling one's primary duty. But to do one's primary duty can be a *practical necessity* only if one has to do it for the sake of some end "beyond duty". On this point Kant, I think, was in serious error.

On So-called Practical Inference

I

Since the publication of Anscombe's *Intention* (Basil Blackwell, Oxford, 1957), practical inference has been a live topic in philosophy. There can be little doubt that the topic is important. I would claim, for example, that practical inference as a schema of explanation plays a comparable role in the human sciences to that of nomological deductive explanation in the natural sciences.¹ To vindicate this claim is difficult, however. This is so, for one thing, because the logical nature of practical reasoning is much more obscure than that of deductive and other forms of "theoretical" argument.

The division of inferences ("syllogisms") into theoretical and practical stems from Aristotle. The most exciting thing which Aristotle has to say about the topic is that a practical inference leads up to or ends in action, that its conclusion is an action.² About the premisses he is not very clear. The example he gives of the major premiss often mentions something generic which is good or ought to be done, e.g. that dry food suits every man or that sweet things ought to be tasted.³ The minor premiss mentions some particular thing which "falls under" the generic label, e.g. that this particular dish is dry or this particular lump of white stuff is sweet. In these examples the conclusion would be that the person who argues proceeds to eat the stuff.

According to Anscombe the first premiss of a practical inference mentions something wanted.⁴ This characterization, it seems, does not fit Aristotle's version of practical inference very well. It fits better another type of inference. This is an inference in which the first premiss mentions an end of action and the second premiss some means to this end. The "practical" conclusion which results from the premisses would consist in using the means to secure the end.

Aristotelian practical inferences, one could say, subsume a particular thing or action under some general principle or rule about what is good for us or is our duty. The study of reasoning of this kind is relevant to the question—central not least to Plato's and Aristotle's thinking—how knowing the good is related to being good and to right action.

The second type of practical inference is not concerned so much with

¹ On my attitude to this claim see above Introduction, p. viii. (1983)

² *De Motu Animalium*, 701a 12–14.

³ *Ethica Nicomachea*, 1147a 6–7 and 1147a 28–30.

⁴ *Intention*, Sect. 35.

right conduct as with purposive behaviour and intentional action generally. Its study is relevant above all to the problems of explaining and understanding conduct—both of individuals and of groups of men.

Of both types of practical inference it is right to say, with Anscombe, that their study has been much neglected. But perhaps not as thoroughly neglected as she seems to have thought. In his *Logic*, Hegel construed purposive action as an inference, leading from the subjective setting of an end through insight into the objective connections of natural facts to the objectivation of the end in action.⁵ (This is a very summary description of a complex and perhaps not very clear idea.) And in his lectures on the philosophy of history Hegel applied the same schema to the historical process as a realization, through the actions of individual and collective agents, of the "aims" immanent in the absolute mind. The Hegelian conceptualization of action as inference also left an impact on Marx and Marxist thinking.⁶

In this paper I shall discuss only practical inference which views actions as the use of means to attain ends.

II

Let it be assumed that the premisses of a practical argument, when conducted in the first person, are:

I want to attain the end *E* (e.g. make this hut habitable).

Unless I do action *A* (e.g. heat the hut) I shall not attain *E*.

It may very well be the case that, if I have this end in view in combination with that opinion of the means to its realization, then I shall actually proceed to act in the appropriate way. But what sort of *connection* would this signify between want and thought on the one hand and action on the other? Can I say that wanting and opining *make* me act? If so, would this be a form of causal efficacy? Or would it be more like a logical compulsion?

Before we proceed to answering these questions we must consider the following "objection" to the first person premiss as stated above. Cannot a person at the same time want to attain several ends? And can it not then happen that some of his ends (wants) are mutually incompatible? For example, that in order to attain one of the ends, he thinks he must forbear doing something which he thinks he must do in order to attain another end of his? Then there are two practical arguments for him. He may act in accordance with the premisses of one of them. But could he then, in any sense, be said to have been compelled to act in the way he acted?

⁵ *Wissenschaft der Logik*, Bk. II, Sect. iii, Ch. 2B.

⁶ As shown by Juha Manninen in "Praktisen päättelyn mukaisista ajatusmalleista Hegelillä ja Marxilla" ("On Practical Inference Models in Hegel and Marx"), *Psykologia* 5, 1970.

Suppose we change the word "want" in the first premiss to "intend". One does not normally speak of intending to attain ends. Some further change in the first premiss will therefore be required here.

One speaks of pursuing ends. Pursuit of ends is intentional action. The phrases "pursue an end intentionally" and "pursue an end" I shall regard as synonymous. To speak of intending to pursue an end points to the future. It is like saying that one is planning or has resolved to go after something later on. But an agent in pursuit of an end can rightly be said to intend to make the end-state, i.e. the state which obtains when the end is attained, materialize or come true.

It may now be suggested that, if one pursues an end *E*, then one cannot, for conceptual reasons, at the same time pursue another end *E'* the pursuit of which one considers (causally or logically) incompatible with pursuit of *E*. I shall accept this as a valid point about "the logic of intention".

It seems obvious that "I pursue the end *E*" is a logically stronger statement than "I want to attain the end *E*". The first entails the second, but not *vice versa*. A plausible way of coping with the difficulty caused by possibly conflicting ends would therefore be to replace the first premiss in the practical argument under consideration by

I intend to make it true that *E* (e.g. that this hut is habitable).

and the second premiss by

Unless I do *A*, I shall not achieve this.

Is there a conclusion which can be said to follow logically from the two premisses thus reformulated?

I think there is one for which this claim can be made. It is the conclusion:

I will do *A*.

This is what we call a *declaration of intention*. An inference is normally thought to be between true or false propositions. But it is doubtful whether a declaration of intention could qualify as a proposition. The logical nature of the argument therefore is obscure.

In order to see things in a clearer light, let us shift the argument from the first person to the third person. What we then get is, in the first place:

X intends to make it true that *E*.

Unless he does *A*, he will not achieve this.

Therefore *X* will do *A*.

Here all three components of the inference are clearly propositions. The conclusion is a prediction about a certain agent's conduct.

It is quite clear, however, that this argument is *not* logically conclusive. There are several reasons why this is so. For one thing, both premisses may be true but *X* himself need not know or think that the second is true,

i.e. he need not realize that he will not attain his end unless he does *A*. And then there is, of course, no guarantee whatsoever that he will do *A*. (I shall not here take up for separate discussion the possibility that *X* is not aware (conscious) of his intention and whether this would affect the conclusiveness of the inference.)

We might now try to make the argument approach conclusiveness by expanding the second premiss to something like

X thinks that unless he does *A*, he will not bring about that *E* comes true.

It should be noted that the same expansion may be said to be implicit in the premiss in the first person. For, if I argue to myself "Unless I do *A*, I shall not achieve this or that", then I also subscribe to the *truth* of the statement. Perhaps I do not claim to *know* it to be true, but at least I believe it or consider it highly probable.

In the case of the premiss in the third person, however, its expansion is not merely to make explicit a concealed presupposition. It adds something substantial to the statement. We shift from a statement about "objective facts" to a statement about a person's "epistemic attitudes".

It may be suggested that in order to make the argument in the first and the third person run completely parallel we must not only add the above epistemic clause to the second premiss of the argument in the third person, but also change the conclusion of the argument in the first person to

I shall do *A*.

When cast in this form the conclusion is a prediction about my own future behaviour. This would turn the first person argument into a special case of the third person argument. Such a "reifying" move is possible, but it misses what is specifically "first person" about the first person argument.

III

There is, however, also another course to be tried in order to make the argument in the third person conclusive. We leave the premisses as they (originally) stand and change the conclusion to:

Therefore *X* has to (must) do *A*.

Whether this conclusion follows logically from the premisses depends upon how one interprets the meaning of "has to" or "must" here. The following seems to me a reasonable view to take: A must-do-statement (in the third person) is elliptic. It says that an action is a necessary condition for the attainment of some end which is left unspecified in the statement. On this view the statement "*X* must do *A*" would be short for "*X* is in the pursuit of some end of action of his which he will not attain unless he does *A*". And *this* follows logically from the premisses (by application of the

principle $fa \rightarrow (Ex)fx$ and assuming that an agent's intending to make a state of affairs come true entails that this state is an, intermediate or ultimate, end of his action).

Consider this variation of the argument:

X intends to make it true that E .

Unless Y does A , he (i.e. X) will not achieve this.

Therefore Y must do A .

Shall we call this a valid argument? The answer again depends upon our understanding of the "must" here. The argument is conclusive, if " Y must do A " is short for "*Someone* is in pursuit of an end of action which he will not attain unless Y does A ". Is this a feasible interpretation? Perhaps must-statements are sometimes thus understood. But why not then take an even more "liberal" attitude and say that sometimes " Y must do A " means simply "Unless Y does A , something or other will not happen (be the case)"? Here the necessity of action need not be related to any end at all, nor to anybody's intentions.

Consider finally the following variant:

X intends to make it true that E .

Unless Y does A , he (i.e. X) will not achieve this.

Therefore he must make Y do A .

But must he? Assume that we (or X himself) know that he cannot make Y do A , that nothing he does will move Y to this. Then we should perhaps say that X must give up pursuing E as an end. And if pressed about the meaning of this last "must" we might say: X must do this, or else he will suffer frustration. But maybe X is so keen on his end that giving it up involves a still bigger frustration for him. Then we might conclude that X *must learn* how to make Y do A (which may or may not involve teaching Y how to do A). And this conclusion would follow, logically, if we change the second premiss to "unless X learns how to make Y do A , he will not make it true that E ". We are back at the first pattern.

Generally speaking: If there is something which an agent cannot do and which is such that, unless he does it, he will not attain a certain end of his, then he must *learn* how to do it or else he will not attain this end. Learning how to do something can be a practical necessity in much the same sense as is doing something. Furthermore, that somebody else does something can be necessary (causally or logically), if an agent is to attain his ends in much the same sense in which it may be necessary for this that something should happen. These last two are comparable cases. I propose, however, that the term *practical necessity* be reserved for the first pair of cases. Practical necessity is the necessity of doing something under which an agent is, if he is to attain some end *of his own*.

An agent thus is under a practical necessity of doing everything which it

is necessary that he should do, if he is to attain the ends he pursues. But can an agent intend to do anything (make it true that *E*, for example) without being able (generically) to do this?⁷ Does intention presuppose ability—pursuit of ends ability to obtain them? I would answer as follows:

One can intend to do only such things which one either thinks one can do or realizes that one cannot do but intends to learn to do and thinks one can learn to do.⁸ (This is not a psychological observation about intending, but a rule about the logical grammar of the concept.)

Sometimes an agent thinks he can do something which, in fact, he cannot generically do. By trying to do this thing but failing to accomplish it he may come to realize his inability—and, if the action was a practical necessity for him, the practical necessity of acquiring the ability. But it may also happen that when the agent realizes his inability to do the required thing, he “changes his mind”, i.e. gives up the pursuit of his original end.

IV

Let us again shift back to the first person case. Consider this inference:

I intend to make it true that *E*.

Unless I do *A*, I shall not achieve this.

Therefore I must do *A*.

Can I be mistaken about my own intentions, think that I intend to do something which, in fact, I do not intend to do? Without prejudging an answer to this question, let us just lay it aside.

The second premiss can surely, “objectively speaking”, be false. Then the conclusion too will normally be false. (It will be false, unless I also pursue some end which is different from *E* and for the attainment of which the doing of *A* is required.)

When the second premiss is false, it may still be the case that I (mistakenly) *think* it true. And then I will also think of the conclusion as true, i.e. I will think of the doing of *A* as a practical necessity incumbent upon me. Thus arguing from the two premisses above makes the conclusion *valid for me*, even though the second premiss may be false and therewith the conclusion false too. This is a peculiarity of a practical inference in the first person.

If the second premiss of a practical inference in the first person is true,

⁷ That an agent is (generically) able to do something shall mean that on most occasions, when there is an opportunity for doing this thing and he sets himself to do it, he will also accomplish it. He has the necessary “know how”. This he may, but need not, possess as a result of learning.

⁸ My position here differs slightly from the one which I took in *Explanation and Understanding*. There (pp. 101–3) I suggested that an intention to do something necessarily involves an opinion on the agent’s part that he can do the thing.

then the agent who conducts the argument will *rightly* consider himself to be under the practical necessity of acting which is stated in the conclusion. If it is false, however, he (normally) will *mistakenly* consider himself to be so. The practical necessities under which an agent (rightly or mistakenly) considers himself to be we may call his *subjective* practical necessities. And we can distinguish them from the *objective* practical necessities under which he is, as a matter of objective truth. There can be a discrepancy between the groups, either because the first one does not embrace all the members of the second, or because it contains members which are not members of the second. Each source of discrepancy is a source of frustration for an agent in pursuit of ends.

V

We now leave the "must do" case. The arguments henceforth to be discussed are:

First person case

I intend to make it true that *E*.

Unless I do *A*, I shall not achieve this.

Therefore I will do *A*.

Third person case

X intends to make it true that *E*.

He thinks that, unless he does *A*, he will not achieve this.

Therefore he will do *A*.

The conclusion of the first person inference, we said, is a declaration of intention. This intention may not have been formed until we realized the practical necessities involved in our aiming at a certain end. So its formation may come later, after the first intention was already formed. We can speak of a *primary* and a *secondary* intention here.

The connection between the two intentions is, moreover, a kind of logically necessary connection. The second (epistemic) premiss can be said to "mediate" between the primary intention of the first premiss and the secondary intention of the conclusion. One can also speak of a transfer or *transmission of intention*. The "will" to attain an end is being transmitted to (use of) the means deemed necessary for its attainment.

This principle of "transmission of intention from ends to means" is basically identical, it seems, with a principle which Kant thought analytically (logically) true and which he expressed in the following words: "Who wills the end, wills (so far as reason has decisive influence on his actions) also the means which are indispensably necessary and in his power."⁹

⁹ *The Moral Law*, Hutchinson, London, 1956, transl. by H.J. Paton, pp. 84-5. Cf. above, p. 9.

I said there was "a kind of" logical connection here. For, although I think it obvious that the conclusion of our first person argument follows logically from the premisses, this "following" has a peculiar nature: A declaration of intention is not a true or false proposition. Whether an argument which is actually conducted from premisses of the kind under consideration will terminate in a declaration of intention is, in a certain sense, contingent and not logically necessary. Perhaps such arguments as often as not terminate in a *change* of intention on the agent's part. Realizing what is necessary for him to do, and perhaps feeling a strong aversion against doing this, he changes his mind and gives up pursuit of the original good (cf. above p. 23). And even if the agent does not change his primary intention, he need not *declare* his secondary intention, not even to himself, "in thought". He can nevertheless be said to *have* it and, moreover, to have it necessarily. If he declares his intention, he only "reveals" to the world, or admits to himself, something about himself which is already there.

So what is logically necessary is something which holds between true and false propositions and which is best expressed in the form of a third person inference of the following type:

X intends to make it true that *E*.

He thinks that, unless he does *A*, he will not achieve this.

Therefore *X* intends to do *A*.

This argument, however, moves wholly on the level of intentions and epistemic attitudes. It does not, it seems, link intention with action. Or does it, after all? We shall return to this question presently (in Section VII).

VI

Even after we have added the epistemic clause "*X* thinks that" to its second premiss the argument in the third person, ending in a prediction, remains inconclusive. This is due to the existence of a time gap, separating the premisses from the conclusion. The premisses may be true, but the agent has not yet performed the action. This leaves open the truth-value of the conclusion.

Assume that the premisses are true at a certain time. Normally, they will then remain true for some time. Unless *X* does *A* *within this time span*, we shall have to say that the prediction failed to come true. For, if *X* does *A*, but only after the end of the time span, his action is irrelevant to the argument. So is it also, if after the lapse of time he does *not* do *A*.

It can happen that we wait for the prediction to come true and, having waited for some time, begin to doubt whether the premisses still obtain. Suppose we find that they do no longer obtain. Does it follow that the conclusion was invalidly drawn from the premisses? Not necessarily. For

assume that what *X* intends to achieve is the result of a "long time project" and that the doing of *A* has to be done at a late stage of it, or that the doing of *A* can be postponed to a late stage. There is in any case a *latest time* when *A* will have to be done. Suppose now the agent "gives up" his project before this time arrives. Then it could still be true that, had he not given up, he would have done *A* before the time was up.

So in order to show that the conclusion "he will do *A*" was falsely drawn from the premisses, we should have to show that the agent failed to do *A at the latest time* when he thought the time was up for him to do it and the premisses were still valid. This observation makes it possible for us to eliminate considerations relating to the time gap from considerations relating to the logical conclusiveness of the inference schema here. The problem can be reduced to a question about a given moment in time, we shall call it "now", and the argument reformulated as follows:

X intends (now) to make it true that *E*.

He thinks that, unless he does *A* now, he will not achieve this.

Therefore he will do *A* now.

The mere fact therefore that in ever so many cases an agent will never do what he considers himself under a practical necessity of doing, if he is to attain some end of his, is by itself quite irrelevant to our problem. For in many of those cases the agent had changed his mind before the time was up in which he had to do the thing he never did.

Is the statement that an agent *will* do a certain thing *now* a "prediction"? The answer depends upon what we think of the "now". Is "now" the present *instant*? Very often we conceive of the "now" as having duration, extension in time; sometimes we mean by it to-day or even this year. So what is "now" supposed to mean in the argument we are discussing here? Not necessarily an instant in time, but rather a *temporal location* of some duration. Of this "location", however, the following must hold true: We cannot slice it up into an earlier part of which it is true that the agent did not even *begin* to do *A* then and a later part of which it is true that he then did *A*. The "now" of which we are here speaking covers what may be called the *specious present* of the performance of an action. Most actions take some time to perform. Only when the action is one which can be performed at an instant ("in an inkling") does the specious present of its performance, the "now", shrink to a point. (Whether there are such actions, I do not know.)

Under the above conception of the "now", *will do A now* "reduces" to *does A now*. One could also say that "will do" has no proper application to the "now" case—and that the conclusion to match the two premisses must be reformulated "Therefore he does *A* now". In any case, the conclusion is no longer a prediction. It is a description of what an agent does, is

engaged in doing, now.¹⁰ ("He will accomplish the doing of *A* now" would be a prediction which is compatible with the present tense description "he does *A* now".) So, in non-linguistic terms, the conclusion is an action. This is exactly what Aristotle said it should be.

But what remains open to debate is, whether this conclusion *follows*, i.e. whether an agent's present intentions and awareness of what he has to do right now in order to make them effective will of logical necessity lead, "compel", "prompt" him to action.

VII

From what was said before about "transmission of intention" we should from the two premisses

X intends to make it true that *E*.

He thinks that, unless he does *A* now, he will not achieve this.

be able to conclude:

Therefore *X* intends to do *A* now.

The problematic conclusion was

Therefore *X* does *A* now.

Our next task will be to compare the two conclusions.

If an agent does *A*, is it then also true to say that he intends to do *A*?

Sometimes a man is said to have done various things which he never intended but which happened to be consequences of his action. (He may even have foreseen that they were going to happen.) Perhaps one could call this "accidental doing" and distinguish it from "intentional doing". Here we understand by "doing" only intentional doing. Thus the practical argument would be invalidated, if it could be shown that on some occasion its (above) premisses were true and that the agent did *A*, though only accidentally, by mistake or as an unintended consequence of his acting.

So our question must be given this more precise formulation: Does it follow that an agent intends to do *A*, if he intentionally does *A*?

It is important to distinguish between intentional doing and intention(s) to do. "Intend to do" normally refers to the future. If "now" designates the specious present of the action, then under this normal use "intend to do now" is a self-contradictory phrase of a sort. (Not to be confused with

¹⁰ If instead of "now" we had put in a reference to an arbitrary moment of time "at *t*" in our inference schema, the conclusion "He will do *A* at *t*" or "He does *A* at *t*" can be a prediction relative to the station in time where we happen to be. But this is irrelevant. What matters is that the conclusion is not a prediction *relative to the premisses*—in this case that the agent, at *t*, intends something and considers the doing at that very moment of some action necessary.

"intend now to do" which is logically in order.) "Intend to do", like "will do", has no proper application to the "now"-case. But what happens then to a persisting intention to do, when the moment of action is there? A suggestion would be that it "matures" and emerges on the behavioural level as intentional action.

Shall we say then that the two conclusions which we are comparing are identical? Shall we say that when the transmission of intention argument is applied to the present moment, then the conclusion which emerges is that *X* does *A* now?

The answer to the first question is, I think, affirmative. To answering the second affirmatively there is an objection, however. This objection may appear so grave that, in fact, it shows that action cannot possibly follow logically from premisses about intentions and epistemic attitudes.

VIII

Something happens which *prevents X* from doing *A*. He stumbles or has a stroke or someone seizes him and keeps him immobile.

If intention is for the future and the preventive interference occurs before the agent makes the intention effective, he will normally either change his intention or evaluate the practical requirements of the situation differently. This case therefore does not affect the problem of the validity of the inference schema which we are discussing.

If the preventive interference occurs within the specious present of the action but after the agent has commenced acting (set himself to act, embarked upon the road to the end), then we usually describe the case by saying that he *tried* but failed to accomplish the thing. Failure to accomplish can also be due to a change of intention within the specious present of the action. Then we say that the agent began or tried, but gave up. Failure can finally be due to the agent's not knowing, how to accomplish the thing. This can be called failure of ability. The agent had misjudged his powers. He intended to make it true that *E*. He realized that doing *A* was a practical necessity for this. He thought he could do *A*. But he was mistaken. Learning how to do *A*, or at least perfecting his ability, was a prior practical necessity for him (cf. above p. 22). Since he did not understand this, he proceeded to do things in the wrong order, so to speak. He set himself to do *A* but found that he could not accomplish it. This is another case of trying but failing.

It is a common feature of all these cases of failure to do *A* that the agent *embarks* on the action and that the failure is one of *accomplishing* the action. Embarking on an action involves some physical effort and therefore some behaviour or conduct aiming at the (completed) action. If the agent's embarking on the action is, moreover, a logical necessity under the

premisses, we could still say that Aristotle was substantially right in thinking that a practical argument terminates in (some) action.

But we have also to consider the case, at least as a theoretical possibility, that the preventive interference occurs exactly at the same time as the specious present for the action begins. Then the agent cannot be said to have embarked upon the action. There was no time for doing this. At most we could then say that he *would* have embarked upon it, had he not been prevented.

If we understand "prevented" to mean either prevented right at the beginning of the action or prevented in the course of an attempt to perform it, we can state the conclusion which finally emerges from our considerations of the form of a practical inference as follows:

Therefore *X* does *A* now, unless he is prevented or else cannot accomplish the action.

IX

In order to show that an agent *does* a certain thing *A*, it is not enough to show that *A*, the result of the action, happens as a causal consequence of some movements in the agent's body (rising of arms, twisting of hands, etc.). We must also show that what took place was *intentional*, i.e. was a case of an agent's *doing A* as distinct from merely bringing it about without intending it.

To establish that the agent's causing *A* to come about is a case of his doing *A* is not to establish, in addition to the happening of *A*, a different event which so to speak occurs "inside the agent". It is to *understand* (the meaning of) the agent's conduct, i.e. to see that by certain changes in his body or changes causally connected with changes in his body the agent is *aiming* at this result. If he aims at it without achieving it, we shall have to say that the agent tried but failed—either because of insufficient ability or because he was prevented.

Now we can see more clearly, I think, wherein the claim to logical validity of the practical inference consists. Given the premisses

X now intends to make it true that *E*.

He thinks that, unless he does *A* now, he will not achieve this.

and excluding, hypothetically or on the basis of investigations, that he is prevented, then his actual conduct, whatever it may "look like", either is an act of doing *A* or aims, though unsuccessfully, at being this. Any description of his behaviour which is logically inconsistent with this is also logically inconsistent with the premisses. Accepting the premisses thus forces on us this understanding of his conduct—unless for some reason we think that a preventive interference occurred right at the beginning of his action.

X

What *uses* has the type of argument which I here call practical inference?

One can distinguish between a *retrospective* and a *prospective* use.

When the argument is used retrospectively we start from the conclusion and so to speak reconstruct a set of premisses to match it. Then the "conclusion" normally is the established or hypothetically assumed proposition that an agent has done a certain thing, performed an action *A*. (But it can also be the proposition that he set himself to do a certain thing which he failed to accomplish, either because he changed his mind, or was prevented, or did not possess the needed ability. Here I shall consider only the "normal" case.)

An agent did *A*. Why did he do it? We *explain* his action by placing it in the "teleological perspective" of his aiming at some end and his epistemic attitude to the requirements of the situation, i.e. his judging the action a practical necessity under this end. This is a prototype case of what is usually called *teleological explanation*.¹¹

I did *A*. Someone challenges me: Why did I do this? I may *justify* my conduct by reference to what I was after and that I thought my doing of *A* a practical necessity for me. (This, however, is not the only sense in which a man is said to "justify" his actions.)

When the argument is used prospectively we set out from the premisses and "extract" from them a conclusion. In the first person case the argument ends in a *declaration* (formation) of *intention* to do a certain thing. I *commit* myself to doing it. Such use of the argument takes place particularly in situations where I *ponder* what I have to do in order to attain some already set aim of mine.

In the third person case the prospective (forward-looking) argument produces a *prediction*. Since so and so is the agent's aim and he evidently considers the doing of *A* necessary for its attainment, he will (probably) do *A*.

Is there, in addition to the retrospective and prospective uses of practical inference also something which might be called an "instantaneous" use of it? Is there explanation of action only on the basis of what is *now* the case—and is there intentional action which is *simultaneous* with the construction of a justification for it?

I think the answer to these questions is No. When practical inference functions as an argument it either looks to the past for the reasons or to

¹¹ The term "teleological explanation" has a multitude of rather different uses. For explanations which set human actions in a "teleological" perspective of intentions and epistemic attitudes I would now prefer the term *intentionalist* explanation. Cf. my book *Freedom and Determination*, North-Holland Publishing Company, Amsterdam 1980, p. 28 and *passim*.

the future for an action. So in any use of the argument there is a time gap involved between premisses and conclusion and with this gap there is also a rift in the logical connection between the intention and epistemic attitude on the one hand and the action on the other hand.

When therefore, in Section VI, we closed this time gap, we thereby also obliterated the character of an argument or an inference from the propositional connections which we were investigating. What happened to our "practical inference" was in the end (Section IX) that we turned it into a set of conditions under which the conduct of an agent has to be interpreted or understood in a certain way, *viz.* as the doing of *A* or as aiming at this result. The premisses of the practical inference became the description of a teleological perspective in which conduct is being understood as intentional.

What then will be our final position on the question of conclusiveness of practical arguments?

When the argument is used retrospectively to explain or, in the first person, to justify action, the clause "unless prevented or else cannot accomplish the action" does not appear in it. That the agent did or tried to do *A* is not called into question and is, moreover, something that was logically bound to happen, assuming that the teleological frame, the intention and the epistemic attitude which the premisses attribute to him, lasted up to the moment of action. For, granting the truth of this assumption, the premisses then set the conditions for *interpreting* what happened.

When the argument is used prospectively in the first person, the clause "unless prevented or else cannot accomplish the action" likewise is no part of the argument. Therefore we have no logical guarantee here concerning that which the agent will in fact do. But this is not what the argument purports to give either. Its binding nature stems from the fact that its conclusion declares the intention which an agent is logically bound to have within the teleological frame which in the premisses he acknowledges for his prospective action.

Only when the argument is used prospectively in the third person, is the unless-clause a part of it. We know that if this clause is satisfied and if the teleological frame attributed to the agent in the premisses remains stable up to the moment of action, then conduct of the predicted character is logically bound to follow. (Since this is how the situation will then have to be understood by us.) But whether the clause will be satisfied and whether the teleological frame will remain stable, we cannot know in advance. It is therefore contingent, whether the prediction will be fulfilled or not.

Postscript on Understanding

The "logical core" of the propositional connections which I have been discussing in this essay can be called a schema of interpretation or of understanding conduct as being intentional (action). This schema is what

a practical inference "becomes" when it is applied to the moment of action, the "now"-case—and ceases to be an inference. I tried to show how the validity of this schema is relevant to the validity of the inference.

This schema also has an independent use, namely for purposes of *understanding* what an agent is doing—as distinct from the (third person) uses of the inference which is to *explain* why an agent did what he did or to *predict* what he is going to do. But here a warning is in place.

On no account must it be thought that the understanding of behaviour as action is always based on an interpretation of an agent's conduct in the light of some intentions and cognitions which we attribute to him. To think thus would be to distort the logic of the situation seriously. It would, moreover, be a similar distortion to the one of which philosophers make themselves guilty, when they say that any statement about physical objects is the result of interpreting certain sense-data.

In the normal cases we say off-hand of the way we see people behave that they perform such and such actions—raise their arms, walk or run, open key-locks or hand things over to one another. Many of these actions we ourselves know how to perform; those, and others which we cannot do, have a familiar "look" or "physiognomy" which we recognize. We are further acquainted with innumerable ends for the sake of which these actions may be performed (when not "done for their own sake"). Therefore we need not first seek for the agent's end of action in order to be able to tell *what* he is doing. Only when we are curious, *why* he is doing what we already think we know he is doing, do we look for a specific object of intention with him. (Explanation of action follows *after* understanding behaviour as action.)

These are the normal cases. But there are others. There are first of all the abortive cases, when an agent fails to accomplish the action. Even then we can usually say what, in fact, he did. He, for example, grabbed a handle and pulled, twisted his hand, etc. and in doing these things he was trying, though unsuccessfully, to do a certain thing which one normally does by doing just those things (and perhaps something else beside). In what he *did* we see at once what he was *aiming* at—and therefore we say that he *tried* to do so and so. On the whole only when failure to accomplish is due to insufficient know-how, or to quite false conceptions of how to achieve the result (of the action), may it be necessary to "interpret" what we witness in the light of hypothetical aims and cognitions of the agent. We then say some such thing as: "Evidently he is trying to open the safe but has no idea of how to do it."

Secondly, there are cases when we are unfamiliar with the (kind of) action and therefore do not understand *what* the agent is doing. Perhaps he performs some strange ceremony. Even then we can usually describe various things which he is doing in the course of doing this strange thing. We, e.g., see him move his legs and arms in certain ways and not only see

his legs and arms move. When we nevertheless say we do not understand what he is doing, this is because we feel reluctant to say that what he primarily intends to do is to make those movements with his limbs but suspect that the object of his primary intention is something which is done *by moving legs and arms in a certain way.* (Perhaps he is saluting or dancing.) We must learn what the new action “looks like”, become acquainted with its “physiognomy” before we can understand what a man is doing, who behaves in this way. We can learn this by being told what the agent is doing, sometimes by just being told the *name* of the action. (I can learn to recognize when a man is saluting without having any idea why people salute.)

So not even in the cases of unfamiliar actions is a practical inference always or normally needed to understand what is being done. Only in some, rather exceptional, situations will the construction of a practical argument be helpful here. We see a person go through some movements, the significance of which we do not understand “in themselves”—but we have a strong hunch that he evidently intends to do a certain thing in behaving thus, e.g. that he is saluting somebody in the street. (But then we must also know, not just the name of the action, but also something about the occasions on which it is appropriate to perform it.) “So this is how these people salute”, is our guess. Here a practical inference can be said to guide our understanding.

Sometimes we mistake behaviour for action when in fact it is not. A person's arm goes up and I say I see him raise his arm. Then I learn that what I witnessed was a physiological experiment and that the person's neural system was being stimulated in a way which made his arm go up. (Perhaps the experimenter had even told him not to raise his arm “himself”.) Then I must withdraw the claim, implicit in what first I said, that the person raised his arm. I can redescribe what I saw in a way which does not carry that claim with it by saying that I saw his arm rise. When thus I modify my original description, it is tempting to speak, *post hoc*, of a false interpretation of what I saw. We say: I interpreted what I *really saw*, viz. his arm rise, as a case of seeing him raise his arm—but this was premature. This is like the case, when we withdraw a physical object statement and replace it by a sense-data statement, because it turned out that the object we saw was only illusory—say a column very skilfully drawn on the wall. In both cases, however, is it misleading to say that we interpreted what we saw in the wrong way. For we did not *interpret* what we saw at all.

I can understand something as an action and be mistaken in thus understanding it. I can see something as a physical object and be mistaken in thus seeing it. But this is not to say that what I really *saw* was, e.g., his arm rising and not him raising his arm—or a painting on the wall and not a column. Understanding is compatible with misunderstanding, one could say. Only when we are on our guard against misunderstanding or confused

about the nature of the case, do we *interpret* what we immediately witness. And if what we witness are the movements of some living beings, the interpretation normally consists in the construction of a practical inference to match the case.

Determinism and the Study of Man

I

By determinism I shall understand ideas according to which that which is also (somehow) *had to be*. Such ideas play an important role both in the natural and in the human sciences. Whatever the ultimate answer may be to the question as to how these two types of sciences are related, the following difference between them is striking on the surface:

In the natural sciences, deterministic ideas are connected with such other ideas as those of universal regularity, repeatability, and experimental control. In the human sciences the immediate connections are with ideas such as motivation and social pressure, goal-directedness and intentionality. In the natural sciences determinism serves in a large measure the forward looking aims of prediction; in the human sciences there is a relatively much stronger emphasis on retrospective explanation, or understanding, of what is already a *fait accompli*.

These differences between naturalistic and humanistic study in relation to determinism I would attribute to the following source—well aware that what I say may sound provocative: Natural science can be characterized as a study of phenomena under the “reign” of natural law. Human science again is primarily a study of phenomena under the “reign” of social institutions and rules. I shall argue for what may be termed a “methodological parallelism” between laws of nature and rules of society. That is, I shall argue that deterministic ideas in the human sciences have a relation to societal rules which is *analogous* to the relation in the natural sciences between deterministic ideas and natural laws. I am not, of course, saying that laws of the state and other social rules are, in themselves, like laws of nature. They are, on the contrary, very different. The former are normative, the latter descriptive, as we say. And from this profound difference between the two types of law it follows that, if the “methodological parallelism” I am making is at all correct, determinism in the study of man means something utterly different from determinism in the study of nature.

I can foresee immediate objections. The theoretical aim of the human sciences, it will be said, is to discover laws which state, not regulations and rules of conduct, but factual regularities of individual and collective behaviour and of institutional change. Such laws, if there are any, will “determine” or “govern” the life of men and of society in a similar way to that in which the laws of nature “determine” or “govern” natural phenomena. On this view, determinism in the study of man would mean essentially the same thing as determinism in natural science.

Deterministic claims of this character have often been made for the

human sciences. Sometimes with the qualification that the laws of behaviour are not as "rigid" as the laws, say, of classical mechanics, but are rather of a probabilistic than of a "strictly" causal nature. Or it is said that it has not yet been possible to formulate them exactly. The social sciences are still young and we must not expect too much of them in their infancy.

To view matters in this light is, I think, to be guilty of a serious "methodological misunderstanding". It is symptomatic of an illegitimate transfer of conceptions and ideals from the natural to the human sciences. To say this is not to deny legitimacy to a study which ascertains factual regularities in human and social behaviour. Nor is it to exclude the possibility of relying on such regularities for successful predictions. But it is an invitation to view in a new light the study of man and all theorizing and philosophizing in the social sciences.

This shift in point of view will also have consequences for the age-old philosophic problem of "the freedom of the will". This is, roughly, the question of the relation between the actions of an individual and various forces working from within him: his will, his wants, his passions and sentiments, and his deliberations. There is a deep-rooted tendency to wed this question to the further question of the relation between neural processes and macroscopic reactions in the individual's body. This "internalization" of the problem is another illegitimate transfer to a sphere where it does not belong of an attitude appropriate to the natural sciences, in this case physiology. The way to a solution is what I propose to call an "externalization" of the problem. This consists in working one's way from considerations of individual action to considerations of the factors which influence ("determine") the individual's conduct as a member of a community, in the context of institutionalized human relationships. It would be a slight, but still useful, exaggeration to say that the problem of the freedom of the will, like the problem of freedom generally, is essentially a question of social philosophy.

II

As a basis and starting point I shall present and briefly discuss two patterns of explanation of action.

The first explanatory pattern I shall call *intentionalist* explanation. It is related to a type of reasoning sometimes called "the practical syllogism". In its simplest form this reasoning goes as follows:

A intends to *p* (e.g. go to the theatre tomorrow).

A thinks that unless he *q*'s (e.g. reserves a ticket in advance), he will not be able to *p*.

Therefore: *A* takes steps to *q*.

The inference remains valid, if for "intends" we substitute "has decided"

or "is resolved" or "is determined", perhaps also "is anxious". For "thinks" in the second premiss one can also put "considers", "realizes", "knows", or "believes".

It is easy to see what practical reasoning has to do with action explanation. Assume *A*, as a matter of fact, *q*'s. We are curious to know *why*. It would be a satisfactory answer to this question to point out that he intends to *p* and considers *q*'ing necessary to this end. And the same answer would explain, why he tried to *q* in a situation when he failed.

It is quite natural to say here that *A*'s behaviour was *determined* by his intention and epistemic attitude. Given them, he had to do what in fact he did. We can speak of the intention and epistemic attitude as *determinants* of the agent's action and say that they jointly constitute a (sufficient) ground or *reason* for *q*'ing. The thing for the sake of which *A* undertook to *q* I shall call *A*'s *object of intention*, and the *q*'ing itself, I shall say, was part of the *requirements of the situation*, as *A* saw it, upon his action.

Suppose that *A* considers *q*'ing sufficient, though not necessary, for attaining the object of his intention. He intends to go to town and knows he will have to use a public conveyance to get there—say, either take a bus or a train. He takes steps to catch the bus. Should we here too say that his actual choice of the bus is fully explained, determined, by what he intends to do and knows about the means of making his intention effective?

Obviously we should *not* say this. The behaviour which we can explain on the basis of the facts, as I presented them, is the "disjunctive action" which consists in the agent's taking a bus *or* taking the train. This action he can perform in one of two ways, *viz.* by taking a bus and, alternatively, by taking a train. So, if now he chooses the bus, he performs the disjunctive action. *This* action is then fully determined by the agent's intentions and beliefs—but not his actual choice of alternative.

But could not his choice be determined too? Certainly it could. Various *reasons* might have existed for his choosing to go by bus: perhaps it is safer or cheaper or quicker than going by train. If his choice of means of transportation can be attributed to some such reason, then it is also true to speak of the *choice* as determined. But it is important to note that, although a man's choices between alternative courses of action *can* be in this sense determined, they *need not* be. To insist that they must would be sheer deterministic dogmatism. Choice *can* be completely "fortuitous".

III

Many actions are performed in response to a verbal (or other symbolic) challenge. The challenge can be, for example, an order which I obey, or a request with which I comply, or a question which I answer—or a traffic light to which I respond. Why did I reach out for the salt on the table and

hand it to my neighbour?' Because he asked for it. This can be a complete explanation. His request determined my action, constituted for me a sufficient ground or reason for doing what I did.

Normally, when I respond to a challenge of this kind, I cannot rightly be said to *intend* to respond. I simply *respond*.

Assume, however, that the saltcellar had slipped from my grip and the salt poured out on the table. I hear people cry out: "What are you doing?" Then I could quite truthfully reply: "I intended ('I meant') to hand the salt to X, who asked for it". The fact that I can give this reply, when I fail in the performance, shows that complying with a request is intentional action—and not just a conditioned response to a stimulus.

In our example the object of intention was to hand the salt to my neighbour. This object was, so to speak, set by the request, and only reconstituted by me in retrospect. For this reason I shall call it externally set—and contrast it with the internally set objects of intention which are there when I can say, before proceeding to act, what I intend to achieve.

Similarly, we can divide the determinants of action into internal and external ones. Intentions and epistemic attitudes are of the former, symbolic challenges of the latter kind.

Response to verbal and other symbolic challenges is participation in various *institutionalized* forms of behaviour or *practices*. That the practices are "institutionalized" means that they are shared by a community into which we are reared by being taught to participate.

Response to symbolic challenge is only one form of participation in an institutionalized practice. Another is behaviour in conformity with rules such as the laws of the state or the codes of morality and good manners or customs and traditions. "Why don't you park your car here?" (It would perhaps be convenient.) The answer might be "It is not allowed". Here a traffic regulating rule functions as a determinant of my behaviour. Obedience to it is an externally set object of my intentional acting.

The proportion of our actions which are determined internally and externally respectively, is not fixed. It varies from society to society, and it varies with the position of the individual agent in society. In a social order with many taboos and ritualized ways of life external determinants can steer the actions of men in the minutest details. In such societies the margins of individual freedom are very narrow.

IV

External determinants of our actions are given to us like stimuli to which we react. Such responses have to be learnt—as is the case with conditioned reflexes. Learning to participate in institutionalized forms of behaviour is

¹ I owe this example to Professor Frederick Stoutland.

connected with a characteristic motivation. I shall call this motivational mechanism *normative pressure*.

Non-conformity to legal and moral and other rules of conduct and good manners is likely to have unpleasant consequences for the agent. In the case of legal norms the "administering" of these consequences is itself institutionalized and consists in various coercive measures against the agent. In the case of moral norms the consequences are disapproval, ostracism, loss of esteem or confidence—things which make a man ill at ease in society.

I think it is essential to the idea of participating in institutionalized patterns of behaviour that it should be surrounded by this "aura" of normative pressure. This does not mean that the answer to the question why people participate or conform to rule, is always or even normally teleological. People do not usually conform *in order to* escape the unpleasant consequences of non-conformity. But sometimes they do this—e.g., when participation or conformity is against their personal interests or connected with discomfort or when it seems pointless. Then the answer to the question: Why did you do *x*? could be: Had I not done *x*, *y* would have happened to me and this I am anxious to avoid. Here the determinant of my action is not the invitation to participate—but the internal determinant constituted by my intention to avoid a certain thing which I consider likely to happen to me, if I do not participate.

In educating people, particularly children, to participate in practices and obey rules, rewards also play a characteristic role. When reward is merely an alternative to punishment in making people conform, I shall call reward external. On the whole external reward seems to play a subordinate role in the institutionalization of patterns of behaviour. Philosophers have noted and tried to give an account of this fact. I think the explanation should be sought along the following lines:

Institutionalization of behaviour normally serves a purpose. Institutions have what we call a social function. Without traffic regulations there would be chaos on the roads. This nobody wants. Participation in the practice, by everybody concerned, is therefore supposed to be in the "public interest", i.e., something which will be in the interest of each individual participant, bring him some good. This good can be thought of as a "reward" connected intrinsically with the practice, i.e., with the idea of having it. Therefore it is a further feature essentially connected with the institutionalization of behaviour that the reason for conforming to the set patterns should, on the whole, *not* be the impact of normative pressure, but simply acceptance of the rule. When rules function in this way, they are also said to be *internalized* with the members of the society in question. The more often normative pressure determines behaviour, the more strongly is the coercive force of society felt and the less "free", in a subjective sense, are the individual agents. But internalization is also a loss of

freedom of a kind. For it means that externally given stimuli are allowed to determine the actions. It is on these two forms of non-freedom that social critics play. They question the fact of internalization, e.g., raise and make people reflect on the question whether various institutions and practices *are* in the "public interest" or whether they perhaps only serve to cement interests not at all "public" but, say, those of a ruling class. Thus their criticism contributes to an increase in the normative pressure felt within the society. The society becomes more and more coercive and its institutions malfunctioning. Hereby the ground is prepared for institutional changes.

V

Assume that it were true that *A* *q*'ed *because* he intended to *p* and thought *q*'ing necessary for this. What sort of connection does this "because" establish between an intention and epistemic attitude on the one hand and action on the other hand? This is a question on which philosophers violently disagree.

Some hold that the connection is causal. This position can be understood in two ways. I shall call them the trivial and the non-trivial.

The trivial understanding of the causalist position stresses the fact that intentions are quite commonly *called* "causes of actions". This is in order, and we should not attempt to reform language here. The only objection which one can have to this kind of talk is that it obscures the difference between the sense in which intentions can uncontroversially be called causes of actions and some other important senses in which things are said to be causally related. One of these other senses is often called "Humean". The existence of a Humean causal relation entails that there is a general law connecting instances of logically independent generic phenomena as cause and effect.

The non-trivial interpretation of the causalist position in action theory holds that a specific combination of intention and epistemic attitude is a Humean cause of a specific kind of action.

Defenders of this position sometimes think that it requires a reinterpretation of intentions and cognitive states in neurological terms. The causal relation is then in the first place between certain brain events and certain movements of limbs and other parts of the body. Of this view I shall here only say the following:

We need not doubt that there are causal relations of the kind just mentioned. But the neurological interpretation of volitional and epistemic attitudes is, at best, only a contingent correlation of them with equivalents in the brain and not a necessary connection. And whether the movements caused by those brain events are actions or not is again a contingent matter, depending upon other facts about the agent than movements of

his body and processes in his nervous system. These observations, when more fully substantiated, suffice, in my opinion, to wreck the non-trivial causalist thesis about the relation between intentions and actions. But I cannot argue the point at length here.²

The contrasting opinion holds that there is a conceptual or logical connection between an action and its grounds in an intention and epistemic attitude. This opinion is sometimes called the Logical Connection Argument or the intentionalist view. I think it comes nearer the truth than the causalist view. But it is difficult to argue for it correctly.

Thus I think it is a mistake—of which I myself³ and others have been guilty—to understand the intentionalist view to mean that there is a relation of logical *entailment* between the premisses and the conclusion of a practical argument.

Consider the following example.⁴—A man is firmly resolved to assassinate a tyrant. He has access to his room, aims at him with a loaded revolver—but cannot bring himself to pull the trigger. Nothing which we later find out about him would make us think that he had changed his intention or come to a different opinion about the things required of him to make it effective. Is this conceivable, i.e., logically possible?

It should be noted that we do not assume here that the assassin *in spe* forbore to pull the trigger. Had he done so, it would be, I should think, a contradiction to say that the man (still) intended to shoot the tyrant and knew what he had to do. (It would be a contradiction reminiscent of the so-called Moore paradox.) What is assumed is simply that he did not do anything at all just then which was relevantly related to his resolve to shoot the tyrant. He was “paralysed”—but neither physically nor mentally in a way which would make us revise the description of his intentions and cognitions.

In view of this example I think we should say the following about the nature of the relation between intentions and actions:

An intention and an opinion of what is required for it to become effective, constitute, as was already said, *a sufficient ground or reason* for acting accordingly. If the agent then acts accordingly, we understand completely why he is doing what he is doing, e.g., trying to kill the tyrant by firing at him. No further information can help us understand *this* better. (We may, of course, wonder why he should have had the intention he had or how it was that he thought as he did—quite wrongly perhaps—about the requirements of making his intention effective. But these questions do not concern the determinants of his action but the determinants, if there are any, of these determinants.) If again the agent fails to “act accordingly”

² Cf. *Freedom and Determination*, especially pp. 57–61.

³ Cf. *Explanation and Understanding*, pp. 97–118.

⁴ *Ibid.*, pp. 116f.

we do not understand him at all. His behaviour is incomprehensible to us and in this sense *irrational* or, considering that he had sufficient grounds for acting in a certain way, anti-rational.

The relation between what I have called internal determinants of an action and the action itself is thus neither a relation of entailment nor a causal relation. We must resist the temptation to reduce it to something which it is *not*. But there is a sense in which we can call the relation *conceptual*—and something remains to be said about this before we have a full understanding of its nature.

VI

What is action? One could answer: Action is normally behaviour understood, “seen”, or described under the aspect of intentionality, i.e., as meaning something or as goal-directed.

Intentionality can quite rightly be said to be in the behaviour. But not like a “quality” inherent in the movements of limbs and other parts of the body. For these movements we can describe completely without mentioning intentionality. So what then is the intentionality of behaviour?

To understand behaviour as intentional, I shall say, is to fit it into a “story” about an agent. We see a person walking in the street, carrying a parcel in his hand. He drops it and bends to pick it up. We should normally think of his picking it up as intentional. Why?

We may not know at all why he picked it up. But we can name hundreds of reasons why he might have done this, reasons which are such that had they been his at the time, they would explain his action completely. Perhaps the parcel contained something he was anxious not to lose; or a gift which he had bought for somebody. Or perhaps he took care not to litter the street or maybe was just following an existing regulation requesting one not to do this. We are, in other words, familiar with a number of possible, internal and external, determinants of his action here. We think it likely that some such determinant will be at work. This is what it is to “see his behaviour as intentional”.

There are on the whole reliable ways of coming to know, of verifying, what a person intends to do, what he thinks are the requirements of the situation, and whether he acts accordingly. Verification may not always be conclusive, and ascertaining one of the three things mentioned may sometimes have to rely upon the accepted verification of one or both of the other two. But normally the verificational procedures here are applicable independently of one another.

A standard way of ascertaining a person’s intentions and epistemic attitudes is by asking him. If we doubt whether his answers are reliable, there are usually other checks available. Ascertaining what a person does

again, is usually a matter of simple observation. We literally *see* him do various actions, i.e., we can describe his perceived behaviour under an appropriate aspect of intentionality. We may be mistaken, but normally we take what we see to be conclusive.

Relying on these verificational procedures we can establish a predictive correlation between the premisses and the conclusion of practical inferences. Having verified the premisses, we expect the conclusion to turn out to be true. The reliability of the prediction is in the following sense a function of *time* here:

When a person intends to do something, the object of his intention is in the future. This is so also when he intends to do something "right now"—for "now" then means the time immediately ahead of him. For this reason a person will normally have time to change his intention, and also his opinion of how he has to act, before he proceeds to making the intention effective. The more time he has to change his mind, the greater the risk that he will actually do so. If he does, the prediction may fail.

It is essential, however, that for short time intervals the reliability of the predictions should be high. This seems to be a feature of the way in which the *concepts* of intention, of the various epistemic attitudes, and of action are related to one another. If it were normally the case with a given person that he did not act in conformity with his shortly before professed intention and understanding of the situation, we should doubt either the veracity of his reports or doubt whether *he* knew what it *is* to intend and to believe something. And if this were the case with people generally, it would modify our view of what intentions and beliefs are. We could no longer be taught nor teach to others the present use of the words and therefore we should not *have* the concepts either, at least not in their present form. One could say that the language-games we now play with action-words and with epistemic and volitional terms rest on (presuppose) a high degree of correlation between intention and action in accordance with the understood requirements of a situation.

The prediction of actions from a background of intentions has a certain resemblance with the prediction of effects from knowledge of their causes. But there are also important differences to be noted.

One difference is this. The prediction of action is subject to a clause that no change occurs in the volitional and epistemic background before action is supposed to take place. The frequency of the failure of the prediction, with a given agent—assuming that his intentions and epistemic attitudes are known for certain—is therefore a measure of the degree to which he is, as we say, "capricious" or "unreliable" or even "irrational". The only *hypothetical* element involved in the prediction is, in other words, that a certain volitional and epistemic attitude of the agent should remain constant in an individual case and that the agent should not act "irrationally". This is different from a typical causal context in the natural sciences. Here the

failure of a prediction can always in principle recoil back also on a hypothetical *law* relating the cause to the effect.

In the case of predictions of actions there simply are no such "covering laws" to be confirmed or refuted. To say that such and such intentions and beliefs, assuming they do not change, will normally result in such and such behaviour is not to state an empirical generalization based on observations or experiments. It is to state a necessary truth to which anybody familiar with the concepts involved will agree off-hand. And therefore this truth is very seldom stated—except, perhaps, in philosophic debates.

VII

What is the nature of the relation between actions and their external determinants?

Assume that it were true that *A* *q*'ed *because* he had been ordered to do so. It is quite obvious that the connection between the determinant and the action cannot here be a relation of logical entailment. For, it would then be selfcontradictory ("unthinkable") that *A* had been given the order, had understood it and was able to carry it into effect, and yet did not do so. This, however, is far from unthinkable. Is the connection then causal? (Here it may be worth noting that it is much *less* natural to call an order a "cause" of action than to call an intention by this name.)

If the connection were one of Humean causation, there should exist a *law* connecting cause and effect. This law cannot be that *A* always obeys when he is ordered to *q*. Perhaps this is true of *A*. But most probably it is not true of everybody who has learnt to obey orders. It may, for example, not be true of *B*. Still it could very well be the case that *B* too, on some occasion, *q*'ed *because* he had been ordered to do so. So whether the agent concerned happens to be one who always obeys orders to *q*, or not, is quite irrelevant to the question of the nature of the connection, in an individual case, between the order and the action.

As far as I can see, the "because" does not rest on any law at all here. And if this is so, then the relation between the determinant and the action is not one of Humean causation. What then is it? I propose to call it a relation of *justification*.

Let it be that *A* answers when asked why he *q*'ed: "Because I was ordered". May he not be lying or even mistaken about his own motives? When pressed with further questions, he may admit that really he *q*'ed because he feared the anger of the order-giver, i.e., acted under the influence of normative pressure, and not just in response to the order. But if he does not admit any such other motive—not even "to himself", "in his heart"—then we must take him at his word and say that he *q*'ed because he was ordered. There is no "external" way of deciding the truth of the "because".

statement to which we could concede ultimate authority here.⁵ The connection between the external determinant and the action is, as I have said, not intrinsic in the sense that it were a logical entailment. But it is in a characteristic sense an "internal" relation, dependent upon the agent's judgement of why he acted as he did. Therefore it is not, in any good sense of the word, a "causal" relation.

Just as one can, within limits, predict the actions of an agent from antecedent knowledge of internal determinants, one can make predictions on the basis of knowledge of external determinants. The degree of reliability of such predictions may vary with the agents concerned and also with the society under consideration. One can use this degree as a measure of the responsiveness of an individual to external determinants (of one sort or another, or generally). For example, one could use it as a measure of his obedience to the law or to his superiors. One can also use it to measure the degree of internalization and of normative "cohesion" of a given society. The characteristics thus measured do not explain predictability. Predictability is their criterion. Nor is there any other general law besides a rough statistical correlation which connects the determinants with the actions.

It is of some interest to ponder why we do not willingly speak of such correlations as "laws". Is it because of their unprecise and statistical nature? Or because of their dependence upon individual agents and individual societies? An even weightier reason for not calling them "laws" is, I think, their dependence upon factors, *viz.* norms and institutionalized patterns of behaviour, which are themselves susceptible to change in the course of history as a result of human action. "Scientific laws", we tend to think, must not for their validity be dependent upon historical contingencies. They should hold true *semper et ubique*.

VIII

Determinants of action, I have said, are either internally set or externally given. By referring to their determinants we explain the actions, *i.e.*, answer questions why agents acted as they did.

Such answers, however, are only explanations "in the short perspective". They give rise to further questions. For example: Why do people have the intentions they have?

Sometimes the answer to that question is given in the terms of a further intention. Why did *A* intend to go to the concert tomorrow? The answer could be that he intends, is resolved, to acquire some education in music. Going to concerts here serves as a means to an end. But why should he be

⁵ On this question I have later come to change my opinion. See the essay "Explanation and Understanding of Action", below, pp. 64f.

resolved to pursue this remoter end? The answer in terms of intentions will ultimately take us, I suggest, to one of two main types of determinants of intentions. I shall call them *wants* and *duties*.⁶

Let us first consider wants. Why did *A* intend to go to the concert? A frequent type of answer could be: Because he wanted to hear *B* perform or because he wanted to hear symphony *S* again or simply because he wanted to hear some music. Such answers are normally complete explanations. One wanted *just this*, and there is no further thing for the sake of which one wanted it, as a means to an end.

If the further question is raised as to why one wants what one wants, the answer sometimes is: Because it pleases or one likes it or because it is thought fine or nice or amusing. But such answers do not point to determinants of the want. They merely specify it by setting it against a background which is there independently of my present intention to reach out for the wanted thing. If, for example, I like to listen to music, I shall, given an opportunity and in the absence of any other determinant already at work on my intentions or actions, listen to music. On such an occasion I might say "I want to hear some music, I like music." My liking of music is, so to speak, a latent want which manifests itself in my intention, say, to hear some music now or to go to the concert tomorrow.

One cannot ask why people should want things they like or take pleasure in. It is, one could say, "in the nature" of pleasant and liked things that they should be wanted—as it is "in the nature" of unpleasant and hurtful things, such as illnesses or punishment, that they should be shunned. Shunning things is wanting not to have them or wanting to get rid of them. When we intend (decide) to follow a rule or order because we are anxious not to risk punishment, it is our shunning of something intrinsically unwanted that determines our intention.

A person says: "I intend (have decided) to go to the Canary Islands during my vacation." "Why do you want to go there?", we ask. We are anxious to know the *want* behind his intention. He answers: "Oh, I just want to see the place, it is supposed to be nice." This could be his sole motive force (want) here. But assume he answers: "I think it will do me good, I have been very tired and run down lately." Then the planned action is seen as a means to an end, the end being one's health or well-being generally. This is what one wants to promote.

Health, well-being, and happiness are "natural" objects of want. Other things being equal we pursue them of necessity, for their own sake. In this they resemble the things we like or which give us pleasure. It makes no sense to ask "Why do you want to be healthy?" But it does not follow that a person will necessarily care for his health, or pursue his pleasures or happiness. There can be overriding considerations. He may have "no time"

⁶ For a fuller discussion, see *Freedom and Determination*, Ch. IV.

for his pleasures or "be forced" to neglect or even to ruin his health. The determinants which can override the influence on our intentions and actions even of things which are "by nature" objects of wants, have the character of *duties*.

IX

The word "duty" should here be taken in a broad and somewhat loose sense.⁷ The range of things which I have in mind can be roughly characterized as follows:

As a member of a society any man usually holds one or several positions in which he is expected, or sometimes even obliged, to do various things. Some such position a man can be said to hold "by nature", such as the position of a parent; others he holds, e.g., by appointment or by election. But in either case the actions or types of action expected of him are defined by the explicit or implicit rules (laws, customs, conventions) of the society to which he belongs. I shall call such positions *roles* and the things expected of a *role-holder* *duties*. (The etymology of the word then suggests that they are things which he "owes" to the rest of the society by virtue of his position in it.)

Thus a head of state is expected to care for his country's prestige, its power and prosperity. This will make him form intentions and take decisions which, as a "private citizen", he neither could nor would contemplate. The objects of these intentions form part of what he and others consider his duties. Failure to perform need not have legal implications for him, but will surely have consequences which it is in *his* ("personal") interest to shun, such as loss of popularity or an unfavourable "verdict of history". So, failing a motive "from duty", there will be a motive "from want" ("self-interest") to make him have the action-guiding intentions which are appropriate to his role.

A policeman is seen jumping into a car and speeding away. Why this behaviour? We are told that he intends to catch the thief who was seen running in the street. Why should he intend this? As a private citizen he may even have felt inclined to let the poor man escape. But his role as policeman "imposes" this intention on him with all the actions following upon considerations about the means of making the intention effective. If he does not realize this and act accordingly, he runs the risk of being fired or even punished.

Similar considerations apply, *mutatis mutandis*, to all holders of roles. It is an essential part of the picture here that roles should be surrounded by an aura of normative pressure which, when needed, makes people

⁷ For a different notion of "duty", cf. *The Varieties of Goodness*, Routledge & Kegan Paul, London, 1963, Ch. IX.

perform in their roles—perhaps somewhat “against their will” but still in agreement with what they want, lest something worse happens to them. In this regard there is a parallel between roles as external determinants of intentions and rules and symbolic challenges as external determinants of actions. But as with rules it is also the case with roles that it is essential to our notion of a functioning society that role-performance should on the whole *not* be motivated by normative pressure, but should be “internalized”, that is: the duties unquestioningly accepted as an ultimate determinant of what we intend.

It is perhaps right to say that duties implicit in various roles more than any other determinants mould men’s intentions and therefore indirectly guide their actions. But the extent to which this happens is different within different societies and with different roles. Therefore questions of role-distribution within a society are inseparably connected with the problems of individual freedom.

So-called “free-time” is that part of a man’s life when he can do what he *wants* and temporarily forget the demands on him of at least some of his assigned roles. When a man has no time for his wants, only for his duties, he is a slave to his roles. This he can be both as servant in the meanest position and as a master endowed with the greatest power.

X

In addition to wants and duties there is also a third type of factor which determines a man’s intentions—and through the intentions his actions: his *abilities*.

Unlike wants and duties which “prompt” people to action, abilities determine actions negatively, restrictingly. They delimit the “horizon” or “domain” or “range” of a man’s freedom to act. This range will then wax and wane with variations in ability.

To have an intention to do something presupposes that the agent thinks, rightly or wrongly, that he can achieve the object of his intention. What he does not think he *may* accomplish, he will not attempt either. To say this is to make a *conceptual* observation on the relation between the volitional and epistemic attitudes involved in a practical argument. There is not, be it observed, a corresponding relation between “want” and “can”. A man may want to do things he knows he cannot do. But if his want is not to remain merely an “idle wish”, he will have to form an intention to acquire the ability. He may, for example, be resolved to learn to do it. This again presupposes, logically, that in his subjective estimate, he *can learn* to do the thing in question.⁸

Abilities are either “innate” or else determined by biological and physical

⁸ On the relation between want, intention, and ability cf. above pp. 6ff and pp. 20ff.

factors, or they are acquired. Intelligence and memory, health and bodily strength are gifts of nature—and nature endows people *unequally* and also makes the abilities which depend upon these endowments *vary* within the lifespan of the same individual. But within rather broad limits people are *roughly* equally endowed in these regards, and the differences which exist between them can to some extent be equalled out as a result of care or training.

Acquisition of abilities happens through learning, instruction, and education. These are largely socially institutionalized processes. But new abilities are also acquired thanks to individual inquiry into the possibilities of doing things and thanks to creative efforts. Of abilities, thus acquired, those which have the character of *technological innovations* occupy an important and peculiar place, chiefly because of their consequences on the social level.

The fact that learning and education are parts of the social fabric can be responsible for great inequalities in the possibilities which men have of acquiring abilities. A man may not be able to afford or his social position may not allow him to avail himself of facilities for education which some other men enjoy. The range of things he is able to do may on this account remain very restricted. Then it is also likely that within this range his duties, "the pressing necessities of life", much more than his wants, will determine his intentional actions.

There are thus a great many abilities which a man "by nature" could acquire but which for deontic reasons, i.e., reasons built into the normative structure of the community of which he is a member, he cannot acquire.

Many abilities, moreover, require for their exercise *equipment* in the form of instruments or machinery. This is true of all abilities which are conditioned by technological innovations, but particularly of the ability to use technology for the production of commodities. This fact is the greatest source of inequalities in the freedom of men and also the greatest urge to make men equal by changing the institutionalized restrictions on what each one of them *can do*.

The prospects of social philosophizing which open up here are easily recognized—but their further exploration must remain outside the scope of this paper.

XI

What a man can do in a given situation is, however, only partly conditioned by his abilities. An equally important condition is formed by the *opportunities*. A child may have learnt how to open a window, but if the windows in his surroundings are already open, it *cannot*, in that situation, open a window. The ability is a generic feature of an agent; the opportunity, again, an individual feature of a concrete situation.

Every action by any man creates and destroys opportunities for actions—by the agent himself and by other agents. By shutting a door I create an opportunity for opening it; by leaving the room I may destroy the opportunity for another man to request my help in an important job.

The opportunities are thus in a constant flux as compared with the *relative* stability of abilities and wants and duties—not to speak of their background in the institutions of society. Intentions fall in a middle zone here. As the situations change, creating new opportunities for action, intentions articulate under the already existing wants and duties and within the frame of given abilities. This interplay between situational change, intentionality, ability, and a motivational and normative background I shall call the *logic of events*. It constitutes the cogwheels of the “machinery” which keeps history moving.

Sometimes the changed situation which makes new actions possible, or imperative, results from the working of natural forces alone. This is the case when, for example, an earthquake or a flood upsets human conditions. The intentions for acting formed under the impact of such changes are often the outlet for wants (and shunnings) shared by practically all men at all times and which might also be called a “will to survival”. People seek refuge or migrate to new abodes—or they join hands to take various countermeasures such as building walls against floods or protecting the environment against industrial pollution. Such measures may in their turn require (in the sense of the second premiss of a practical argument) changes on the institutional level of society.

A very different type of logic of events is exemplified by changes in situations which primarily result from people acting in roles, for example when the actions are those of a government, or a corporation, or the army, etc. A country conquers a province from another to safeguard its borders. “Unless we do this, they might invade us” is now the reason-giving premiss of a practical syllogism. In the new situation, created by the conquest, the neighbouring country, i.e., its government, thinks it necessary to safeguard its independence by entering into an alliance with a third power. The two now jointly constitute a threat to the first and further actions become imperative for its rulers. And so forth. Each new action by one party “triggers off” the conclusion of a “latent” practical argument by another party—the “latency” of the argument consisting in the fact that the goal-structures, both the duties of the role-holders and the wants (“hopes and fears”) of those who vest their expectations in the role-holders, are fixed, and the requirements of the situation appear univocal in the light of past experience or of traditional standards of assessing them.

Such chains of successively created sufficient reasons necessitating action are particularly impressive when events are heading for disaster or towards decay. The origins of imperialistic wars, the fall of empires, the decay of an economy often follows this pattern and thereby assumes an

Again a somewhat different type of "logic" is presented by the great creative innovations, particularly in the realm of technology, which open new possibilities of action and thereby become an outlet for latent wants rather than for pre-existing institutionally determined duties. The chain-reactions "let loose" by such changes often create a tension between the duties of role-holders and external objects of intention set by the rules on the one hand, and the direction given to the action-opportunities by the flux of situational change on the other hand. Internalization of the institutional forms becomes more difficult and more dubious, normative pressure increases and the "internal contradictions" of the community crystallize into grounds for changes in institutions.

XII

The description which I have tried to give of the motivational mechanisms and the working of chains of necessitation of action should help us answer the question whether there are "laws" in history.

In fairness to the question it should be said that if there existed law-like connections between concrete historical events, we should rather think of them as instantiations of general laws of sociology, and perhaps of economics, than as "laws of history" proper.

A number of so-called laws of economics easily come to the layman's mind: Say's or Gresham's laws, the law of supply and demand or of diminishing marginal utility, etc. In sociology there is much less unanimity about what deserves the name of a law—but candidates are not lacking, for example the several Marxian principles concerning the dependence of social structure upon productive forces and relations.

My suggestion now is that such laws are applications to specified types of activity and types of historical situation of the very general conceptual patterns which I have outlined in my paper. Even the most elementary laws of economics presuppose some institutionalized forms of exchanging commodities on a market and of rough standards of measuring the value of goods to a producer and a consumer. It is usually not too difficult to see under which assumptions concerning the institutional frame these laws are conceptual necessities about the way in which wants and abilities regulate behaviour. Within different frames different laws are valid. This means that different frames require different laws, if the logic of events is to be correctly described. The complexities of theoretical economic analysis largely consist in devising conceptual instruments appropriate to the description of economic behaviour within the institutional structure of a historically given situation. Thus, for example, in the strongly "manipulated" market of late capitalist societies the laws of "classical" market economy cannot be expected to hold good. For this reason the

said that the laws of economics and sociology are themselves subject to historical change—unlike the laws of nature which are valid *semper et ubique*.

Social “laws” are not generalizations from experience but conceptual schemata for the interpretation of concrete historical situations. Their discovery, or rather, invention, is a matter of the analysis of concepts and their application a matter of analysis of situations. On this account one can say that social study occupies an intermediate position between philosophy and history. It can move in the direction of the one or the other of the two poles, but it cannot live a self-contained life divorced from either of them.

History, when it is “scientific” and not mere chronicle or narration, is an inquiry into the logic of events in a fragment of the past with named actors and institutions. It is a study of history from a deterministic point of view in as much as it studies the interplay between historical change and the determinants of human action. As we have seen, these determinants have to a great, not to say overwhelming, extent their roots in the structure of the social fabric: in the distribution of roles and the institutionalization of behaviour-patterns. With changes in these societal determinants of actions, actions too will be different. But changes in the determinants are in their turn the results of action—except for the cases when they are man-independent changes in nature. Thus the actions of men are determined by their historical situation, but the historical situation is itself the result of the actions of men. There is no circularity of a logically vicious kind in this fact that mankind is both slave and master of its own destiny.

The determinants of natural change are causal laws—and them man cannot change. But he can use his knowledge of the laws to steer natural change by producing and suppressing opportunities for causes to work. Man’s foresight, however, is limited and what further causal consequences his manipulations of nature will call forth may be humanly impossible to foresee. We are reminded of this by the eroded landscapes in lands of ancient cultures—but also by the ecological problems facing modern industrial society. That man has made himself master of nature to the extent he has is one of his greatest achievements as a species. To exercise the restraint and skill needed in order not to be dethroned is the most serious challenge facing him today. It is unlikely that it can be successfully met without profound changes also in that law-regulated realm in which man’s mastery can never be challenged and where he is for ever sovereign, viz. his societies.

Explanation and Understanding of Action

I

To explain an individual human action is, one could say, to give a truthful answer to the question *why* the action was undertaken. "Why?" means "from what motive?" or "for what reason?". *Can* it also mean "from what cause?"? In a trivial sense, certainly. Motives and reasons are often called "causes" of actions. But in a non-trivial sense the question is problematic. The problem will be discussed later. Let it, however, be clear from the beginning that by "cause" I here mean what is also called a *nomic cause*. A nomic causal relation between two terms, *C* and *E*, should satisfy *at least* the following two conditions:

- i. *C* and *E* are *logically* independent of one another.
- ii. *C* and *E* are connected by a *universal law*. Whenever *C* occurs (in the frame or setting of some circumstances *F*), *E* follows.

II

It is worth saying a few words about the *motives* for action explanations. Why do we ask, why this or that was done?

The reason is seldom a "scientific" interest in action. Action explanation is not a species of "scientific explanation". Its motive is usually a wish to *evaluate* the action or the agent. Does what he did deserve praise or blame? We may not know until we have investigated the agent's motives. How does the action reflect on his character? What can we expect of this agent in the future? Is what he did perhaps something we should do ourselves once we know the reasons *he* had for doing it? Such questions reflect typical interests behind our ques' for action explanations.

There is, however, one "science"—if it can be called by that name—in which explanations of individual actions play a role. This is the study of history. Why did Caesar cross the Rubicon? Why did Queen Christina of Sweden abdicate her throne? These may be interesting questions. But as a species of historical explanation action explanation is not, I should say, of great prominence.

III

Reasons for actions have been extensively discussed in recent philosophic literature. But to the best of my knowledge, no systematic inquiry has been made into what might be called "the varieties of reasons". What *sorts of thing* count as reasons for actions?

Many actions are done *for the sake of an end*. The end is something the agent intends or wants to reach (produce, secure, enjoy). The action is thought to be either needed for or conducive to or in some other way "promotive" of the end. Such a volitive-cognitive combination of a "will" to an end and the thought of some means to its realization constitutes a *type of reason for action*.

For example: I hear voices in the corridor outside my room. I want to know what is going on. This constitutes a reason for me to open the door and look out. I believe that this will help satisfy my curiosity.

Reasons of this type I shall call *inner* or *internal*. They are *necessarily* reasons for action. By this I mean the following: No-one who is familiar with action discourse could, without committing an inconsistency, deny that aiming at something and thinking a certain action promotive of this aim is *a* reason for doing it. There may exist reasons *against* doing it—or one may know better means to the end.

Other actions—perhaps the majority—are done *in response to a* (symbolic) *challenge*. Two sub-types of challenge may be distinguished. One consists of challenges presented in what may be called *communicative action* patterns. Examples are orders, requests, questions. Traffic lights and (many) other signals also belong here. The other type consists of (prescriptive) rules or norms, and of norm-like things such as customs, fashions, or traditions within a community.

Example: Why did you stop your car? Because the traffic light turned red. Why do you stop in front of the red light? Because there is a rule prohibiting one to drive against this signal. Such answers give reasons for action.

Challenges I shall call *outer* or *external* reasons for action. Unlike internal reasons, challenges are *contingently*, and not necessarily, reasons. This means the following: Even though an agent recognizes the challenge and has learnt or otherwise knows how to respond to it, he need not *acknowledge* it as a reason *for him* to act upon. External reasons can thus be said to "exist" in two different senses. As *instituted* and presented to members of a community they exist, so to speak, "objectively". As *acknowledged* by individual agents as reasons for their acting they exist "subjectively". Their subjective existence cannot be inferred, in the individual case, from their objective existence.

IV

Agents are taught to participate in communicative action and to obey rules. When they are thus brought to acknowledge outer reasons as reasons for their (own) actions, the reasons will be said to be *internalized* with the agents in question. In order to understand how internalization comes

about we must note the following connections between outer and inner reasons:

Outer reasons are often, though certainly not always, instituted to serve some ends. If the agents to whom the challenges are addressed share those ends, the agents have inner reasons for responding to the challenges. Example: Traffic rules are there to secure safety on the roads. Whoever wants to contribute to this end, has an internal reason for obeying the rules.

Such internal reasons may not exist or may not be effective for all agents, however. Therefore external reasons of a secondary type will have to be instituted which threaten the agents with some evil (coercive measures, punishment, reprobation, sanction) in case of non-conformity. Challenges and rules are thus being surrounded by what I have elsewhere¹ called an "aura of normative pressure". Agents who intend or want to avoid these evils have a secondary type of internal reason for conformity.

Internalization has taken place when the agent acknowledges an external reason as a reason for him to act on, independently of the "pull" of social ends or the "push" of normative pressure. Internalization admits of degrees and may be more or less well accomplished. In the individual case it may, for example, be difficult or even impossible to tell whether the agent obeyed an order *because* he had been ordered, or *because* he feared punishment for disobedience. His motives might have been "mixed". But it would be a distortion to think that his action *must* have had internal reasons and *could not* have taken place on purely external grounds.

V

Often reasons which explain actions are overtly neither of the kind I have called inner nor of the kind I have called outer. Such reasons I shall call "oblique".

Examples: Why did you take an overcoat when you went out walking? Because the sky looked threatening. Why did you bring your wife flowers? Because it is her birthday. These are explanations in terms of reasons. Normally we should understand them and regard them as exhaustive. But in order to understand why this is so we must establish a relationship between those reasons and reasons of an overtly internal or external nature.

The sky looked threatening and there might have been a rainstorm and I should have become wet. Getting wet is unpleasant, something I *shun*, i.e. want to avoid. My raincoat will protect me. So, I have an internal reason for taking it with me.

¹ See above, p. 39.

It is my wife's birthday—I want to do something which will please her. Bringing her flowers will make her happy. This constitutes an internal reason for my action. But also: It is customary to honour people on their birthdays. I am anxious to behave politely; I acknowledge the external reason constituted by the custom. Was it the inner or the outer reason that moved me to action? Perhaps my motivation was a "mixture" of both.

Compare: "I screamed because it hurt so badly" and "I shouted because I got angry with him". In the first case we should perhaps say that I could not help doing what I did. My screaming was more like a reflex than an action; the pain a cause rather than a reason. The second case *could* be like the first. He had been teasing me for a long time and at last I "exploded", could no longer control myself. Was this an action? But perhaps I thought "This cannot go on, I must stop him", and therefore I shouted at him angrily. Then what I did was clearly an action of mine and I had reasons for it.

VI

So far we have only spoken of reasons *for* an action. Reasons, however, may also be *against* an action. A reason against doing something can be defined as a reason for omitting the thing; reciprocally, a reason for doing is a reason against omitting.

Even though the two types of reason (*for* and *against*) are thus inter-definable, what we ordinarily think of as reasons against are considerations pertaining to some action which the agent *also* has or may have some reason for doing. Reasons against are typically "second thoughts". They are thoughts of something we either *shun* or *ought not* to do and which are consequences of or prerequisites for a contemplated action (*for* the doing of which something speaks).

Things we shun constitute *inner*, things we ought not to do *outer* reasons against doing something.

VII

When an agent had several reasons for an action of his, and no reason against it, his action will be said to have been *overdetermined*. When this is the case, one cannot tell for which reason the action was undertaken. A truthful answer to the question *Why?* will have to mention *all* the reasons actually present.

Of more interest is the case when there are reasons for but also against the action. In many, but not in all, such cases there is *deliberation* terminating in a *decision* to act.

Examples: I am about to go for a walk; shall I take an overcoat? It is heavy to carry—this is a reason against taking it. But it may be raining—this is a reason for taking it. I take it. *Why?* Because I would rather suffer

the discomfort of carrying the coat than risk the discomfort of becoming wet. I *prefer* the first to the second.

Shall I visit my old mother this afternoon? She is expecting me; I have a duty towards her and I acknowledge this. Or shall I take part in the excursion which has been announced? The weather is fine and the company will please me. Here an outer reason for a certain thing is balanced against an inner reason for doing something else. This is the classic case of conflict between "duty" and "inclination".

Assume that duty conquers inclination. This may but need not mean that I want, or like better, to do my duty, i.e. that some inner reason "behind" the outer one carries the victory over another inner reason. But it would be correct to say, in any case, that I *prefer* to do my duty rather than follow my inclination.

Is a preference a reason (for action)? To call a preference a reason seems to me highly misleading. In the normal cases, a preference goes together with the existence of several reasons: some for, some against a certain action or some for one, others for an alternative action. These reasons are being "weighed", taken into consideration, and a "balance" formed. The preference is the resulting balance—not an additional reason to be thrown into the scale, so to speak. Therefore one should not call a preference for a certain course of action over another a reason for choosing the first.

How is such a balance formed? One can think of this as a decision-theoretic problem: we assign probabilities and values to the outcomes of alternative courses of action and calculate the "expected utilities". *Sometimes* something approximating to this takes place in real life. But the peculiar talent required for making choices which also in retrospect will seem to us the best or right ones is a "practical wisdom" without which decision-theoretic calculations, when at all possible, will be of no or little avail.

VIII

I began by saying that to explain an action is to answer truthfully the question why the action was performed (p. 53). When the motivation for an action is, in the sense explained above, *complex*, we cannot in a straightforward manner answer the question *why* the agent did it or say that he did it *because* he had such and such reasons for and such and such reasons against doing it. Another agent may have articulated a different preference on the basis of the very same set of reasons—and accordingly acted differently,

These limitations on action explanation are, I think, significant. It is a prejudice to insist upon answers to the question *why* or upon accounts in terms of *because* in the case of all *actions for reasons*. Many actions have a

complex motivation which makes such accounts impossible. Still we *understand* the actions in the setting of their motivation. They are not "incomprehensible" or "unintelligible"—and *in that sense* not "inexplicable" either.

There is, however, another and entirely different case when the question why an action was done has no answer. This is the case of fortuitous or gratuitous action. I did something, but I did it *for no particular reason*. Yet, it was an action of mine—I am not willing to say in retrospect that I "could not help" having done it. I may even say that I then intended or wanted to do just this—though for no particular reason.

Cases of this sort occur. But they are relatively rare and we sometimes find them puzzling. If they occur repeatedly with an agent, we may look for "subconscious reasons" or try to find "medical causes" for his behaviour.

If we could not, *on the whole*, account in terms of reasons for what people do, it would be difficult for us to understand them *qua* agents. If this were the case with ourselves, we should cease to feel responsible for our actions, since we could not then on the whole account for them. We should perhaps think that we are at the mercy of uncontrollable outer or inner forces—maybe of a causal nature.

Free action and action for reasons are twin concepts. "Determination" of action through reasons is, one could say, a precondition of human freedom. Without this type of determination our very notions of agent and action would not exist, or be quite different from those we have.

IX

In spite of the above "limitations on action explanation" I therefore think it is right to say that there is an answer to *most* questions why, i.e. for which reasons, this or that action was undertaken. A good many of the actions we perform we *learn* to do by being trained to respond adequately to various challenges or conform to rules. When asked why we did this or that we often *justify* our action by relating it, as an adequate response, to a challenge or a rule. A justification of action has the form of an explanation (in terms of reasons) of the action. But it need not be a correct explanation. There is an important discrepancy to be noted between justification and explanation:

The reasons given in justification of an action can be there, exist as a matter of objective fact, and yet not be the reasons why the action was undertaken. For example: I pay a visit of condolence to Mrs A who lost her husband. Why? It is customary to do such a thing when one is a close acquaintance of the family. The custom justifies my action. But it need not be the reason why I paid the visit. Perhaps the reason was that I had *selfish* designs on the rich and beautiful widow.

If I acknowledge the custom as a reason for my visit but also had such designs, my justification of the visit was a *partial* explanation. A *full* explanation would then show that my action was overdetermined.

If, however, I had the selfish motive but do *not* acknowledge the custom as a reason, then the action had an acceptable justification which was *not* even partially a valid explanation. That I do not "acknowledge" the custom as a reason means the following: I have been taught that one is expected to observe this custom and recognize that many people do this but I should myself never conform to it unless I had *some independent reason* for doing so. If I conform and say that I do this because it is the custom, I am hiding the true motive of my action. I hide it from others—but maybe I even hide it from myself.

When *is* a justification an explanation? The answer obviously is: When the reason *R* given in justification of the action not only was there but moved the agent to action.² When it is true to say of the agent that he did it *because of R*.

But when *is* this true? To this question some philosophers would wish to answer that the statement is true when (the existence or occurrence of) *R* is, or is somehow "correlated" with, a *nomic* cause of the action. Those who give this answer are sometimes said to adhere (in one form or another) to a *causalist* theory of action. I shall try to show why I think this theory will not help us solve the problem.

X

It is useful to distinguish two aspects under which the relation between an action and its grounds may be viewed. I shall call them the aspect *ex ante actu* and the aspect *ex post actu*. The two aspects answer, roughly, to the distinction between *prediction* and *explanation* of action.

Prediction is *ex ante*: Given causes or reasons, we predict what will take place. For example: *A* intends to go to the theatre next Saturday. He knows that, unless he books in advance he will not get a seat. Therefore, we predict, he will take steps to book in advance. Or: *A* has been asked to report to the police within 48 hours. We predict that he will do so.

Explanation is *ex post*. The action is a *fait accompli*. We look back for causes or reasons. *A* turns up at the booking office. Why? Maybe he

² The terms "reason" and "motive" (for actions) are commonly used as rough synonyms. One could, however, try to separate their uses, and such separation might serve a philosophic purpose. One way of separating them would be as follows: *Reasons* are given in *justification* of an action—*motives explain it*. The motives, in other words, are those reasons which actually "move" the agent to actions. Other reasons which are objectively present but are not subjectively acknowledged are not motives. I am indebted to Professor Chaim Perelman for conceptual observations in this area.

intends to attend Saturday's performance and thinks it necessary to book tickets in advance.

In writings on the philosophy of science it often used to be said that the predictive and the explanatory relations are *converses* of each other. What is prospectively a prediction is, if true, retrospectively an explanation, and *vice versa*. This view is an oversimplification even in the natural sciences. But it holds true, *by and large*, for prediction and explanation which subsume an individual case under a general law. Therefore it is true, on the whole, in the natural sciences. In my opinion, the situation in the human sciences is quite different. Many of the difficulties and obscurities in action theory arise from failure to appreciate the conceptual difference between the predictive and the explanatory point of view.

XI

If the *explanans* of an action is a (nomic) *cause*, then from knowledge of the presence of the cause one can predict the action "with certainty".

Predictability, however, does not presuppose causal laws. Reasons which are not causes may also provide good grounds for predictions. What is then predicted is, strictly speaking, not the *action*. In the case of internal reasons, the prediction is that the agent will not "change his mind" before proceeding to act, i.e. will not give up his intentions or alter his opinion of how to make them effective. If the reasons are external again, the prediction is that the agent will *acknowledge* such and such reasons, the presence of which is an objective fact, as reasons for him to act upon. Such a prediction does not rely on a *law* connecting the reasons with the action; but it may rely heavily on our *knowledge of the* (particular) *agent*.

The reliability of predictions of this kind may vary from agent to agent, and with the same agent in the course of his life history. When the motivational situation is simple, for example because all the known reasons point in the same direction, reliability may be very high. When again the motivational web consists of a great many reasons for and also against the action, prediction may be next to impossible. Another factor which affects the reliability of predictions is whether the motivational situation is "typical", i.e. in relevant features like a situation in which the agent has repeatedly been before, or whether it is a rare or even unique case, or a case in which it is difficult to tell whether the *known* reasons (for and against) the action are *all* the reasons which there in fact are.

The predictive relation which we are considering thus holds between a compound of *reasons* and (other) *facts* about the agent and the circumstances of the case on the one hand and the agent's proceeding to *action* on the other. I shall call these three factors *R*, *F* and *A* respectively. Our problem is: *Can* this relation acquire *nomic force*, i.e. can there exist a

causal law connecting *R* and *A* as cause and effect in the frame or setting of the circumstances *F*?

For example: Someone at the dinner table asks me to pass him the salt.³ The request is an outer reason for me to do something. Can this outer reason in combination with facts about me (my abilities, character, education, position at the table) nomically "force" me to proceed to a certain action?

One can discuss this question under two aspects. I shall refer to them as the "macroscopic" and the "microscopic" aspect respectively. The first concerns the existence of something which might, in some broad sense, be called a *psychological law* connecting *R* and *A* within the frame *F*. The second concerns the existence of a *physiological law* connecting events or states in my neural system, somehow "corresponding" to *R*, with bodily movements of mine, somehow "corresponding" to *A*? We shall consider in turn these two aspects of the problem.

XII

Let us assume that the description and identification of *R* causes no problem. We may think of some very simple case when there is only one reason present. How can we then specify the frame of accompanying circumstances *F* so as to make the purported psychological law testable? How can one make sure, for example, that the agent does not suffer from some "inhibition" which will prevent him from proceeding to the action *A* although this would be the adequate response in the presence of *R*? Can one be confident that no such "counteracting" factors are hidden in *F* except by making the action *A* itself a criterion of this? In other words: how can we ensure the *logical independence* of "cause" and "effect" here? It will suffice to pose the question—leaving open the answer.

Assume next that the specification of *F* is no problem, but that *R* is a compound of reasons for and against *A*. The agent has to form a "balance" of reasons before acting. Must not this balance, once formed, be a commitment either to do or to omit *A* such that if the agent, in the first case, does not proceed to action or, in the second case, proceeds to action, we must deny that he was committed? The difficulty, in other words, is again that of securing the logical independence of cause and effect.

Assume, finally, that the specification of neither *R* nor *F* is problematic. An example could be the appearance of a traffic light under entirely normal circumstances of driving or walking. The adequate response seems to follow without exception. If wanted, we can test the regularity experimentally. Could we not then say that the reason *R* under the circumstances *F* causally necessitates the action *A*? So that, viewing things

³ Cf. above, p. 38.

prospectively, the agent cannot help initiating action or, viewing them in retrospect, he could not have acted otherwise? Is this possible?

I think we ought to be generous and admit that such cases are "just possible". But they are almost certainly *marginal*.

Suppose somebody suggested that they are marginal only in relation to our limited knowledge. In fact *all* cases are like this. If only we could specify *R* and *F* with ideal completeness it would necessarily be the case that those two components taken together causally necessitate (the initiation) of *A*.

The one who suggests this, however, has shifted the point of view from *ex ante* to *ex post actu*. Not the law itself but the *existence* of a law is now a tautologous after-construction. Whenever an agent proceeds to action we think that this could not have happened had it not been for some reasons which, under the circumstances, causally necessitate the action. But we do not claim to be able fully to specify which these reasons and these circumstances are and therefore do not claim to be able to test the causal law either. Those who are intellectually wedded to "a causalist point of view" may draw a certain amount of comfort from this "existentialist" construction. Those who are not thus prejudiced will think it idle.

XIII

A consistent statement of the "microscopic" point of view must, I think, be given in terms of external and internal stimulation of the neural system resulting with nomic necessity in certain bodily movements. The description of the stimuli and the movements should moreover, be "drained" of intentionality, i.e. the stimuli should not be described in the language of reasons, nor the movements in the language of actions. But there ought to exist a "correspondence" between the two levels of discourse, the "actionist" and the "physicalist" one. Some would prefer to speak of "alternative descriptions" of the same thing. The criteria of correspondence and of sameness constitute difficulties for this approach. We can, however, bypass these problems here.

We need not doubt that there exist nomic connections between stimulations of the neural system and bodily movements. Nor need we deny that such connections "work" when action for reasons takes place. That a causal law "works" or is "operative" shall mean that the cause is present within the frame of circumstances under which the law holds true. In what way would these causal connections now *explain* how the action is connected with the reasons?

If there are reasons for and against the action, a balance of reasons will have to be formed before the agent proceeds to action. If there are only reasons *for* the action, but the reasons are external, the agent will have to *acknowledge* the reasons (or some of them) as reasons for *him* to act upon. In neither case will the mere *existence* of the reasons move the agent. He

will, in forming a balance, have to "take a stand" on them or, if they are external, "make (some of) them" his reasons in proceeding to action. Only when this has happened can we refer back to the reasons in an explanation of the action. But how does *this* happen? In what does the "making" the reasons reasons (for him) consist?

I submit for consideration that it is an urge to answer these questions that is the motive force behind the "microscopic" version of a causalist theory of action. To the reasons becoming operative there corresponds a "click" in the brain, the neural system, which is released by some stimuli corresponding to the reasons and which results in some movements corresponding to the action. Does such a "click" occur in the brain whenever an agent proceeds to action? We surely do *not know* whether it does—but perhaps we can imagine the possibility. I say "perhaps" because it is not entirely clear what we are supposed to "imagine". Granted that we can do this and that the possibility is an actuality, *must* it be this? That is: must it be the case that when an agent proceeds to action for such and such reasons there is a "correlated" physiological law simultaneously operating? Here the difficulty is with the "must". *In what sense* could this be a requirement? Merely postulating the correspondence seems an idle move—just as postulating a psychological law connecting the reasons with the action seems idle (cf. above p. 62). The answer must be that we could not otherwise *understand* how the reasons connect with the action.

Assume, for the sake of argument, that there is this parallelism between acting for reasons and the operating of physiological laws. Does this mean that an agent could not have acted for *those* reasons had it not been for the operating of *that* law? Or shall we say, on the contrary, that the law would not have operated had the agent not acted for those reasons? There is, I think, a presumption—deeply ingrained in and strongly nourished by the "scientism" of our culture—to say the first. But who can tell that this is not sheer prejudice and superstition?

My conclusion then is that the question of the "click in the brain" is *irrelevant* to action explanation. The truth of the statement that an agent did something for certain reasons is *independent* of the truth or falsity of the statement that to the existence of those reasons there "corresponds" events and states in the agent's body nomically connected with some bodily movements "corresponding" to the performance (or initiation) of the agent's action.

XIV

An agent did something, *A*, and maintains that he did this for the reason(s) *R*: If he is not lying, i.e. saying what he does not believe to be true, then in explaining or justifying his action he evinces his *self-understanding*. The reasons he gives constitute a "perspective" in which he sees his action.

Assume, however, that we doubt the agent's own account of the case. That is: we do not doubt that he did the action, nor that the reasons were "objectively" there. But we doubt whether he acted for *those* reasons.

How would one try to settle such doubts? Certainly *not* normally by examining his neural system. (In exceptional cases of a pathological nature a neurological examination may be relevant.) What one would do is to examine the agent's motivation in a *wider setting* either of his life history or of the present circumstances or both. We should press him with further questions, call in "witnesses", study records of his past.

Consider again the visit of condolence (above p. 58f). In the light of what we know about the visitor's character and conduct in the past, and also maybe in the light of things which took place *after* the visit, we may be convinced that he acted from selfish motives and not out of respect for the custom. The "coherence" of the picture we form ourselves of him "requires" this interpretation of his action.

When the agent is confronted with the way we understand him he may *confess* that he had been lying to us. But he may also be led to scrutinize himself and come to see that he had, in fact, been lying to himself, too. By way of his self-scrutiny and, maybe, under the influence of our opinion of him and of further facts about him to which we drew his attention he is *converted* to a new understanding of himself. Perhaps he says: "had I been sincere I should have admitted that I did this for selfish reasons—observance of the custom was only a pretext".

But what if he stubbornly refuses to see his action in the light we are convinced is the right one? Perhaps we say: "his lips refuse to confess—but in his heart he admits the truth". With what right can we say this? Hardly with any right at all, unless we foresee the possibility of a conversion. But what sort of "possibility" is this? Is it the possibility that the agent becomes "brain-washed" and a new self-understanding thus forced on him—or is it the possibility that he comes to see the Truth (which was always there to be seen)?

One must be cautious with the answer to these questions. One cannot separate the *truth* about the connection between the action and the reasons from the *understanding* of this connection. Therefore there is no unique way of deciding between the truth-claim of an agent's self-understanding and that of an observer's understanding from without.

There was a time when I believed that in the case of conflicting claims the agent's self-understanding must "in the last instance" be accorded priority.⁴ The qualification "in the last instance" indicated that an agent's self-understanding may be defective and a "conversion" is needed to make him see the truth about himself. To think otherwise would, it seemed

⁴ Cf. *The Varieties of Goodness*, Routledge & Kegan Paul, London, 1963, p. 190. See also above, p. 45.

to me then, be to assume an unjustifiable authority on the part of one man in matters relating to the "inner" life of another. The subject must be supreme judge in his own case, I used to say. But I am afraid that this position has to be abandoned. Self-understanding cannot be accorded ultimate authority. Understanding "from outside" *may* overrule its verdict.

XV

Behaviour sometimes is seemingly "causeless", "incomprehensible", "irrational"; we may even doubt whether it should be classified as "action" at all. But in a wider context it may assume a different appearance.

A person kicks another person in a crowd. He knows the person whom he kicked. He apologizes and says it was a "mistake". Perhaps it was. Next time they meet the first says something which hurts the feelings of the second. Unintentionally, he maintains. But we may begin to wonder. We learn that the second person once did something which made the first feel deeply wounded. A similar incident occurs for a third time. If the agent insists that it was all unintentional and by mistake, we may say: " 'Mistakes' simply are not all that often repeated. You kicked him because you wanted to pay him back, to revenge the wrong you thought he had inflicted on you". Even if the person were never to admit this, we may be convinced that we had understood his behaviour better than he does himself. Further consequences may confirm our belief. If we say that we have seen the Truth in the matter, this easily suggests a misleading picture of the case. But we may say that we have seen the truth *to the extent that there is a truth to be seen*.

XVI

It may be illuminating to draw attention to some distinguishing features between typical cases of action explanation, causal explanation, and deductive explanation.

In giving a deductive explanation we show that a proposition follows logically from a set of grounds or premisses in which its content is "implicit". In order to "extract" the conclusion from the premisses we ought to "analyse" or "explicate" their content.

In a causal explanation we *isolate* a fragment of nature from a wider context of "accompanying circumstances", and link a piece within this fragment or frame as "effect" with another piece as "cause". The ideal test of correctness of a causal explanation is *experimental*. We reproduce the cause within the given frame and watch whether "nature" will reproduce the effect. This requires that the frame too is reproducible and that what happens inside it is not affected by outside differences in the situations under which the reproduction takes place. A failure of the experiment is

sometimes interpreted as a refutation of the causal law, sometimes as a "disturbance" caused by outside factors. If we can control the presence and absence of these factors we can reformulate the frame-conditions under which the law holds and continue to use it for causal explanations. If, however, we simply reject the law we must look round for a new explicative principle.

The characteristic pattern of an action explanation in terms of motives and reasons is different. Here we are never interested in *isolating* the context but sometimes in *widening* it so as to make the action to be explained match maximally well with the rest of the agent's life history. Usually we can stay content with a narrow context. All that is needed for our understanding of the case may be that the agent deemed his action necessary for something he coveted or thought adequate for meeting some requirement. It is on the whole only when the objects of his intentions or wants seem queer or "suspect" or his beliefs distorted or the requirements which his action is supposed to meet abnormal that we are anxious to widen the context so as to get a more coherent picture of the case. Some cases remain inexplicable. Usually they are laid aside as being of no consequence. Sometimes, however, they continue to puzzle us—for example if they are enigmatic but consequential actions of historically important personalities. Historians will perhaps again and again dig into the background, revising our previous understanding of the case and our assessment of its significance. The criteria for when we have succeeded in explaining or understanding an action may also change and vary. What satisfies our craving for an answer to the question *Why?* is not uniquely determined by the brute facts of the case, but depends also on the requirements which we put on an acceptable explanation.

The Foundation of Norms and Normative Statements

I

By a *normative statement* I understand a statement to the effect that something ought to or may or must not be done. Such a statement is often, but not necessarily, elliptic or incomplete unless it says, by whom this thing ought to (may, must not) be done, and whether it ought to (may, must not) be done generally or on some particular occasion, unconditionally or provided certain conditions are fulfilled.

I use the term "statement" here in that which I propose to call its "strict" sense. A statement in the strict sense is true or false.

By the *foundation* of a given normative statement I shall understand the content of a truthful answer to the question, *why* the thing under consideration ought to or may or must not be done. The foundation, in this sense, of a normative statement can also be called its *truth-ground*.

To answer the question of the foundation or truth-ground of a normative statement is to tell, what *would make* the statement true. The foundation of the statement is thus some "possible fact". If this possible fact is actual, the statement is true.

Very often, when a normative statement is made, a foundation for it is also supplied. "You ought to (may, must not) do this, because . . .", we then say. If a foundation is not supplied, it is often asked for. "Why ought I to (may I, must I not) do this?", we then ask.

People are far from always clear over the foundations of the normative statements, which they hear. And not infrequently are people unclear also about the foundation of normative statements, which they make themselves. This is true above all with regard to statements concerning so-called moral matters.

The philosophic *problem of the foundation* of normative statements is, in the first place, the question of the general form and nature of the truth-grounds of such statements. It will be seen that there are different forms or kinds or types of foundation for normative statements. Moreover, one and the same normative statement may have more than one foundation of the same or of different types. (Plurality of foundations.) The question will also become urgent, whether a normative statement could be true and yet lack a foundation, i.e. whether it could be true to say that something ought to or may or must not be done and yet not possible to tell *why* (on what grounds, for which reasons) this is so.

II

One important type of answer to the question "Why ought (may, must not) this or that to be done?" is the following: *There is a norm* to the effect that this thing ought to (may, must not) be done. The existence of the norm is here the foundation or truth-ground of the normative statement.

For example: I am told that I must not park my car this side of the street. Why? The answer could be that there is a by-law or regulation prohibiting this.

The word "norm" in English has many meanings. Here I use the word to mean something, which can also be called a prescription or a regulation or, sometimes, a law. Commands and orders are a species of what I here call norms. But a norm can be a permission; and a permissive norm is not ordinarily called "command" or "order".

Examples of norms, in the sense in which I here use the word, are the laws of the state, or by-laws and regulations issued by a magistrate. But also orders, which parents give to their children, or officers to soldiers in the army, are norms in this sense.

Norms are given by some agent to a certain other agent or agents. I shall call the first agent norm-*authority* and the second norm-*subject(s)*. Calling the norm-authority "agent" should be compatible with the possibility that this authority is not a physical person, but e.g. a law-court or magistrate or legislative assembly.

The giving of norms requires the use of language. The norm-authority promulgates or makes known by means of signs to the norm-subjects, what he wants them to do or forbear. The signs which are used to announce the norm, I shall call *norm-formulation*. A norm-formulation can be a sentence in the imperative mood. Or it can be a sentence using the auxiliary verbs "ought to", "may" or "must not". We may call such sentences *deontic* sentences. Very often, however, the norm-formulation is an indicative sentence of the ("ordinary") declarative or descriptive type.

It is important not to confuse the norm itself with the norm-formulation, nor the existence of a norm with the uttering or other production of a norm-formulation. I shall not here discuss the logico-semantic nature of the relationship of the norm to its expression in words. I should only like to say, in passing, that I do not think it right to call the norm either the reference or the meaning or the sense of the norm-formulation.

Norms are neither true nor false. The notions of truth and falsehood do not apply to them. Norms fall outside the category of truth. In this respect they differ importantly from normative statements.

A *norm* is not true or false, but *that there exists a norm*, ordering or permitting such and such acts or forbearances, is true or false. The existence of a norm is a contingent, empirical fact. We shall not here inquire closer into the nature of this fact—of which we have already said that it cannot be identified with the existence of a norm-formulation.

III

A normative statement to the effect that a certain thing ought to (may, must not) be done may sometimes become founded on the existence of a norm, which is *expressly* to the effect that this thing ought to (may, must not) be done. This case is rather trivial. It is, however, not the only and perhaps also not the most common case.

Sometimes the truth-ground of a normative statement is *deduced* from (the existence of) one or several norms. For example: Let there be a norm to the effect that it is obligatory to do *p* and another to the effect that whoever does *p* ought also to do *q*. From these two norms we can deduce an obligation to do *q*, i.e. an (actual) truth-ground of the normative statement that *q* ought to be done. This is obviously a valid inference. Yet it is not an inference according to the "laws" of traditional logic. In order to account for the nature and validity of this inference an extension of the province of logic is required. I have elsewhere coined the name *deontic logic* for this new branch of logical theory. Deontic logic thus is important also to the question of the foundation of normative statements.

Norms are usually to the effect that acts of a certain category or kind ought to (may, must not) be done. A normative statement again is often about some individual act. The question may then arise, whether an individual act, about which a normative statement is made, falls or may become *subsumed* under some category of acts regulated by norm. We could call this a question of *interpretation* ("interpreting law").

The final answer to such questions of subsumption and of interpretation, it should be observed, is never set by "theoretical considerations" relating to the nature of individual acts. It is set by the reaction of the norm-authority to the disputed case. If the case is doubtful and the norm-authority has not yet reacted to it, the nearest we can come to an answer to our question may be a *conjecture* about this reaction and therewith also about the normative status (obligatory, permitted, forbidden) of the act. When a person or a corporation takes legal advice about some planned action of theirs, the advice often has the form of such a conjectural normative statement. A prospective agent may be anxious to know, what the ruling of a law-court would be, if a case were brought against him—say for unfair competition or for having injured somebody's privileges. In unclear cases it is for the law-court to decide, whether acts should be counted as obligatory, permitted or forbidden according to existing norms (laws).

What has just been said means that the existence of a norm is partly dependent upon the hypothetical reactions of norm-authorities to the conduct of norm-subjects. These reactions are not always easy to foresee in the individual case. It is probably these two facts in combination, which are responsible for the view—entertained by some legal philosophers—that the (legal) norms themselves were a kind of *prediction* (mainly about the reactions of law-courts). This is, I think, an utterly mistaken view of

the nature of norms—legal and other. But the existential statements, on which normative statements are founded, sometimes have to be enunciated in the form of predictions about the future reactions of norm-authorities.

IV

From the problem of the foundation of normative statements one must distinguish the problem of the foundation of the norms themselves. Many things can be meant by the "foundation" of a norm. I shall here mention and briefly discuss *two* of the senses, in which one may speak of the foundation of a norm, when "norm" is understood as we do it here.

A foundation in the first sense is again an answer to the question "Why?". Someone may wish to know, *why* there is (exists) a norm to such and such effect.

This question "Why?" is not a question of truth-grounds. It is a teleological question, a question of ends. It concerns immediately, not the norm itself, but the act or activity of the norm-authority, when giving or issuing this norm or rule. Why, i.e. for what reason or with a view to what end, has the norm-authority ordered or permitted this action to the norm-subjects? Not of every truthful answer to this question would it sound plausible to say that it provides a "foundation" of the norm. But if the answer satisfies certain requirements, it is natural to say that it *justifies* the act of giving the norm and thereby provides a *foundation* of the norm itself.

I shall call the giving of a norm *normative action*. The problem of foundation and of justification may now be put as follows: Which conditions must the answer to the question "Why?" satisfy in order to justify normative action?

That something is done *for the sake of an end* can mean two things. Firstly, it can mean that this act is considered necessary for the attainment of the end. That the act is necessary means that unless it is done, the end will not be attained. In other words: the act *must be* done, if the end is to be attained. Action, which must be done for the sake of an end, I shall call a *practical necessity*.

Secondly, that something is done for the sake of an end can mean that the act is thought to favour or promote the attainment of the end, without being necessary for this. The act is then said to be a *good* or a *useful* thing to do with a view to the end.

What has been said also applies to action of the type, which I have called "normative". A norm-authority may consider it useful or even necessary with a view to some end of his to prescribe certain acts or forbearances to certain norm-subjects.

That something is necessary or useful for the attainment of an end is an "objectively" true or false proposition. Since this is so, an agent may also be mistaken in thinking that something is necessary or useful with a view

to some end of his. The end may be, for example, that people should behave in a certain way and the mistake consists in thinking that one could make them behave in this way by commanding them. Gentler means of persuasion may be effective, where normative action is ineffective.

We can make these observations a ground for distinguishing between a "subjective" and an "objective" justification of normative action. The giving of a norm is subjectively justified, we shall say, when the norm-authority gives it, because he believes this to be useful or necessary with a view to some end of his. And it is objectively justified, when the belief which (subjectively) justifies it is true to fact.

When normative action has, in the sense just explained, an objective justification, I shall say that the norm, in which it results, has a *teleological foundation* in the norm-authority's ends.

We have so far viewed the norms in the perspective of the norm-authority's ends. One can, however, also approach the problem of the justification of normative action and the foundation of norms from the point of view of the norm-subjects' ends.

As the giving of norms may be necessary or useful for the ends of the norm-authority, in a similar manner the having or receiving of norms may be useful or even necessary for ends of the norm-subjects. There is, needless to say, no "pre-established harmony" between these two sets of ends, from the point of view of which the necessity and usefulness of norms may become judged. What the norm-authorities want the norm-subjects to do or forbear may be badly in conflict with what the norm-subjects want to do themselves. It seems to me to be a useful mode of approaching some of the central problems of political philosophy, to formulate them as questions concerning this harmony of ends in normative relationships. This approach to the problems is not altogether new, but I think it can, within a modern theory of norms, be given a clearer expression than before.

One could give a new sense to the terms "subjective" and "objective" as attributes of a justification of normative action. Normative action might be called *subjectively* justified, when it has a justification ("subjective" or "objective" in the old sense of these terms) from the point of view of the ends of the norm-authority. Whether normative action, which is subjectively justified, is also *objectively* plausible, might then be made dependent upon considerations relating to the ends of the norm-subjects, or upon some other considerations.

V

From the teleological foundation of norms in ends one must distinguish that which I shall call the *normative* foundation of norms. A foundation in this second sense is not the content of an answer to the question, *why* there is such and such a norm. It is provided by an answer to the question

whether the authority of the norm has a *competence* or *right* to issue the norm.

The normative, like the teleological, foundation of a norm is in the first place relevant to the *act* of giving the norm—and only *via* this act to the norm itself.

Like any other type of human action, normative action can be subject to norm or rule. There can be a norm to the effect that a certain agent may or may not issue regulations of such and such a kind. Norms regulating normative activity can conveniently be called *competence norms*. The subjects of such norms are themselves authorities of other norms.

That which I call the normative foundation of a norm can now be defined as follows: A norm has a normative foundation if, and only if, there exists a competence norm which permits the act of issuing this norm to the authority, who actually has issued it.

It is *one* of the uses of the terms “valid” and “invalid” in legal and other normative contexts to say of a norm, the giving of which is permitted by norm, that it is *valid* and of a norm, the giving of which is prohibited by norm, that it is *invalid*. In particular this use of “invalid” is common. That a norm is invalid very often means that the authority, who issued it, transgressed its competence as determined by the law.

There is also another use to be noted of “valid” and “invalid” as attributes of norms. Then the terms mean the same as “in force” and “not in force” respectively. To say that a norm is in force is another way of saying that it exists. To say that it is not in force is usually to say that it no longer exists—e.g., to say of a law that it has passed out of existence because of *desuetudo*.

As far as I can see, a great deal of confusion in legal philosophy has been due to failure to keep distinct these two, completely different, notions of the validity of norms.¹ Every norm, which has come into being within a given legal or other normative order, either is valid in the sense that it is in force or invalid in the sense that it has become repealed or otherwise passed out of existence. But it is not the case that every norm within a given legal or other normative order is either valid in the sense that it was issued by a competent authority or invalid in the sense that the authority, who issued it, transgressed its competence. For all I can see, it is even logically impossible that every norm within a given normative order should possess that which I have here called a normative foundation. There must, it seems, exist at least one supreme competence norm, which is void of this kind of foundation.

One can, however, also define “competence” or “right” to issue norms in a way, which is independent of the existence of norms concerning this competence. I shall not here discuss, how such a definition may be given.

¹ For a fuller treatment of the two notions of validity see *Norm and Action*, pp. 194ff.

Be it only said that an aspect of the question of the right to prescribe rules for the conduct of others is intimately connected with the question, which we discussed in Section IV, of the justification of normative action and therewith also with that which I called the *teleological foundation* of norms.

VI

I now move on to discussing another type of foundation, which normative statements may be given, i.e. another type of answer to the question, *why* something ought to or may or must not be done. It will be seen that this second type of foundation, although different from the first type of foundation of normative statements, is in a characteristic way related to the first type of foundation of *norms*, which we have been discussing.

Consider the following case. A builder orders beams from a carpenter for a house, which he is building. He says the beams ought to be so and so many inches thick. If he is asked, why the beams ought to have this thickness, his answer is perhaps that, unless they are thus thick, the roof will (or may) break down. This statement of the builder's could also be cast in the following form: In order to support the roof (safely) the beams ought to be thus and thus thick. Instead of "ought to be" we could here also say "must be" or "have to be".

I shall call the "unless"-statement, and its alternative form the "in-order-to"-statement, *anankastic* statements.

In an anankastic statement there is mention of two states of affairs or events, of which the one is said to be (or not to be) a necessary condition of the other. For example: the roof being safely supported and the beams being thus and thus thick. Now it may happen that the conditioned state or event is an end of human action, something we want to achieve. Then the conditioning state or event becomes something which, for the sake of attaining this end, we ought to produce or bring about. For example: we want the beams to support the roof safely, and therefore we ought to take care that they are of the needed dimensions.

Of the sentence "If you want the roof to be safely supported, you ought to make the beams thus and thus thick" I shall say that it expresses a *technical rule*. Technical rules must be distinguished both from anankastic statements, which are statements of (logical or natural) necessity, and from norms or rules of action, which are somebody's order (or permission) to somebody else.

Now consider the normative statement that the beams ought to be made—by whoever makes them—thus and thus thick. Let the question "Why?" be raised for this statement. And the answer given "Or else the roof will (may) collapse." That this could be a satisfactory answer will, I think, be readily admitted. Yet it is only provided two things are taken for granted

that the answer is satisfactory. The one thing is the truth of the anankastic statement that unless the beams are of certain dimensions, the roof will (may) collapse. The other thing is the fact that preventing the roof from collapsing is an aim or end of the person, who orders the beams to be made. These two things in combination, the end and the necessary connection, I shall say, constitute a (teleological) *foundation* of the normative statement. Alternatively, we could say that (pursuing) the end and (believing in or being aware of) the necessary connection *justify* the (making of the) normative statement.

Given the end and the normative statement, we can tell what is supposed to be the necessary connection "linking" the two. The technical rule "if you want, etc." we can therefore regard as a way of justifying a normative statement aiming at the given end.

There is an easily recognized logical relationship between the foundation of a normative statement in ends and necessities on the one hand and that which we have previously called the teleological foundation of norms on the other hand. The relation, briefly speaking, is this: Whenever ends and necessary connections justify a normative statement to the effect that a certain thing ought to (may, must not) be done, then a command (permission, prohibition) to somebody to do this thing, given by somebody in pursuit of these ends, has a teleological foundation in the ends of the norm-authority. The converse of this, apparently, does not hold. Considerations of utility (goodness) may provide a teleological foundation of norms even in the absence of necessary connections linking means to ends (cf. above, Section IV).

We have thus become acquainted with two principal answers to the question, *why* a certain thing ought to or may or must not be done. The one is to say that there exists a *norm* ordering or permitting or prohibiting the doing of this thing. The other is to say that *ends and necessary connections* make (or do not make) the doing or forbearing of this thing a practical necessity.

On the basis of the above we could also say that the idea of "ought" has two main sources. The one source is in the will of a commanding agent or norm-authority. The other is a double source in ends of human action and necessary connections between things.

In themselves, the two sources are of a rather different nature. But they are related to one another through the notion of a foundation of a norm (as a manifestation of the will of a norm-giver). Norms are frequently, perhaps one could say: normally, given for the sake of some ends. For this reason it may happen that the "ought", which flows from a commanding will, becomes supported by the "ought" of a technical rule and will rest on this latter "ought" as on its foundation.

VII

The question may be raised: Do all normative statements have a foundation of either of the two types, which we have so far been discussing? In particular, one may wish to know, whether normative statements relating to so-called *moral* matters, may be given a foundation of these types.

The answer to the second question will partly depend upon what one understands by a *moral* normative statement (or by "moral matters"). Trivially, one can define a moral normative statement as a statement to the effect that doing or forbearing a certain thing is morally obligatory or morally permissible. This leaves open the non-trivial question, what makes action, *in a moral sense*, obligatory or permissible. I shall return to this question in Section X.

An *example* of a normative statement which, I think, nobody would hesitate to call "moral", is the statement that promises ought to be kept. It is, moreover, an example of that which I shall call a statement of *moral principle*.

Only some moral normative statements, i.e. statements to the effect that something or other is morally obligatory or permissible, state moral principles. Consider, for example, the statement that I ought to help my neighbour to paint his house. Let the question "Why?", i.e. the question of its foundation, be raised. And the answer given "Because I have promised."

This answer to our question would normally be regarded as completely satisfying. Only under special circumstances would we raise the further question "Why ought you to do that which you have promised to do?". To raise it would be to raise a question of moral philosophy.

The answer "Because I have promised" is, of course, not the *only* possible answer to the question, why I ought to help my neighbour to paint his house. There could, in some community, exist a law or regulation to the effect that one ought to help one's neighbours with this kind of job. Or the reason for the "ought" here could be some "technical" considerations, relating to ends and wants and to what will happen, *if* . . . For example, that unless I help my neighbour now, he will not help me on some other occasion, or my reputation will suffer, or . . . These answers need not be exclusive of one another nor of the answer "Because I have promised."

That fact that I have promised to do a certain thing is thus a possible foundation of the normative statement that I ought to do this thing. As a foundation of the statement it is, moreover, a truth-ground of it. (See above, Section I.)

It is, however, plain that what makes the fact that I have promised to do *p* to be a foundation of the normative statement that I ought to do *p*, is the moral principle that promises ought to be kept. It is because promises ought to be kept that I ought to do *p*, if I have promised to do *p*. The normative

statement is founded on the *principle* of promise-keeping in combination with the *fact* that I have promised. We could also say that the moral statement has become *subsumed* under the moral principle.

Subsumption under a moral principle is thus a type of foundation of normative statements. Many moral normative statements have this type of foundation. It is, however, impossible that they should *all* become thus founded. At least some statements of moral principle must, for logical reasons, be void of this type of foundation.

Subsumption of moral normative statements under moral principles is an analogue to the foundation of normative statements on (the existence of) norms (see above, Sections II and III). Moral principles are analogous to norms. They are indeed quite commonly called "moral norms". There is no objection to this terminology. But it is important to realize that, *prima facie*, moral principles are very different from norms in the sense of prescriptions for action, given by some norm-authority to some norm-subject(s).

As a foundation of moral normative statements subsumption under moral principles may be said to possess only a "relative" value. For the question will instantly arise, how statements of moral principle then may become founded. Until we have answered this question, we do not even fully understand the nature of the foundation, which the moral principles themselves provide for the various moral normative statements that are subsumed under them.

VIII

There are many answers to the question, which we raised at the end of the preceding section. The answers correspond to well-known positions in moral philosophy.

One answer is to take the view that moral principles *are* that which we have here called "norms"—or something very much like this. One could term this the *legalistic* or law-conception of morality. The historically most important version of this conception is the view that moral principles are the commands (commandments, laws) of God to men.

This view of morality cannot be dismissed as being an "anachronism". Maybe we must reject it. But if no other attempted foundation of morality satisfies us, we are in a serious predicament. As Dostoevsky put it: "If God does not exist, then everything is permitted." If there is no satisfactory answer to the question, why one should do, as the moral principle says one should do then why should one be moral? This is no purely "theoretical" problem, but also a "practical" one.

To say that no moral normative statement is founded on the existence

of a norm is tantamount to saying that the fact that somebody has commanded or permitted a certain conduct to somebody is never a reason, why this conduct is *morally* obligatory or permissible.²

If we accept this thesis, we must look for a foundation of moral normative statements in some other source beside the normative activity of a God or of men. But the thesis is in no way incompatible with the fact that norms and human normative activity are, to a very considerable extent, concerned with moral matters, i.e. with regulating conduct which is considered morally obligatory or permissible. This is true both of the laws of the state and of the rules of conduct, which parents and teachers try to make children obey.

Consider, for example, the principle of promise-keeping. Agreements and contracts are kinds of promises. The *legal* obligations, which people have under contract, are therefore obligations to keep a kind of promise. But we are inclined to think that the *moral* obligation to keep promises, even when *safeguarded* by law, is not *founded* on law. This means: We think that the proper answer to the question, why it is *morally* obligatory to keep promises, is *not* that there is some norm or norms to this effect.

One could, in a broad sense of the word "legal", call those normative statements, which have a foundation in norms, *legal* normative statements. Then one could, if one rejects the law-conception of morality, say that it is a characteristic logical difference between legal and moral normative statements that they hold opposite relationships to norms. Legal normative statements are *secondary* to norms in the sense that the existence or non-existence of norms determine their truth-value. Something is a legal obligation or right, *because* the norm (law) so requires. Moral normative statements are not secondary to norms, and nothing is morally obligatory or permissible, *because* it is prescribed. But many *norms* are secondary to moral principles in the sense that we issue and enforce the norms, because we want people to behave morally.

Rejecting a law-conception of moral principles thus does not mean that the logical ties between morality and norm (law) are completely severed. This must be taken into account, when discussing the possibility of a moral without a religious foundation.

² Note the difference in meaning between "permitted" and "permissible" in ordinary usage. "Permitted" suggests that somebody has given permission to somebody. "Permissible" does not suggest this. It does not sound natural to call something "morally permitted", but it is perfectly in order to call something "morally permissible". Perhaps this observation can be instanced as showing a certain unnaturalness of a law-conception of morality.

IX

Another possible answer to the question of the foundation of a statement of moral principle would be to say that such statements have no foundation—at least not in that sense of “foundation”, which we are here discussing. On this view, it is not possible to tell, *why* that thing ought to or may or must not be done, which according to the moral principle is obligatory or permissible. This impossibility does, of course, not exclude that one could answer the question, how people *have come to think* that certain things are morally obligatory or permissible. But the answer to this question would not provide a foundation of moral principles—but a (causal) explanation of certain moral attitudes and opinions.

If statements of moral principle lack a foundation, does it then follow that they have no truth-value? In other words: If it is “in principle” impossible to tell *why* something or other ought to (may, must not) be done, is then the statement *that* it ought to (may, must not) be done neither true nor false? Statements in the “strict” sense of the term are true or false, we have said. Thus, if we accept an affirmative answer to the above questions, we should have to say that statements of moral principle are no statements in the “strict” sense. They do not express “propositions”, are not accepted or rejected on the basis of acts of “cognition”.

The view that statements of moral principle, because void of a foundation, are neither true nor false, would be a form of the position, which has become known as *non-cognitivism* in ethics. It is much in fashion among contemporary writers.

It should be observed, however, that the mere fact that the question, *why* it is thus and thus, cannot be answered for a given statement to the effect *that* it is thus and thus, is not sufficient to show that this statement has no truth-value. There are a great many statements which are true or false, but which have no truth-ground “outside their own content” so to speak.

Perception-statements, for example, may be regarded as being of this category. Consider the statement that I hear my own voice, when reading this paper to my audience. There may exist a physiological explanation, why it is that I hear my own voice. But the truth of the statement that I hear my voice, is not logically grounded on any further fact beside the fact that I hear it. There is no answer to the question, *why*, i.e. for which reason, it is true. The statement is true or false, depending, as we say, upon whether I hear or do not hear my voice.

In view of this, could it not be the case that moral principles too lack a foundation but yet possess a truth-value? To give an affirmative answer would be to think that the statement that something or other ought to (may, must not) be done could be true for no other reason but that this thing ought to (may, must not) be done.

The opinion that statements of moral principle are true or false, but not grounded on anything outside their own content, is known as the *deontologist* position in ethics. Adherents of it sometimes wish to make an analogy between moral normative statements and perception statements, and sometimes between moral normative statements and mathematical statements. The truth of statements of moral principle, they say, can be "intuited" or is a matter of "self-evidence".³

When viewed in the perspective of the problem of foundation of normative statements, the deontologist and the non-cognitivist positions thus come to be related. They share the opinion that for an important category of normative statements one cannot answer the question, why it is as the statements say that it is. In other regards the two positions differ sharply. Both are, in my opinion, unsatisfactory as attempts to give an account of the nature of moral principles.⁴

X

The chief alternative, in the history of ethics, to a law-conception of moral principles is a *teleological* view of them. The so-called utilitarianism is a variant of teleological ethics.

One can characterize the difference between the legalistic and the teleological conception of morality as follows: The first regards the foundation of normative statements on the existence of norms as the pattern, according to which statements of moral principle have to be founded. The second again regards the foundation of normative statements in necessities and ends of action as being the proper pattern. On the first view, the moral "ought" is the "ought" of a command. On the second, it is the "ought" of a technical rule. Just as the command behind the "ought" must be a very special sort of command (e.g. by a God) in order to constitute a *moral* "ought", in a similar manner, the technical rule behind the "ought" must be of a very special sort in order to justify a statement of *moral* obligation.

³ The position sometimes called an *ethics of conscience* may be regarded as combining elements of a law-conception and a deontologist conception of morality. My conscience "bids" me to do and forbear certain things. As such an inner voice it is like a norm-authority. But by consulting my conscience I also come to "see" or understand that such and such is my duty. As a supposed source of knowledge of duty, conscience may be said to resemble perception as a source of knowledge about the sensible world.

⁴ For assessing the correctness of non-cognitivism as a position in ethics, it is important to distinguish clearly between norms and values, and between normative statements and value-judgments. Unfortunately, this is seldom done with sufficient clarity by adherents of the position. It seems to me that as a theory of the nature of valuations and value-judgments, non-cognitivism contains much that is both true and important. To call statements of moral principle, e.g., that promises ought to be kept, "value-judgments" would, however, be to blur important logical distinctions. To point this out is not to deny that moral principles may have logical connections with valuations.

We shall here consider the problems confronting a teleological view of moral principles in the light of an example. The example will be the principle that promises ought to be kept.⁵ This principle has presented those philosophers in the past, who have been inclined towards a teleological view, with notorious difficulties. It is therefore a good test-case for the possibility of giving to statements of moral principle a foundation in ends and necessary connections.

That the institution or practice of promising is of value to those who participate in it, is not only an obvious fact. It is, moreover, true for reasons of logic. This is so, because the objects of particular promises, i.e. the things which the givers of promises promise to do, are things in which the receivers of promises take an interest. If I am indifferent to a thing, which is held forth to me as the object of a promise, I do not receive the promise. And if I loose interest in it, after having received the promise, I—as we say—“release” the promisor from the obligation to fulfil his promise. Only promises, the objects of which represent an interest (a “good” or “value”) to the promisee, oblige. Hence every man who participates in the practice of giving and taking promises, necessarily has an immediate interest that those, who have promised *him* something, should keep their word. This is a common interest of all participants. It is not, however, a contingent interest of theirs, but a necessary interest or an interest inherent in the concept of promising.⁶

The *anankastic statement* underlying the principle of promise-keeping can be formulated as follows: People ought to keep their promises, lest others should forego the benefits of participating in the practice of promising and suffer the evil of being deceived. (This is not to say anything “over and above” the fact that, unless people keep their promises, others will lose a benefit and suffer some evil.)

From the anankastic statement must be distinguished the *technical rule* underlying the moral principle. The anankastic statement speaks of a necessary connection between a certain interest (“good”, “value”) and a certain practice or mode of conduct. The technical rule extracts an obligation to behave in a certain way from the pursuit of this *interest* as an end. It says that, if you want not to do evil to your neighbour, you ought to keep your word to him. The interest, of which the technical rule speaks as an end of my action, is, be it observed, not *my* interest (self-interest), but my neighbour’s interest. It is for the sake of *my neighbour’s good* that moral conduct is required of *me*.

By love of one’s neighbour we can understand pursuit of one’s neighbour’s good “for its own sake” or, as it can also be called, as an “ultimate end” of action. It should be clear, in which sense love of our neighbour

⁵ Cf. the essay “On Promises” in this volume.

⁶ Cf. below, pp. 97ff.

can be said to demand or require moral conduct of us. Whether love can itself be required, imposed in the form of an obligation, we need not here discuss. What is certain, however, is that moral conduct, observance of moral principles, may become demanded of us otherwise than by our love of our neighbour. The way in which this may happen, is, moreover, essential to the logical picture we are trying to draw.

The common interest of all participants in the practice of promising that others should keep their word to them engenders the further common interest to *make* the givers of promises keep their word. This we try to achieve in a variety of ways: by teaching children the importance of being trustworthy, by urging people to keep their word, sometimes by calling for the arm of the law to enforce the obligation, and by the punitive measure of not trusting the faithless again, i.e. by depriving them of the benefits of further participation in the practice. The activity, in which men engage in order to make people keep their promises, is to a large extent of the type, which we have called normative activity. It includes both the commanding of children and the making of laws for the citizens of a state.

It is thanks to the fact that the interest at the source of the moral obligation engenders various kinds of normative activity that men *also* become self-interestedly concerned about obedience to moral rules. The self-interest, which men have in behaving morally, is the interest, which they themselves have to escape from the various evils which they threaten to visit on others, who behave immorally. In the case of promising, one of these evils, to quote Hume, is "the penalty of never being trusted again".

The interest which men have to escape the evils consequent upon immoral action is essential to the "mechanism of interests", whereby men become morally obligated. But pursuit of this interest is *not* the end of action, on which a statement of moral principle is founded. Not to be trusted again would not be a penalty for breach of word, unless continued participation in the practice of promising were a benefit. And participation would not be a benefit, unless the objects of promises were goods to the receivers. *This* is the interest, on the pursuit of which as an end the obligation to keep promises is founded.

My suggestion now is that a statement to the effect that something is *morally* obligatory or permissible is true if, and only if, it may be given a teleological foundation of the general form, which I have here illustrated using as an example the principle of promise-keeping. The interest at the source of a moral obligation is an interest, which each individual participant in a certain practice necessarily has that the other participants behave towards him in a certain way, defined by the practice. This is self-interest. Its pursuit does not oblige the interested party to the behaviour in question. But it may come to oblige his neighbours, his co-operants in the practice. The interest, from which the obligation on him arises, is not *his* but *theirs*. The obligation, which the moral principle imposes on him, is to

observe the same mode of conduct which *he* self-interestedly demands of *others*.

This view of the nature of moral principles is not, in its spirit, new. It is a view which makes an idea of *justice* the core of morality.⁷ It has an obvious affinity to the so-called Golden Rule. It is related too to Kant's maxim of universalization of action as a touch-stone of morality. (This is so notwithstanding the fact that Kant himself would have been strongly opposed to attempts to derive moral principles from considerations of interest.) But neither the Golden Rule in any of its usual formulations, nor Kant's maxim can be regarded as satisfactory expressions of the conditions, which a normative statement has to satisfy in order to qualify as stating a moral principle.

XI

My paper thus ends with a suggestion. This suggestion is a proposed definition of the concept of a moral principle in terms of a teleological foundation of morality. The suggestion could also be called a contribution to the moulding of the notion of a moral principle. Philosophical ethics, as I see it, is essentially such a moulding or shaping of the ethically relevant notions.⁸

Is there a test of truth for the results of this sort of inquiry? In some important sense, I think, there is no such a test of truth. But there are various tests of what may be called acceptability. The acceptability of the suggestion concerning the nature of moral principles which I have ventured to make, can be put to test in at least three ways. The first test is that normative statements, which we ordinarily regard as stating moral principles, can be given a foundation, which satisfies the suggested formal requirements. The second is that no normative statement, which has this sort of foundation, would commonly and naturally be regarded as *not* stating a moral principle. The third is connected with the fact that the suggestion provides a method for deciding, whether a proposed rule of action satisfies the requirement of being a moral principle. These are questions, on which moralists have notoriously disagreed and which may become thus decided. But on the further question, whether our suggested criteria are an acceptable basis for such decisions, moralists may nevertheless continue to disagree. The existence of such disagreement need not be deplored. It seems to me to constitute the life-nerve of philosophical ethics.

⁷ A fuller statement of this view is given in *The Varieties of Goodness*, Ch. X.

⁸ Cf. *The Varieties of Goodness*, Ch. I, Sect. 3.

On Promises

I

What is a promise? I.e., what sort of things are promises?, under which category do promises fall?

A promise is not an *utterance* or a *sentence* or anything which could be reasonably called a "linguistic" category. Nor is a promise an *act(ion)*, nor a *relation* between a promise-giver and a promise-receiver. It would also be false to say that a promise is a kind of *obligation*. But "utterance", "sentence", "act(ion)", "relation", and "obligation" are all of them needed for a satisfactory account of promises.

Promises belong to the same category as *agreements* and *contracts*. There seems to be no name for this category in ordinary language—and we need not here coin a name for it.

A promise is not an action, but promising or the giving of a promise obviously is a human action. Of this sort of human action the following features are characteristic:

(a) The action cannot be performed *solo*. It requires two parties, a promise-giver (promisor) and a promise-receiver (promisee). They are human beings.

I shall here disregard the case, when a promise is given collectively or jointly by several agents and also the case, when a promise has more than one receiver. It seems that these cases are essentially similar to the simpler case, which we are here discussing.

If acts of promising cannot be performed *solo* in the sense explained, then one cannot "strictly speaking" give a promise to oneself. Yet we sometimes do things which we describe as giving promises to ourselves. It seems to me, however, that promises to oneself are promises in an *analogical* sense only. My reasons for this opinion will not be given here. They have to do with the notion of having an obligation to somebody.

Vows and *oaths*, be it observed in passing, are related to promises. They differ from promises, it seems, among other things in the nature of their receivers, who need not be human beings.

(b) The "performing" of the act of promising, its physical aspect so to speak, normally consists in that the promisor addresses the promisee—either orally or in writing—using a characteristic *form of words*.¹ We

¹ I adopt Hume's phrase "form of words". It seems to me better suited for the purpose of a logical clarification of promises than the term "promise-sentence". Prichard talks of "a

could call this verbal performance an utterance (in a broad sense of the word).

In English, the standard form of words used for promising is "I promise" followed by an expression for the thing promised, e.g. by the phrase "to come and see you tomorrow morning". The standard form admits of variations. "I give my word" is another form. I shall later mention a possible alternative form, which differs both from "I promise" and "I give my word" in that it is not of the form of an indicative sentence.

It is worth mention that the verbal performance of the promisor in giving a promise often consists merely in an affirmative answer to a question on the part of the promisee "Do (will) you promise to — — — ?" It seems to me that philosophers, who have written on promises, have tended to ignore the role of the promisee in the act of promising, or to view promising one-sidedly from the standpoint of the promisor. If someone addresses me with the words "I promise you to do *p*" and if I have not asked him to do *p* or otherwise manifested some interest in *p* being done, I should probably feel bewildered or embarrassed. "What does he mean?" "Is he perhaps joking?" are questions which arise. If I cannot understand, why he should be promising me to do this thing, I should probably refuse to enter into a relation of promisee to promisor with him. (The receiving as well as the giving of promises is a mode of voluntary action.) His verbal performance would then fail to constitute an act of promising.

The thing promised I shall call the *object* of the promise. The object of a promise is one or several future actions (and/or forbearances) on the part of the promise-giver. Promises necessarily refer to the future. One cannot promise to have done something. Nor that something *is* thus and thus, e.g. that twice two is four. But one can make *assurances* about such things. (Assurances sometimes have the appearance of "promises referring to the past".)

(c) In the act of promising the promise-giver acquires an obligation to the promise-receiver. This obligation is to do the promised thing (the object of the promise).

That *x* has an obligation to *y* to do (or forbear) certain things can be called a "normative relationship" between *x* and *y*. The successful performance of an act of promising thus establishes a normative relationship between two agents.

It should be observed that the obligations, to which acts of promising give rise, are of the peculiar kind which I here call *obligations to persons*. Not every obligation, under which a man may be, is an obligation to some

certain noise in connection with the phrase for some action". See his essay "The Obligation to Keep a Promise" in H.A. Prichard, *Moral Obligation*, Oxford at the Clarendon Press, 1949.

other man (or men). When *x* has an obligation to *y*, then *y* is said to **have a right** as against *x*, viz. the right to claim or demand that *x* fulfil his obligation (e.g. keep his promise). Such rights sometimes have a legal significance. This notion of a right, it seems to me, is logically rather different from that notion of a right which consists in a *permission* to the right-holder to do a certain thing in combination with a *prohibition* to others to prevent the right-holder from doing this thing, should he choose to avail himself of his right.

II

When, under which circumstances, do we say that a promise has been given, i.e. an act of promising successfully performed?

The urge to establish (decide, find out), whether a promise has been given or not, usually comes from the claim on the part of an alleged promise-receiver that an alleged promise-giver should keep his word. The urge to establish the existence of a promise, one could also say, usually comes from a claim that a certain agent is under an obligation to another agent. The existence of the promise is the justification of this claim.

The mere fact that certain words were uttered or written down on some occasion can never be *sufficient* for concluding that a promise has been given. The words, first of all, must have been *used* and not *mentioned* (e.g. by some philosopher in a discussion of the nature of promises). The words, moreover, must have been used by somebody to *address* somebody else.

It is sometimes said that the appropriate form of words must be used "in earnest" in order to constitute a promise. But this is not quite true. If an agent demands that another agent should do something on the ground that he promised, then this other agent cannot refuse to carry the burden of the obligation *merely* on the ground that he had been joking. Not the fact that he was joking makes a man's use of the words "I promise to - - -" not constitute an act of promising. But often, when it is clear that a man, who says "I promise to - - -", is not promising, we conclude that he is (perhaps) joking. If a person addresses me with the words "I promise to make you Emperor of China", he would (normally) not be promising me anything. Why should he nevertheless have said this? The answer may be that he was joking.

It does not seem possible to tell exactly, *how* words must be used in order to constitute a promise. Not even the use of words *with the intention of giving a promise* is sufficient to secure that a promise has been given. For the addressee of the words may fail to respond in the adequate manner. He is perhaps not even *capable* of entering into a relationship of promisee to promisor with another agent (cf. below).

Even though the use of words is not, by itself, sufficient to accomplish an act of promising, it seems that the use of words or signs is *necessary* to

this end. Promising "essentially" involves the use of language. The giving and receiving of promises is therefore possible only among beings, who can communicate with each other by means of symbols.

I shall here mention and briefly discuss three more things, which may be regarded as necessary conditions (beside the use of a "form of words") for saying truly that an act of promising has been (successfully) accomplished or that a relation of promisor to promisee exists between two agents. It is not maintained that the four necessary conditions jointly constitute a sufficient condition. Nor is it maintained that it is possible to state necessary and sufficient conditions for deciding, when the four necessary conditions are, in the given case, fulfilled.

The proposing of necessary conditions is a contribution to the *moulding* or *shaping* (some would say "defining") of the concept of a promise. Rejecting some of the proposed conditions or replacing them by others would lead to different concepts of a promise. It cannot be claimed that the concept as shaped by us here is the "true" one, or the one which is "really" meant when we speak of promises in ordinary parlance. The concept of a promise is obscure. This is why it challenges philosophic reflection. And this is why the philosopher is to some extent free to shape the concept so that its logical features and relations to other concepts become clear. Nothing more will here be said about the justification of our procedure.

(a) A necessary condition of the origination or coming into being of a promise is that the parties concerned, i.e. the promisor and the promisee, should *know what they are doing* when giving and taking promises. They must, as we could also put it, be familiar with the institution or practice of promising. That a child has learnt to talk and also to say "I promise — — —" does not *ipso facto* mean that it can give and/or take promises. It is a logical peculiarity of the act of promising that it is possible to master the physical performance, which is an aspect of it and which consists in the production of some words or signs, without being able to accomplish the act. Ability to accomplish the act presupposes a certain "conceptual maturity" on the part of both agents, who engage in promising. It may be impossible to give necessary and sufficient conditions for deciding whether this presupposition is fulfilled in the individual case. Its character of necessary condition for saying that an act of promising has been accomplished is, however, not affected by this.

(b) From the fact that acts of promising give rise to *obligations* we may deduce the following necessary condition of saying that such an act has been performed: The promisor must be able to perform an action *of the kind* (category), to which the promised action belongs. (The question, whether he can perform the individual action, which would fulfil his promise, is not relevant to the question, whether he has given a promise or

not.) If it is clear that the promisor cannot do the kind of action he says he promises to do (e.g., swim across the Atlantic), we dismiss his words as nonsense and do not regard them as constituting a promise. If again it is doubtful, whether he can do it or not, we sometimes say that he is under an obligation to *try* to do it. We then interpret what he did as the giving of a promise to try to do a certain thing. Explicit promises to try to do things are not uncommon. The objects of such promises are acts of trying to do.

Consider the case, when the giver of a promise knows that he cannot do the thing he "promises" to do, but the receiver of the promise believes that the promisor can do this. The promisor can then rightly be accused of having cheated the promisee. The cheat, however, does not consist in the promisor's breaking his word, but in his making the promisee believe that something has been promised. This seems to me to be a natural way of describing the situation. Yet we must not be pedantic about the use of language here. The cheated party would perhaps want to insist on saying that he had been given a promise, but that the promise was not kept. He would then be using "promise" in such a way that the promisor's use of a form of words, which is characteristic of promising, in combination with the promisee's belief that he has been promised something, entails that a promise has been given. His use of "promise" would correspond to a slightly different concept of promising from the one, which we are here moulding.

(c) Finally, from the fact that acts of promising give rise to obligations of the peculiar kind, which I have called *obligations to persons*, one may extract the following necessary condition of saying truly that an act of promising has been performed: The object of the promise must represent an *interest* to the promisee, must be *for him a good*, something he *wants* or *welcomes*. (I shall regard these phrases as alternative ways of saying the same thing.) It seems to me to be a logical feature of the concept of an obligation to a person that one can have such an obligation only on condition that its fulfilment is in the interest of the person, to whom one has this obligation. I shall not here discuss the criteria for judging that something is somebody's interest. The question is complicated. Parents, e.g., have obligations to their children which may involve the doing of things, which the children shun or hate in their ignorance of their "true good". But parents cannot have obligations *to their children* to do things which are either harmful or totally irrelevant to the welfare of the children.

In order to see the importance of interests to promising, it may be useful to consider the circumstances under which promises normally originate. The "initiative" is usually with the promise-receiver—and *not* with the promise-giver. For example: One man wants another man to do something. The second says he will do it, but the first is not quite sure. "Will you promise?", he asks. If the second man then gives his word, the first will,

as a rule, feel more confident that he will get his wanted thing. Very often promising is linked with an *exchange of services* (goods). Then the connection with interests is in the nature of the case. One man asks another to help him with a certain task. The second agrees, but only on condition that the first promises a service in return on another occasion.

It sometimes happens that an act, which is in itself of no interest to the promisee, acquires an interest to him as a consequence of being mentioned in the form of words "I promise to — — —". For example: Someone asks me for an appointment and promises to come to see me tomorrow morning. It is of no interest to me, let us assume, to grant him the appointment; nor am I anxious that it should be tomorrow morning and not at any other time. But having taken his word, I arrange my day accordingly. Perhaps I make an appointment to see another person later in the day. If the first person does not turn up at the appointed time, his failing to do so may cause considerable annoyance and maybe even damage. I can insist that he should come in the morning, as promised, and I can accuse him of breach of promise, if he does not turn up then.

If something, in which a man takes no interest at all, is held forth to him by another person as the object of a promise, the first man will either acquire an interest in the thing and enter into the relation of promisee to promisor with the second man, or he will refuse to enter into this relation. In the second case, what the prospective promisor said and did failed to constitute an act of promising.

If a person addressed me with the words "I promise to make you Emperor of China", he would (normally) not be promising me anything (cf. above). One reason, why he *cannot promise* to make me Emperor of China is that, as we both know, he *cannot make* me Emperor of China. But even if he could make me Emperor of China, his words would fail to constitute a promise with me, i.e. with the writer of this essay on promises. For I am not in the least attracted by the prospect and my attitude to it would not change as a consequence of its being held forth to me as the object of a promise. I should *refuse* to enter into the relation of promisee to promisor with my interlocutor.

Sometimes it happens that the promisee loses or gives up his interest in the object of the promise before the promise is fulfilled. Then the promisor is *released*, as we say, from the obligation to keep his word. The relation of promisor to promisee ceases, dissolves. The promise, so to speak, passes out of existence before its fulfilment.

* * *

The form of words "I promise to — — —" is sometimes used for addressing a person, knowing that the action mentioned in the form of words *is* a thing which the addressee shuns. This sort of "promise" we call a *threat*. A threat, I would suggest, never creates an obligation *to the person, who*

is being threatened, to make the threat effective. For this reason a threat, even when it is "dressed up" in the form of a promise, is not a (genuine) promise.

Normally, it does not make much difference, whether a person says to another "I shall punish you, if — — —" or "I promise to punish you, if — — —". But it makes a *great* difference, whether a person says to another "I shall do as you ask me" or "I promise to do as you ask me". In the first case, it is only a declaration of intention that carries weight—irrespective of which form of words has been used in announcing the threat. A threat of punishment *cannot* assume the obliging character of a promise (to the threatened person).

III

Some words should be said about certain features, which are *not* essential to promises or to deciding, whether a certain verbal performance constitutes an act of promising. The reason for mentioning them is that they may *appear* essential.

It is, first of all, not essential to the existence of a promise that it should be fulfilled or kept, i.e. that the promise-giver should actually do what he has promised to do. He may try but fail, or he may neglect the promised act altogether. *Broken* promises are a species of *unfulfilled* promises. Both are promises.

It is, moreover, not essential to the question of the existence of a promise that the promise-giver, at the time of promising, should *intend* to do what he promises to do. A promise is not a declaration of intention. Nor is it, of course, a *prediction* of what the promisor is going to do. Promising without intention to keep the promise can aptly be called giving a *false* promise. But it is not making a false statement (to the effect that a promise is being given). False promises, like broken and unfulfilled promises, are promises too.

The question may be raised, whether it is not essential to the (existence of the) *institution or practice of promising* that false and broken promises should be *exceptions*, i.e. that promises should on the whole be given with an intention of keeping them and also on the whole be kept. I shall not discuss this question here. Nor shall I discuss the problem, whether an attempt to promise, which uses some such form of words as "I promise you to do *p*, but I do not intend to do *p*", would be self-refuting, i.e. fail to constitute a promise, or otherwise "paradoxical".

IV

To the act of promising, we have said, the use of a characteristic form of words is essential. We shall now examine in some detail this form of words and its function.

The form "I promise to — — —" is that of an indicative sentence in the first person of the present tense. With it may be compared sentences of otherwise the same form but in different person and/or tense. For example: "He promises to — — —", "I promised to — — —", "He will promise to — — —".

The normal use of all members of this class of (forms of) indicative statements, with the exception of sentences in the first person of the present tense, is for *stating that* somebody is promising or has promised or will promise to do something. The normal use of the sentence in question is thus to make statements, which are either true or false as the case may be.

The sentences of the class under consideration—again with the exception of those in the first person of the present tense—are *not* normally used for *giving promises*. Sometimes, however, they are used for that purpose too—or for some closely related purpose. Thus, for example, "I promised to — — —" might be used to reinforce or renew a promise already given; and use of the words "I shall promise to — — —" might be understood to constitute a promise to promise.

The form "I promise to — — —" stands out from all the rest of the class of (forms of) indicative statements under discussion in that its normal use is for *giving promises*. Considering the peculiarity of its function, its outward similarity to the other members of the class may be regarded as "philosophically misleading". This means: its form may lead the thoughts of someone, who reflects on the nature of promises, in a wrong direction. It may therefore be helpful to consider the possibility of replacing the characteristic form of words, actually used in promising, by some other form.

The conventional form of words used for promising could be, for example, "On my honour, I shall do — — —". Is this an indicative sentence? I do not know what grammarians would say. I think there are some reasons for *not* calling it an "indicative sentence". It is, for one thing, not clear what its different tenses would be. Sentences of this type could not without the introduction of special linguistic conventions be used for stating that a promise has been or is or will be given. If "On my honour, I shall do — — —" were the standard form of words used for promising (and "I promise to — — —" were not), then philosophers would perhaps never have been strongly tempted to think that the use of words for making promises was a use of words for making statements.

To sum up: It is a *contingent* fact about the use of language that the form of words used for promising is that of an indicative sentence. But it is *necessary* that the form of words used for stating that promises have been or are or will be given either is that of an indicative sentence or can be translated without alteration of meaning into indicative sentences.

V

Can the form of words "I promise to — — —" be used for *stating that* I promise to — — —? Or, to put the question in a more precise form: Can the very same token of this form of words be used both for giving a promise and for stating that I give a promise?

One must distinguish between the possibility that the words "I promise to — — —" are used for making a statement, and the possibility that they express a true proposition. That they can do the second need not be doubted, I think. The words "I promise to — — —" express a true proposition if, and only if, it is the case that I promise to — — —. From the fact, however, that the words actually happen to express a true proposition it does not follow that they have been used for making a (true) statement. Words, as they stand, may express a (true or false) proposition, even though nobody is there to "contemplate" or "entertain" it. But words, it may be argued, are not used for making a statement, unless their user *intends* this use of them.

Accepting this, the answer to the question, whether the very same words can be used both for promising and for stating that a promise is being given, would depend upon whether one can have and carry into effect "at once" the two intentions of giving a promise and stating that one is giving a promise. (The problem, whether one can do this, is not a psychological problem, but a conceptual or logical one.) Now granting that one *can* have the two intentions in one act, it is fairly obvious that *normally* one does not have both intentions. Only on rare occasions, therefore, would I, in uttering the words "I promise to — — —", both accomplish an act of promising and state that I am accomplishing this act.

It would, of course, be utter confusion to say that a sentence (either type or token) of the form "I promise to — — —" can be both a promise and a true statement and a true proposition. It, the sentence, can be neither of the three. But a token of a sentence of this form can be *used for giving a promise*. And it can *express a true proposition*. Whether it can also be *used for making a true statement* is, as we have seen, somewhat problematic.

VI

Consider the third person sentence "*x* has promised *y* to do *p*". Its normal use would be for making a (true or false) statement about something, which has occurred between two persons. By an *analysis* of (the meaning of) this sentence one could understand a formulation of its truth-conditions (when used for making a statement). This analysis would throw light upon the question, *what x* has done, when he promised *y* to do *p*, or—as we could also put it—what it *means* (to say) that *x* has promised *y* to do *p*.

This would be a possible analysis of the above third person sentence: "By addressing *y* with a certain form of words (such as 'I promise you to do *p*'), *x* has put himself under an obligation to *y* to do *p*". This would be an "analysis" also in the sense that it mentions two main "components" or "parts" of a promise, *viz.* the use of a certain *form of words* and the *obligation* which the user of the words henceforth (and on account of having used them) has to another person to do a certain thing.

"With these words I put myself under an obligation to you to do *p*" is another *form of words*, which could be used for giving a promise. (It probably sometimes *is* used for this purpose.) This form of words is *equivalent* to the form of words "I promise you to do *p*" in the sense that the two forms may be used for doing exactly the same thing, *viz.* giving a certain promise. And one could call the longer form of words an *analysed* version of the shorter form of words, on the ground that the first sentence contains explicit mention of things which, though essential to promising, are not overtly mentioned in the second.

The notions of an obligation in general and of an obligation to a person stand themselves in need of "analysis" or of some form of philosophic clarification. This, however, is not in conflict with the fact that the notions are needed for an analysis or clarification of the notion of a promise.

Can one not teach a child what promising is without teaching it what obligations (to others) are? One can teach a child to give and receive promises, without mention of the word "obligation" or some equivalent word, e.g. "duty". But teaching a child what promising is *is* making it familiar with a kind of obligation.

Philosophically, the concept of an obligation may be said to be highly controversial and obscure. And the same may be said of the concept of a promise. The fact, however, that the nature of obligation and promise puzzles philosophers, in no way impairs the commerce of men (including philosophers) in giving promises and in entering into various normative relationships with each other.

VII

In the rest of the present essay I shall discuss some problems relating to *the obligation to keep promises*. Considering the obvious philosophic importance and interest of this notion, it is surprising that it has received relatively little attention. Hume dug deep into the problems connected with it.² The

² In *A Treatise of Human Nature*, Bk. III, Pt. ii, Sect. 5, entitled "Of the Obligation of Promises".

most interesting recent treatment, of which I am aware, is by the late Professor Prichard.³

The problem in this region, which chiefly occupied Prichard, could be formulated as follows: How can an act of promising or the production of "a certain noise in connection with the phrase for some action", to use Prichard's words, create (give rise to, bring into existence) an obligation? Prichard's way out was to suggest that the obligation must exist antecedently to the individual act of promising. In order to account of this antecedently existing obligation, he invented the fiction of an antecedent *promise* (agreement) never to cause a noise of a certain kind in connection with the phrase for some action without also going on to do the action.⁴ This "second-order" promise, unlike ordinary promises, does not require the use of language.⁵ Therefore, strictly speaking, it is not a promise at all.⁶ Prichard's essay ends with the question, what this thing is, which must be presupposed in promising in order to make intelligible the obligation to keep a promise.

In order to resolve Prichard's difficulty, it is helpful to make a distinction which Prichard failed to make. It is the distinction between the (general) obligation to keep promises and the (particular) obligation to do *p* on account of having promised. The second can also be called the obligation to keep a particular promise.

It is clear that acts of promising have not created (cannot create) the obligation to keep promises. But individual acts of promising can quite correctly be said to create obligations to keep particular promises, i.e. obligations to do certain things *qua* promised. There is, as far as I can see, no difficulty or problem about this "creation" of obligations. Before we had promised to do *p*, it was irrelevant to our fulfilling of the obligation to keep promises, whether we did or neglected *p*. Having promised to do *p*, this is no longer irrelevant. The act of promising changed the relevance of our future doing or neglecting of *p* to the general obligation to keep promises. In this change consists the "creation" of the new obligation.

Instead of saying that an act of promising "creates" an obligation to do the promised thing, we can also say that an agent in giving a promise "puts himself under" an obligation to do the promised thing. The obligation, under which he is thus putting himself, is not the *general* obligation to keep promises. This is not an obligation, under which a man can put himself or from which he can withdraw or become released. He is, in a sense, always under it. The obligation, under which he can put himself

³ The essay called "The Obligation to Keep a Promise", published posthumously in *Moral Obligation* (1949).

⁴ *Op. cit.*, p. 172.

⁵ *Op. cit.*, p. 179.

⁶ *Ibid.*

through an act of promising, is the *particular* obligation to do a certain thing *p* on account of having promised it. But since the particular obligation did not exist antecedently to the giving of the promise, it seems to me better to say that one "acquires" it than to say that one "puts oneself under" it.

The answer to Prichard's question is thus *not* that the obligation to keep a promise presupposes a *promise* to keep promises. The answer is that the obligation to keep a promise presupposes—the obligation to keep promises. The answer may raise the further question whence this second obligation arises, what is its "foundation". As a solution to the (genuine) difficulty, which was puzzling Prichard, the answer nevertheless seems to me correct and satisfying.

The (general) obligation to keep promises, be it observed, is not an obligation of the kind which I have here called an *obligation to a person*. But the obligation to keep a particular promise, e.g. to do *p* on account of having promised, is an obligation to a person. Thus the obligation which is created (acquired) in promising is an obligation to a person, although the obligation to keep promises is not.

VIII

What is the "content" of the obligation to keep promises? What demand does it make on human action? The answer which immediately comes to our mind is this: The obligation to keep promises is the obligation never to promise to do a certain thing without also going on to do this thing.

I think that this answer is correct. Yet accepting it is connected with a certain logical difficulty. This can be stated as follows:

If the range of actions, which are the objects of promises, is unrestricted, then there is nothing inherent in the institution of promising to prevent a forbidden action from being the object of a (genuine) promise. A forbidden action is one which must not be done, i.e. which it is obligatory to forbear. A promised action is one which it is obligatory to do—not *simpliciter*, but *qua* promised. Now if somebody has promised to do a forbidden action, then it looks as though one and the same action had become both obligatory and forbidden, i.e. obligatory to do and obligatory to forbear. This, as such, can hardly be called a contradiction of logic. But it is plausible to regard it as a logical feature of the concept of an obligation that one and the same action cannot be both obligatory to do and obligatory to forbear (for one and the same agent on one and the same occasion). Promising the forbidden therefore appears to lead to something, which might be called a *contradiction of deontic logic* (the logic of obligation-concepts).

At this point someone may wish to say: Since promising the forbidden obviously is possible, how *could* it lead to a contradiction? But saying thus would be to take a much too simpleminded view of what promising

is. Of course, the fact that p is something, which must not be done, does not in the least affect the logical possibility of addressing a person with the words "I promise you to do p ". The serious question is, whether use of those words would constitute a promise, i.e. create an obligation for me to do p , if p is something which I must not do.

How shall that which looks like a contradiction here be resolved? A suggestion would be that the range of possible objects of (genuine) promises must, in the name of logic, be restricted to acts which are permitted. But in fact no such restriction is needed. This may be shown by elementary considerations of a logical nature.

What it means that the action p is obligatory, not *simpliciter* but *qua* promised, we can understand in the following way: It is obligatory *simpliciter* either not to give a promise to do p or to do p . Or, which means the same: It is forbidden *simpliciter* to promise to do p and leave p undone. From the prohibition *simpliciter* to promise to do p and leave p undone one cannot extract a prohibition *simpliciter* to leave p undone, i.e. an obligation *simpliciter* to do p . Therefore, that one and the same action is forbidden *simpliciter* and is obligatory *qua* promised, is, in fact, no contradiction at all.

The following logical point about promising the forbidden is worth noticing:

Consider an action which is forbidden (*simpliciter*). With regard to it we have two obligations. The first is not to do it. The second is either not to promise to do it *or* to do it. Under the first obligation we are not allowed to promise to do the action *and* do it (since we are not allowed to do it at all). Under the second obligation we are not allowed to promise to do the action *and not* do it (since this would be breaking our word). Thus under all circumstances, i.e. independently of whether we do or do not do the action, we are not allowed to promise to do this action. *The act of promising to do the forbidden is, for reasons of (deontic) logic, itself a forbidden action.*⁷

Although promising the forbidden does not lead to a *contradiction*, it can quite correctly be said to lead to a *conflict of obligation*. A man ought to do x , because he has promised it. But he also ought to forbear x , because x happens to be a forbidden action. What shall he do in such a situation? This is an ethical problem, often of considerable interest. It cannot be solved by means of considerations of a formal logical nature. It can be solved only on the basis of considerations of an *axiological* nature.

⁷ This principle is a consequence of what might be called the Jephthah Theorem of deontic logic—alluding to the story of Jephthah in the Book of Judges. For further comments on the theorem and on the notions of conflict of obligation and predicament, cf. my *Essay in Deontic Logic and the General Theory of Action*, North-Holland Publishing Co., Amsterdam, 1968, pp. 78ff.

By these I mean considerations relating to the question, whether it is *worse* (a greater bad or evil) to break the promise to do *p* than to do the forbidden thing *p*. It may be doubted, whether a general answer to the question can be given.

IX

Is it true that we ought to keep our promises? Ought we really to keep our promises? Whence, if from any source at all, does the obligation to keep promises get its "binding force", its "validity"? I shall say that these and similar questions refer, in a confused and unprecise way, to the problem of the *foundation* of the obligation to keep promises.⁸

This problem, as far as I can see, presents two rather different aspects. I shall here allude to them with the two questions "*Ought* promises to be kept?" and "*Must* I do *p*, if I have promised to do it?" respectively. The first question refers to the (general) obligation to keep promises, the second (directly or primarily) to the obligation to do a particular thing on account of having promised to do it.

Ordinary usage does not maintain a *sharp* distinction between the sense of "ought to" and "must". A distinction can, however, be made and is important to observe.

I shall say that something *must be done*, when doing this is (causally) necessary for the attainment of some end. I shall call such necessity of action a *practical necessity*. The question, whether I must do *p*, if I have promised to do it, is thus a question whether promise-keeping here is a practical necessity (with a view to some end).

A useful mode of attacking the question, whether keeping his word is a practical necessity for the giver of a particular promise, is to consider what the consequences would be of a breach of promise. If the consequences are something, which the promisor shuns, would not want to befall him, then it is necessary for him to keep his word in order to escape this evil. (Whether keeping his word is also sufficient for the attainment of this end, is, of course, another question.)

It is clear that a breach of promise *may*, in the individual case, not affect the welfare of the giver of the promise at all. It is also clear, however, that it may affect his welfare considerably—and in various ways. The party, whom he has wronged, may try to revenge himself. Or he will perhaps resort to the aid of legal justice to force the promisor to keep his word or to punish him for not having kept it. Even if the giver of the promise has to fear neither of these evils, he may yet by not keeping his word risk "the penalty of never being trusted again"—to quote Hume's words.

⁸ On the obligation to keep promises, see also the essay on the foundation of norms in this volume, above, pp. 80ff.

The penalty of never being trusted again could also be called excommunication from further participation in the practice (institution) of promising. Excommunication is a *penalty*, i.e. an evil for the individual concerned, only provided participation is a *benefit*. To most men participation is a benefit, permanent forfeiture of which will make their lives most unhappy (cf. Section X). The facts that forfeiture is not always permanent and that it is not always consequent upon a single breach of promise, tend to make people forgetful of the severity of being refused the benefits of participation in the practice of promising.

Considerations, when they arise, as to whether a particular promise must be kept, are usually conducted in terms of the *risks* (probabilities) of various unwanted consequences of a breach of promise. An agent may be prospectively willing to take risks, which he retrospectively regrets having taken. He may come to think that he should have kept his word, considering what happened as a consequence of his breaking his promise. He had been judging falsely of a practical necessity.

To try to argue that *every* promisor is *always* under a practical necessity to keep his word does not seem realistic. To think that *practically all* agents are for *the most part* well advised to keep their promises is good common sense.

X

Consider the following situation. A person maintains that he ought to do a certain thing. We ask him: "Why is it that you ought to do this?". He answers: "I have promised to do it."

There is a way of understanding "ought to" here, which is such that the above answer (if truthful) to the above question would be *completely* satisfying. On this understanding of "ought to", if the questioner went on and asked "Why is it that you ought to do that which you have promised to do?", we could say that he has not grasped what promising is. Promises *ipso facto*, by their very nature, ought to be kept.

If, however, the questioner went on and asked "Ought you really to keep your promise?", he *could* also be asking something, which is fully compatible with understanding the nature ("meaning") of promising. He could be asking, for example, whether the other man *must*, whether it is necessary for him to keep his word. He would then be viewing the question of promise-keeping in a new perspective. He would be asking, whether that which is, of course, an obligation is *also*, perchance, a practical necessity.

When the question "Ought promises to be kept?" is so understood that "I promised" is a completely satisfactory answer to the question "Why ought you to do this?", we view the obligation to keep promises as *inherent* in the nature of promising. It is a feature of the concept. The obligation, we

could also say, exists as a necessary part of the institution or practice of promising.

Consider now a particular promise, of which it is true that the giver of the promise is not under a practical necessity of doing what he has promised, that he *need not* do this. We can then still say truly that he is under an obligation to do what he promised, that he *ought* to do this. The question, however, may be raised: Is saying this anything "over and above" stating the bare fact that the agent has promised, given his word to do a certain thing? If not, we may feel that the obligation to keep a promise, when lacking a foundation in the practical necessity of action, is only an "empty word" and has no "binding force".

The feeling of uneasiness, which may be felt here, is relieved, I think, by further considerations pertaining to the *interests*, which are associated with the obligation to keep promises, generally and in the particular case. Of one such interest we have just been talking. This is the interest which a man may have to escape some bad thing, consequent upon a breach of promise, and which sometimes assumes the force of a practical necessity of keeping one's word.

The evil which a man may have to suffer as a consequence of not having kept his word, is inflicted upon him by the actions and reactions of other men. It is in the nature of things that a man should be interested in avoiding that which is, for him, an evil. But why should men be willing and even anxious to visit promise-breakers with evil?

It will here be necessary to remind of the fact that the obligation to keep a particular promise is an obligation of the peculiar kind, which I have called an *obligation to a person*. The fulfilling of an obligation of this kind, we have said, represents an *interest* (a good) to the person, to whom the obligated agent has this obligation. If the obligation is not fulfilled, the "wronged" party suffers disappointment. The bad of suffering such disappointment is sometimes minimal. But sometimes it amounts to grave damage.

It is because the objects of promises are of value to the promise-receivers that sharing in the practice of promising is a good thing and exclusion from the practice a bad thing to people. The good of participating in the practice of promising and the bad or evil of being punished for breaking one's word by becoming excommunicated from the practice are thus *correlative values*.

It is because sharing in the practice of promising is a good thing that people are interested in *making others keep* their word. This they try to accomplish in various ways: by teaching children the importance of being trustworthy, by urging people to keep their word, sometimes also by calling for the arm of the law to enforce the obligation, and by the punitive measure of not trusting the faithless again. These and similar measures to make

people keep their promises "institute" promise-keeping as an *obligation*, as something which we think *ought to be done*.

The interest which men have in making others keep their word I shall call the *foundation* of the obligation to keep promises. I have tried to show, how this interest and the interest, which makes dutiful action a practical necessity, are linked with one another thanks to the correlative value of the good of participation in and the bad of excommunication from the practice of promising..

On the Logic of Norms and Actions

1. DEONTIC LOGIC AS MODAL LOGIC—ANALOGIES AND DISANALOGIES

I

In what may be called the prehistory of modern deontic logic one can distinguish two main traditions. The one goes back (at least) to Leibniz, the other (at least) to Bentham.¹

Bentham entertained a grand idea of a Logic of Imperation or of the Will. It was going to be a new branch of logic, "untouched by Aristotle". Bentham did not develop it systematically. This was left to the Austrian Ernst Mally in his work *Grundgesetze des Sollens, Elemente der Logik des Willens* (1926). Mally seems not to have been aware of Bentham's pioneer work which remained practically unnoticed until the late mid-twentieth century. As an aftermath to Mally's work one may regard discussions in the 1930s and early 1940s on the logical nature of imperatives—including some constructive efforts at developing a Logic of Imperatives and of Optatives.

The discipline which now goes under the established name Deontic Logic did not evolve in the tradition of Bentham and Mally. It was born as an off-shoot of Modal Logic. None of its founding fathers, however, seems to have been aware that their leading idea had been anticipated, quite explicitly, by Leibniz who in the *Elementa juris naturalis* (1672) wrote: "Omnes ergo Modalium complicationes, transpositiones, oppositiones ab Aristotele et Interpretatibus demonstrare, ad haec nostra Iuris Modalia non inutiliter transferri possunt". With these words the birth of deontic logic can truly be said to have been announced.

II

By the *Iuris Modalia* Leibniz meant the deontic categories of the obligatory (*debitum*), the permitted (*licitum*), the prohibited (*illicitum*), and the facultative (*indifferentum*). And by saying that to the deontic modalities may be transferred all the "complications, transpositions, and oppositions" of Aristotelian modal logic, Leibniz was in the first place thinking

¹ Actually, the history of the formal logical study of norms and normative concepts can be traced back (at least) to the fourteenth century. See the interesting essay by Simo Knuuttila "Deontic Logic in the Fourteenth Century", in *New Studies in Deontic Logic*, ed. by Risto Hilpinen; D. Reidel Publishing Co., Dordrecht, Holland, 1981.

of the relations of interdefinability which obtain between the traditional (alethic) modalities.

I shall refer to these observations by Leibniz as *analogies of interdefinability* between alethic and deontic modalities. The analogies can be exhibited in the following table:

it is possible, M	it is permitted, P
it is impossible, $I = \sim M$	it is forbidden, $F = \sim P$
it is necessary, $N = I \sim$ $= \sim M \sim$	it is obligatory, $O = F \sim$ $= \sim P \sim$

We shall omit from special consideration here the category of the facultative, i.e. the neither-obligatory-nor-forbidden, answering to the alethic category of the contingent. Then interdefinability, as shown by the table, means that one can, taking *one* of the concepts as basic or primitive, through a process of "double negation" define or engender the other concepts of the triad. Which of the three one regards as *Grundbegriff* is indifferent.

As far as the interrelatedness of the basic deontic categories is concerned, Bentham seems to have been of the same opinion as Leibniz. But Bentham did not note the analogies with the modal concepts. The first author to study in detail both the analogies and the interdefinability relations seems to have been the Austrian Alois Höfler in a paper written in the 1880s but not published until 1917.²

III

The analogies of interdefinability do not, by themselves, suffice for the construction of an (interesting system of) deontic logic. To this end some logical laws or principles governing the deontic notions must be found or suggested.

The additional observation which gave a decisive impetus to my efforts in the area, concerned the distributive properties of the alethic and deontic modal operators. For the notion of possibility we have the distribution law $M(p \vee q) \leftrightarrow Mp \vee Mq$. It seemed to me then that an analogous principle holds for the notion of permission (permittedness). Accepting this *and* the analogies of interdefinability gives us the following *analogies of distributivity*:

$M(p \vee q) \leftrightarrow Mp \vee Mq$	$P(p \vee q) \leftrightarrow Pp \vee Pq$
$I(p \vee q) \leftrightarrow Ip \& Iq$	$F(p \vee q) \leftrightarrow Fp \& Fq$
$N(p \& q) \leftrightarrow Np \& Nq$	$O(p \& q) \leftrightarrow Op \& Oq$

² "Abhängigkeitsbeziehungen zwischen Abhängigkeitsbeziehungen", *Sitzungsberichte der kaiserlichen Akademie der Wissenschaften, Philosophisch-historische Klasse*, 181, 1917.

The first to pay attention to the distributive properties of the deontic concepts was, as far as I know, Mally. In Mally's *Deontik* the law (answering to) $O(p \& q) \leftrightarrow Op \& Oq$ holds. The analogy with modal logic, however, passes unnoted.

IV

By the *Minimal System* of deontic logic, I shall understand a calculus which can be characterized syntactically as follows:

Every tautology of propositional logic (PL) is a valid formula of the system when the propositional variables are replaced by deontic formulas. The sole (additional) axiom of the system is the formula $P(p \vee q) \leftrightarrow Pp \vee Pq$. The definitions $F =_{df} \sim P$ and $O =_{df} \sim P \sim$ are accepted. In addition to the usual inference rules of substitution and detachment we have a transformation principle to the effect that provably equivalent PL-formulas are intersubstitutable *salva veritate* in deontic formulas. This last principle may be regarded as a version of what is sometimes also called Leibniz's Law.

Bentham regarded it as a law of his Logic of the Will that if something is obligatory (Bentham says "commanded") then it is not also prohibited. In our symbolism above: $OP \rightarrow \sim Fp$. In the minimal system this is equivalent with the formula $\sim (Op \& O \sim p)$ which is equivalent with $Op \rightarrow Pp$ which again is equivalent with $Pp \vee P \sim p$. By virtue of the distribution axiom, finally, this last is equivalent with Pt where "*t*" stands for an arbitrary tautology of PL.

Bentham's Law is also valid in Mally's *Deontik*. Mally, moreover, recognized the role Leibniz's Law plays in the proofs of deontic theorems. His system, one could say, has all the ingredients of a "sound" deontic logic, but also contains additional ingredients which, unfortunately, from a formal point of view vitiate the whole undertaking. The "unsound" features of Mally's system have to do with his way of treating *conditional* obligations.

The system of deontic logic which I presented in my 1951 paper³ was the Minimal System embellished with one additional axiom. This was the formula $Pp \vee P \sim p$. I coined for it the name Principle of Permission. Accepting the relations of interdefinability this, as we have just seen, is but another form of what above I called Bentham's Law.

For the Minimal System embellished with Bentham's Law I shall, *faute de mieux*, propose the name the *Classical System* of deontic logic.

Any normal modal logic accepts as valid the formulas $Mp \vee M \sim p$ and $\sim (Np \& N \sim p)$ and their "contracted" form *Mt*. Thus Bentham's Law too has an analogue in modal logic.

³ "Deontic Logic", *Mind*, 60, 1951.

From the Classical System we reach what I shall, following Bengt Hansson,⁴ call the *Standard System* by making the following two modifications: The deontic operators are understood as operating on and resulting in propositions. Leibniz's Law is replaced by a stronger inferential principle which says that, if f is a valid formula of the deontic system, then Of is so too.

Both modifications strengthen the analogy with modal logic. The analogue of the inferential principle allowing the inference from f to Of is in modal logic known as the Rule of Necessitation.

The Standard System may be said to stretch the analogy between modal and deontic logic to its utmost limit. The only significant deviation lies in the fact that whereas traditional modal logic accepts as valid the formulas $p \rightarrow Mp$ and $Np \rightarrow p$, a "sound" deontic logic must reject their analogues $p \rightarrow Pp$ and $Op \rightarrow p$. It has to rest content with the weaker forms of those analogical formulas which are represented by the Principle of Permission and by Bentham's Law respectively.

When building the Classical System I took the view that the deontic operators operated on names of (categories or types of) action, and not on propositions. In the Classical System, therefore, "mixed" formulas, such as e.g. $p \rightarrow Oq$, or "higher order" formulas such as e.g. OPp , were not regarded as well-formed. To logicians these restrictions may seem impediments to the development of deontic logic. Their removal within the Standard System has gained more or less universal acceptance. There may nevertheless have existed some good and serious reasons against taking the step from the Classical to the Standard System—and even against the step from the Minimal to the Classical System.

V

Suggestive and, from a formal point of view, fertile as the analogies between the alethic and the deontic modalities may be, they are also open to doubts. The more I have reflected on the nature of norms and normative concepts, the stronger have these doubts grown with me. I shall next mention some points on which one may focus such doubts.

(a) *Disanalogies of interdefinability*

It seems much more plausible to regard the operators (concepts) O and F as being interdefinable than to regard P and O , or P and F , as being so. One can ask: is permission to do something simply the absence of prohibition to do this same thing? That permission should entail the absence of a "corresponding" prohibition seems clear. But does the reverse entailment

⁴ "An Analysis of Some Deontic Logics", *Noûs*, 3, 1969.

hold? Is not permission something "over and above" mere absence of prohibition?

This question is in fact a classic problem of legal philosophy and theory. Do permissions (rights) have an independent status in relation to prohibitions (obligations) or not? I think it is correct to say that opinions continue to be *very much* divided on this issue.

To say that prohibition to do is tantamount to obligation to forbear (omit) one and the same thing, and to say that obligation to do is tantamount to prohibition to forbear, seems uncontroversial. What is not clear and uncontroversial, however, is whether the relation between doing and forbearing (omission of action) is simply the relation between something and its *negation*. This is a grave problem for a "logic of action".

(b) Disanalogies of distributivity

Doubts concerning the analogies of interdefinability may, but need not, affect the analogies of distributivity. One might, for example, reject the formulas $\sim Fp \rightarrow Pp$ and $\sim O \sim p \rightarrow Pp$ and yet regard disjunctive permissions as being disjunctively distributable and conjunctive obligations as conjunctively distributable. In building a calculus or system of deontic logic one would then have to lay down independently in axioms the distributive properties of permissions and obligations respectively.

One can, however, for other reasons doubt the analogies of distributivity. Since its beginnings, deontic logic has been beset by some "anomalies" or "paradoxes". The best known and most discussed one is Ross's Paradox. Two others are the Paradox of Derived Obligation and the Paradox of the Good Samaritan. The last two ones may be regarded as variant formulations of the first. And all three have their roots in the formulas $O(p \& q) \rightarrow Op$ or, alternatively, $Pp \rightarrow P(p \vee q)$ of "traditional" deontic logic—whether in the "minimal", the "classical", or the "standard" version. Therefore, in a deontic logic which rejects the implication from left to right in the equivalence $O(p \& q) \leftrightarrow Op \& Oq$ while retaining the implication from right to left, the "paradoxes" would not appear.⁵

Analogous "paradoxes" are known from modal logic. The Paradox of Derived Obligation, for example, is an analogue in deontic logic to what is known as the Paradox of Strict Implication in modal logic. But the conflict between "intuition" and "formalism" of which the paradoxes are symptomatic seems to be much more serious in deontic than in modal logic. In this fact I would see an indication that the analogy between the two logics is not as perfect as many people have thought.

On a normal understanding of the word "or" in normative language, disjunctive permissions are *conjunctively*, and not *disjunctively*, distribut-

⁵ See A. Straninger, "Ein paradoxenfreies deontisches System", *Forschungen aus Staat und Recht*, Band 43: *Strukturierungen und Entscheidungen im Rechtsdenken*, 1978.

able. If someone is told that he may work *or* relax this would normally be understood to mean that he is permitted to work but also permitted to relax: it is up to him to *choose* between the two alternatives. Disjunctive permissions of this character I have called Free Choice Permissions.⁶ Opinions on their logical status differ considerably. Some logicians think that they only apparently conflict with the distribution law $P(p \vee q) \leftrightarrow Pp \vee Pq$. Another attitude is to reject, at this point, the analogy with modal logic and build a deontic logic which incorporates a distribution principle $P(p \vee q) \leftrightarrow Pp \& Pq$.⁷ Such a deontic logic, however, will have to differ in many other features as well from the traditional systems.

(c) *Disanalogies in the rules of inference*

In a "normal" modal logic the contradiction is pronounced impossible and the tautology necessary. A "normal" modal logic, moreover, accepts the Rule of Necessitation (above p. 103) and all tautologies of PL. From the Rule of Necessitation and the distribution principles of the modal operators one easily derives Leibniz's Law (above p. 102). Since $p \vee \sim p$ is a theorem, $N(p \vee \sim p)$ is a theorem, too. By virtue of Leibniz's Law and the interdefinability of the modal operators, Nt and $\sim M \sim t$ are theorems. However, taking $N(p \vee \sim p)$ or $\sim M(p \& \sim p)$ as *axioms*, one can, with the aid of Leibniz's Law, derive the Rule of Necessitation.

What I have called the "standard" system of deontic logic accepts the deontic analogue of the Rule of Necessitation. That which is, in deontic logic, provably true is also obligatory. This always seemed to me highly counterintuitive, sheer nonsense. Most logicians, however, seem willing to swallow the absurdity—presumably for reasons of formal elegance and expediency. I cannot regard this as an acceptable ground. The "classical" system, therefore, did not accept the necessitation rule and, since it accepted Leibniz's Law, it did not regard Ot , the deontic analogue of Nt , as logically true. This still seems to me a sound attitude.⁸

(d) *In standard modal logic*

Here, the operators operate on propositions. The expressions are read "It is possible (impossible, necessary) *that* . . .". The place of the blank is taken by a descriptive *sentence*. The deontic phrase "it is permitted (forbidden, obligatory)" is sometimes, in ordinary parlance, continued "that . . .". Equally often, however, or maybe more often, it is continued "to . . .".

In "it is permitted to . . ." the place of the blank is taken by a *verb* (or

⁶ *An Essay in Deontic Logic and the General Theory of Action*, p. 22 and *passim*.

⁷ This is done in my paper "Normenlogik" in *Normenlogik*, ed. by H. Lenk, Verlag Dokumentation, Pullach bei München, 1974.

⁸ However, as the reader will see from the next essay "Norms, Truth, and Logic" I have since come to take a different view of the acceptability of the "obligatoriness of tautologies". (1983).

verb-phrase) for (a category or type of) *action* or *activity*. For example: "to smoke" or "to walk on the grass".

As mentioned above (p. 103), in my first effort to build a deontic logic I regarded the variables in the deontic formulas as standing for *names* of actions. This suggests yet a third reading of the deontic operators. By names of action one could understand nouns such as "smoking" or "trespassing". On this conception, *Pp* might be a schematic representation for "smoking permitted", and *Fp* for "trespassing prohibited". It is feasible to think, however, that such phrases can be translated into the idiom using verbs for actions. "Smoking (is) permitted" and "it (one) is permitted to smoke" seem to say very much the same thing.

The readings of the deontic operators with "that" and with "to" respectively may be said to answer to two different types of deontic logic. The one is a logic of that which ought to, may or must not *be*, and the second a logic of that which ought to, may or must not *be done*. To use a terminology which has become established in German, it is a difference between a deontic logic of the *Sein-Sollen* (-*Dürfen*) and the *Tun-Sollen* (-*Dürfen*) type.⁹ The Classical System was intended to be a logic of the *Tun-Sollen*; the Standard System is by its very nature a logic of the *Sein-Sollen*-type. It follows from what has been said above that only a deontic logic of the second type *can* preserve a perfect analogy with modal logic.

It is problematic whether deontic sentences prefixing the operators to verbs of action can be "translated" into sentences prefixing the operators to sentences. Consider, for example, the sentences "it is permitted to smoke" and "it is permitted that everyone smokes". It is doubtful whether they mean the same. Another rendering of the first sentence might be "everyone is permitted to smoke"—replacing the impersonal "it" by the universal quantifier "everyone". This sentence then says that permission is given to everyone. But the sentence "it is permitted that everyone smokes" seems to say that a certain state of affairs is allowed, *viz.* the one when everyone is (maybe at the same time) smoking. This is something different from permission given to everyone.

There is of course no objection to thinking that the variables *p*, *q*, etc. of the Standard System represent sentences describing actions ("action sentences"). But this by itself, as seen from the above considerations, does not mean that the *Sein-Sollen* nature of the Standard System could capture and do justice to the *Tun-Sollen* logic which the Classical System intended to formalize. As a formal system, the Classical System is much poorer than the Standard System. But from the point of view of intended content the former aims at embracing something which seems out of reach of the latter.

In the third part of the present paper I shall try to sketch a new type of

⁹ On this distinction more will be said in the next essay. See below, pp. 196–209.

deontic logic which I hope will do justice to the intentions implicit in my first venture into the subject. But first we must say something more about action.

2. ON ACTION SENTENCES AND THE LOGIC OF ACTION

I

As the basic type of action sentence one may regard one which says that an agent *a* on an occasion *o* does a certain thing *p*. The content of such a sentence can often be viewed, alternatively, under two aspects. I shall call them the aspect of *achievement* and the aspect of *process*. The two aspects are related, loosely, to the ideas of *making* and *doing* respectively. They are also related to the distinction between *act(ion)* and *activity*.

a on *o* opens a door, say. By his activity he achieves the opening of the door and it is, at least for a short time, open. What he thus achieves is the *result* of his action in opening the door. He makes the result come about, happen. (Whether we think of the result as the *event* of the door's opening or as the *state* of its being open is, for present purposes, immaterial.)

The connection between an action and its result is *intrinsic*. Had the door not opened, the agent would not have opened the door; this is "logically true".

The opening of the door makes a creak and the noise wakes a sleeping child, say. These effects of the result of the action are also called *consequences* of the action. The connection between an action and its consequences is *extrinsic*.

The phrase "the opening of the door" is ambiguous. The process denoted can be *the door's opening*, the fact that the door opens. This can be the achieved result of an action. But it may also come about independently of action, as when the wind blows the door open. The process, however, can also be *the agent's opening the door*. This consists, for example, in his seizing the handle, pressing it down, and pushing. These are *bodily movements* and *muscular activity* displayed by the agent on the occasion of his acting. I shall call them, for short, *bodily activity*. By "the action process" I shall mean the bodily activity involved in the performance of the action.

Every action which can be viewed under the aspect of achievement also presents an aspect of process. But some actions seem to consist solely in bodily activity, for example running or walking. They need not "result" in anything, produce any state in the world which remains, at least for a short time, once the activity has ceased.

It may be argued, however, that also such "pure activities" can be viewed under an aspect of achievement. The activity of running, for example,

manifests itself in the transportation of a human body through a stretch of space. That a body was thus transported is an achievement which may be said to have resulted from the activity. Moving a limb, e.g. raising one's arm, results in a change of position of the limb in question. In the activity of moving a limb "back and forth" this position is restored but the (repeated) transportation of the limb through a stretch of space has been achieved. These observations seem to support a view according to which "pure activity" also presents an aspect of achievement and result. But whether this is a correct view or not I shall not try to decide here.

I shall use the symbol $[p] (a, o)$ as a schematic representation of the sentence that a certain agent on a certain occasion does a certain thing. The symbol " p " will have a different significance, depending upon whether we view what is being said in the sentence under the aspect of achievement or under that of process. When the sentence is viewed under the aspect of achievement, " p " is a schematic representation of a *sentence* describing either some state of affairs or some event, for example that the door is open or that it is opening. When again the sentence is viewed under the aspect of process, " p " is a schematic representation of a *verb* or verb-phrase denoting some type of action or activity, for example door-opening.

Adopting the achievement point of view, the schema $[p] (a, o)$ may be read " a on o makes (it so) that p ". Adopting the process point of view, the reading could be " a on o is p 'ing".

It should be noted that the sentence represented by " p " in our schema does not express a true or false proposition, but describes something which I propose to call a *generic* state or event. A generic state (event) is one which may or may not obtain (occur) on a given occasion o . That the state, e.g. that this door is open, obtains on the occasion o is a true or false proposition.

II

The meaning of $\sim [p] (a, o)$ is obvious, both on the achievement and the process view of action sentences. If, for example, $[p] (a, o)$ says that a on o opens the door, then $\sim [p] (a, o)$ says that a on o does *not* do this. On the achievement view this means that the agent does not make the door open—and on the process view that he does not engage in the bodily activity of door-opening.

Negation, however, need not be of the whole schema. It can also be of the part " p " in it. If " p " stands for "the door is open" (or for "the door is opening"), then " $\sim p$ " stands for "the door is not open(ing)", i.e. it says that the door is (stays) closed. This is entirely obvious. If, however, " p " stands for "opening the door", how shall we then understand " $\sim p$ " or the phrase "not-opening the door"? This is not immediately clear since the phrase can hardly be said to have a settled place within ordinary usage.

We need not, however, reject it as meaningless. It can be understood as another, somewhat "primitive", way of saying that the agent in question abstains from or *omits* opening the door. As far as he is concerned, he leaves the door closed, lets it remain closed.

One could say that, on the achievement conception, the symbolic form $[\sim p](a, o)$ signifies the un-doing of a certain existing state of affairs or its suppression on an occasion when otherwise it would come to be, and that, on the process conception, $[\sim p](a, o)$ signifies omission to engage in a certain activity.

As seen, it is possible to give a sense to the negation-sign when it stands in front of a verb or verb-phrase. It is obvious that "and" and "or" may be used for joining verbs (for example, "read and write", "read or write"). Junctors applied to sentences expressing true or false propositions, junctors applied to sentences describing generic states or events, and junctors applied to verbs or verb-phrases should be distinguished from one another. It cannot be taken for granted that they all behave in the same way logically. But we shall nevertheless use here *the same symbols* for the three kinds of junctor.

III

The notion of *omission* of action is notoriously tricky. Omission is a non-action—and yet it is, at the same time, a "mode of action or of conduct". It is something for which an agent can be held responsible. Omissions are *imputed* to agents. A logic of action, clearly, has to take this into account and treat omission as something different from mere *not* doing something.

It may be suggested that omission could be defined in terms of not doing and the notions of *ability* ("can do") and *opportunity*. On this view, *a* omitted to do a certain thing on *o* if he could have done this thing but did not do it. The expression of this view in a symbolic language requires some kind of "modal operator". If $M[p](a, o)$ means that *a* can do *p* on *o*, then $M[p](a, o) \& \sim [p](a, o)$ says that *a* on *o* omits doing it.

The notion of omission thus defined may be called "omission in the widest sense". In ordinary language, the word "omission" would hardly be used for a good many omissions in this widest sense. On most occasions there are innumerable things which I could do then but which I do not do simply because it does not occur to me to do them. (They do not fall within what may be called my "horizon of intentionality" on *o*.)¹⁰ Normally we should not say that I omitted to do these things then. But if I had a *reason* for doing or was *expected* to do some of the things, for example because it

¹⁰ On this notion see *Freedom and Determination*, p. 25 and *passim*. Cf. also the related idea of a "preference horizon" in the essay "The Logic of Preference Reconsidered" in the second volume of my *Philosophical Papers*.

was my duty or because I had promised to, then we may say that I omitted what I did not do. Such omission is often called *neglect*. If again, upon deliberation, I decide not to do some action, we call the omission *forbearance*, sometimes also *refrainment* or *abstention*.

IV

For sentences of the schematic type $[p] (a, o)$ one can build a calculus, "Logic of Action". Such calculus is "based" on ordinary propositional logic (PL) in the sense that all tautologies of PL are theorems of the calculus when action sentences are substituted for the variables (in the formulas of PL). The inference rules are those of PL, i.e. Substitution and Detachment, and no others.¹¹

When action sentences are being viewed under the aspect of process the following four principles intuitively seem valid:

- A1: $[\sim p] (a, o) \rightarrow \sim [p] (a, o)$
- A2: $[\sim \sim p] (a, o) \leftrightarrow [p] (a, o)$
- A3: $[p \& q] (a, o) \leftrightarrow [p] (a, o) \& [q] (a, o)$
- A4: $[\sim (p \& q)] (a, o) \leftrightarrow [p \& \sim q] (a, o) \vee [\sim p \& q] (a, o)$
 $\vee [\sim p \& \sim q] (a, o)$

The question may be raised whether one could not replace the fourth axiom by a weaker distribution principle to the effect that $[\sim (p \& q)] (a, o) \leftrightarrow [\sim p] (a, o) \vee [\sim q] (a, o)$. But consider what it means that an agent *omits* engaging in two different activities on one and the same occasion. The answer most in agreement with intuition seems to be that this is to omit engaging in both or omit engaging in one of them *while engaging in the other*. For example: What is it to omit (abstain from) reading-and-writing? The best answer seems to be that one reads but neglects to write or writes but neglects to read or neglects the one as well as the other. Why not simply say that it is to omit at least one of the two activities? One could say this—but it may strike one as "unnatural". For it would mean that if an agent on some occasion omitted to read, which, say, he was expected to be doing, then one could say *a fortiori* that he omitted to read-and-write although perhaps he was not expected to be writing or he cannot write or could not have written on that occasion. We need not try to decide which one of the two views on the nature of a "conjunctive omission" is the right one. If one takes the more restrictive view which also strikes me as the "natural" one, then one would have to accept for action sentences when viewed under the aspect of process the principle A4 above.

¹¹ There is a somewhat fuller description of this type of action logic in my papers "Deontic Logic Revisited", *Rechtstheorie*, 4, 1973 and "Handlungslogik" in the anthology *Normenlogik*, ed. by H. Lenk, Verlag Dokumentation, Pullach bei München, 1974.

We may define the notion of a "disjunctive activity" as follows:

$$[p \vee q](a, o) =_{df} [\sim (\sim p \& \sim q)](a, o)$$

An agent is engaged in the disjunctive activity of, say, reading or writing, if, and only if, he omits the conjunctive omission of both.¹² By virtue of A4 and A2 this is equivalent to saying that the agent engages in both or engages in the one while omitting to engage in the other.

The calculus of action sentences with the axioms A1–A4 is decidable and semantically complete.¹³ Every formula of the calculus may be shown to be provably equivalent with a formula which is a truth-functional compound of "constituents" of the simple types $[\](a, o)$ or $[\sim \](a, o)$ where the place of the blank is held by a single variable p, q , etc. Truth-values may be distributed over the constituents in a truth-table subject to the sole restriction that constituents $[\](a, o)$ and $[\sim \](a, o)$ of the *same* variable cannot both be given the value "true". This is a simple consequence of A1 or of the truth that one and the same agent cannot on one and the same occasion both commit and omit the same thing. If, observing this restriction, a formula gets the value "true" for all distributions of truth-values over its constituents, it will be said to be an action(-sentence) *tautology*. All such tautologies are provable in the calculus and provable formulas of the calculus are action(-sentence) tautologies.

We may introduce quantification into our action logic. This can happen in steps. We can quantify the sentences with regard to agents and let the sentences refer to the same (arbitrary) occasion o ; or we can quantify them with regard to the occasion and let the sentences refer to one and the same (arbitrary) agent a ; or we can combine these two modes of quantification. Finally, the calculus may also become quantified in the "proposition-like" variables p, q , etc.

For action sentences when viewed under the aspect of achievement one can also build a logic. This will have to have a somewhat more complex structure than the above "logic of action as process". In its fully developed form the variables p, q , etc. would stand not only for sentences describing results of action but also for sentences describing states which are, or are not, transformed through the action. Only then can one, for example, express in the formal system the important distinction between *productive* and *preventive* action.¹⁴

¹² I am indebted to Professor Carlos Alchourrón for a correction of a previous attempt of mine to define the notion of a disjunctive activity.

¹³ Cf. the papers mentioned in fn. 11 above.

¹⁴ A logical study of action under the achievement aspect is found in *An Essay in Deontic Logic and the General Theory of Action*, Ch. II. At that stage, however, I did not see clearly the relevance to a logic of action of the distinction between the two aspects of achievement and of process, respectively. For a new attempt, see the next essay in this volume.

V

One can distinguish between act-categories or *generic* actions, such as door-opening, murder, or smoking and act-individuals or *individual* actions, such as for example the murder of Caesar by Brutus.

Opinions differ on the question whether the deontic attributes *primarily* apply to generic or to individual actions. If one takes the view that they apply primarily to act-individuals, then the question will arise: Do they apply as operators to action-sentences or are they genuine attributes or properties of some individual things ("logical individuals")? Those deontic logicians who have opted for the second alternative in answering the first question have almost invariably opted for the first alternative in answering the second.

The question whether one can make good sense of the conception of actions as logical individuals is not uncontroversial.

Consider the schematic form of a sentence describing an individual action "*a* on *o* does *p*". (The action could be, for example, that of opening a door.) One cannot individuate this as "the action performed by *a* on *o*". It is logically possible to do more than one action on one and the same occasion.¹⁵ From the name of an individual action it must also be plain *what* it was that *a* did on *o*, i.e. we must mention a generic characteristic of the action. The phrase "the opening of the door by *a* on *o*" names an individual action or, in pure schematic form, "the doing of *p* by *a* on *o*".

Individual actions have various properties (attributes, features). The individual action of *a*'s opening a door on *o* has the "property" of (being a case of) door-opening. This is trivial. But when an action is being individuated or identified as an act of a certain category or kind the question will sometimes arise whether it may not also be classified as an action of a certain other category or kind. This question is often a preliminary to evaluating the action or to qualifying it deontically. For example: A child has been ordered to stay at home for the afternoon studying, perhaps as a punishment for a minor offence. It stays at home reading a book. Is this studying? If the child in reading was doing its homework for the school, its activity would probably count as studying. If the reading was of a novel, the child's activity would probably not count as studying. A person spat. Was what he did perhaps an act of insulting somebody? Was the killing of *b* by *a* a case of murder? When such questions are considered

¹⁵ The notion of "occasion" is vague. In this there is nothing objectionable. One could make the notion sharper by stipulating that the occasion must be restricted to the time-span of one single action. This would exclude that an agent on some occasion *first* does a certain thing and *then* another. But this restriction would not remove the possibility that at the same time as the agent did a certain thing he also did a certain other thing.

and decided, properties are in a nontrivial sense being attributed to individual actions.

The property which, as we said above, "trivially" belongs to an individual action I shall also call its *essential* property or characteristic. It is the property which we use for *identifying* ("picking out") the act-individual under consideration. Some of the properties which an action may have in addition to its essential characteristic belong to it by virtue of *the way in which* the action was performed.¹⁶ Suppose that the agent opens a door by pressing a button and pulling. Then his action, *viz.* his action of opening the door, is also a case of button-pressing and of pulling. It has these two additional characteristics.

Further additional properties may belong to an action by virtue of its consequences. Let us assume that the agent by opening a door lets cool air into the room. His action is thus also one of cooling the room. It has the property of being a "room-cooling action".

A non-essential property of an action is not necessarily either a causal prerequisite or consequence of its performance. Suppose *a* on *o* is waving his arms. In doing so he might be giving a signal. His action is thus also a case of signalling. It is this because there is a convention giving a "meaning" to the arm-waving.

The two events of a button sinking down and a room getting cool are different events from the event of a door opening. But the event of a pair of arms moving in a certain way is not a different event from the appearance of a certain signal.

Unintended consequences of an action may also constitute properties of the action. The agent who lets cool air into the room by opening the door may, as a consequence, catch a cold. His action is then a cold-giving action.

Which property of a given individual action is singled out as essentially belonging to it, is to a large extent a matter of choice. The choice may depend on our *interest* in the action, on what is *important* about it. The person who opens a window may "primarily" be ventilating a room. The essential property of his action is then that it is a case of room-ventilation. But because of the way the action was done, it was also an action of opening a window—and because of its consequences perhaps also an action of making a person sneeze.

The (causal) consequences of an action will normally materialize some time after the action was performed. At the time of performance it may

¹⁶ The "way" in which an action is performed here means an (other) action which is instrumental for the achieving of the result of the first action. This sense of "way" must be distinguished from adverbial modifiers such as (doing something) quickly or silently or well, etc. One could distinguish the two senses as "way" and "mode" (of acting), respectively.

therefore not be clear (known) which (all) the consequences will be. *a* runs over *b* with his car in the street. *b* is badly injured and dies soon after. Medical expertise attributes the death of *b* to the injury. It may be a matter of decision for a court whether *a*'s action of running over *b* should be deemed the *cause* of *b*'s death. But if it is thus regarded, *a* can correctly be said to have killed *b*; *a*'s action of running over *b* was a case of killing a man.

If the running over is regarded as the essential property of *a*'s action under consideration, then this action can be said to acquire a property, *viz.* that of being a killing, which initially it did not possess. If again the killing is regarded as the essential property, then this action may be said initially to have consisted in (causing) a car accident.

Is *a*'s running over *b* and *a*'s killing *b* one and the same action? The result of the action of running over *b* is the event that *b* gets under a car, and the result of the action of killing *b* is that *b* dies. Getting under a car, and dying are two different events (even if they take place at the same time). But *a*'s action of running over *b* and his action of killing *b* are one and the same action. Some philosophers would say that they are one and the same action "falling under different descriptions".

At the generic level, i.e. as act categories, running over and killing, door-opening and room-ventilation, etc. are, of course, *different* (types of) *actions*.

Could we not say, therefore, that e.g. the person who ran over a man in the street thereby causing his (later) death performed *two* actions? We can do this if thereby we mean that his (one) individual action on a certain occasion exemplifies two (or more) generic actions. But it is important to distinguish this from the case when a man actually performs two individual actions on one and the same occasion, for example, opens a window and closes a door. Even if those two actions take place simultaneously and not successively they would be two different *individual* actions.

Once it is accepted that actions may be regarded as logical individuals there seem to exist no obstacles of a conceptual nature to regarding also deontic attributes as properties of individual actions. One such property would be permittedness. Actions of a certain category are, let us assume, permitted. Then the performance of an action of this category by a certain agent on a certain occasion may (but need not) have been a permitted individual action (cf. below p. 122).

Not all properties of individual actions, it seems, mark generic actions. Let it be granted that deontic status, e.g., permittedness can be a genuine property of an individual action. This seems plausible. But it does not seem plausible to say that there is a generic action "doing the permitted". One cannot *identify* an individual action as being a case of doing the permitted. It must be identified as a case of doing such and such, *the doing of which is permitted*. Since permittedness cannot be used for identifying

individual actions, it cannot be an essential property of an action either. Essential properties can only be those which name act-categories.¹⁷

As noted above (p. 113), an individual action may be identified, now by one, now by another essential property. It may, for example, be identified as a case of flipping a switch or as a case of illuminating a room or as a case of alerting a prowler who was about to enter the room—to allude to a famous example from the literature in the philosophy of action.¹⁸ Depending upon which property is used for identifying the action, the set of its properties is differently divided into a subset of prerequisites and consequences. If we identify the action as one of illuminating a room, then its being prowler-alerting is a “consequential” property of the action. If again we identify it, as the prowler himself may do, as a prowler-alerting action, then both its being a room-illuminating and a switch-flipping action are accidental properties which belong to the action as its causal prerequisites.

If “two” actions have all their properties in common but different essential properties, are they then “the same” action, or not? It seems to me that we are free to mould our criteria of identity so as to answer the question either by Yes or by No. But I should prefer to answer Yes, and I have a surmise that those who prefer to say No are misled by the fact that the *individual* action under consideration exemplifies several (different) *generic* actions.

VI

Omissions too can be individuated and treated as logical individuals. The individuation of an omission is the identification (labelling) of the conduct of an agent on a certain occasion as an omission to do a certain thing. How is such an identification done? We may verify that *a* on *o* did *not* do *p*. The occasion was one when one could have done *p*; the occasion in other words provided an opportunity for doing the action. We know, e.g. from previous experience, that *a* can perform actions of the kind in question, that he has the required ability. If these are established facts, then it is also established that he omitted to do *p*—in the widest (weakest) sense of “omission” (above p. 109). We can now speak of “the omitting by *a* on *o* to do *p*” as of a logical individual. If “omission” is not understood in the weakest sense but in some stronger sense, such as not doing what one is expected or has a duty to do (above pp. 109f), then these additional criteria too will have to be taken into consideration in determining whether the

¹⁷ What is here said of deontic predicates is true also of “moral” predicates of individual actions such as, e.g., an action being “good” or “evil”, “courageous”, “temperate”, or “self-sacrificing”. Cf. *The Varieties of Goodness*, pp. 139ff.

¹⁸ This much discussed example was introduced by Donald Davidson in his influential paper “Actions, Reasons and Causes” in *The Journal of Philosophy*, 60, 1963.

agent should be said to have omitted this or that action on such and such occasion.

An omission of an action is usually "constituted" by the performance of some other action. For example, an agent is engaged in reading and thereby omits turning off the tap from which water is pouring into the bath-tub. As a consequence there is a flood in the bathroom, let us assume. We do not say that by reading he flooded the bathroom. But by omitting to turn off the tap he did so.

The agent's omission to turn off the tap does not "consist" in the tap being on (and the water pouring into the bath-tub). It "consists" in *his* reading in combination with the fact that he *could* have turned off the tap on that occasion—maybe even had a reason or was expected to do so ("instead of reading" as we should say).

There is a sense in which omissions can be called causes. What this means is that something, e.g. the tap remaining on, that happens because something else is omitted, e.g. the tap being turned off, causes a third thing, e.g. a flood, to take place.

An omission may have a less definite dating than a "corresponding" action. But in principle actions and omissions are on a par in this regard. The window was closed and *a* opened it at 11:15 a.m. On another occasion, the window remained closed the whole morning and *a* did not open it although he was there and could have opened the window, maybe even was expected to do so. The occasion for his omission to open the window was that (whole) morning.

It does not follow that the agent's opening, say, a door at 11:15 is identical with his omission that whole morning to open the window in the room. But his opening the door at 11:15 also constituted an omission to open the window at 11:15. And this "bit" of his failure to open the window in the course of the entire morning is, as an action individual, identical with his opening the door then.

Omissions can have further properties in addition to being omissions to do so and so. Someone stands by and sees another person drown. The first could have saved the second but omitted to do so. By his omission he became responsible for the death of a person. Depending upon the circumstances, a court may even pronounce his omission a case of murder.

The question may be raised whether an omission must always be "constituted" on the basis of some *other action* which the agent performs. Perhaps it is usually the case that one omits to do something because one is engaged in doing something else. But I do not think that it must be so. An agent need not do anything at all on a given occasion, he may stay completely passive. Then his passivity is omission to do every one of the things which he is able to do and for the doing of which the occasion of his passivity affords an opportunity.

VII

The Logic of Action which was described in outline in Section III is a logic of *action-sentences* of the schematic prototype form "*a* on *o* does *p*". The actions described in such sentences, we have seen, may be regarded as logical individuals, the prototype name form of which is "the doing of *p* by *a* on *o*". Under this conception of actions we get yet another type of Logic of Action. Its objects of study are sentences attributing properties to individual acts. The prototype form of such sentences is "the doing of *p* by *a* on *o* is *A*" where "is *A*" is a schematic representation of such phrases as, for example, "is (a case of) murder" or "is ventilating a room". One can, if one wishes, call such sentences action-sentences too. But then it should be remembered that, unlike the above prototype form of such sentences, they do not say that something or other is being done, but that something or other which was done has a certain characteristic or property.

Similarly, we shall have to count with sentences attributing properties to individual omissions. "The omission of *p* by *a* on *o* is *A*" might say that *a*'s omission to save a person from drowning was, on that occasion, a case of murder.

I shall use *x*, *y*, etc. as variables for individual actions or omissions and *A*, *B*, etc. as schematic representations for names of properties. Names of properties will also be called *predicates*.

The logic of the sentences now under consideration could be regarded as simply a fragment of "classical" (monadic) predicate logic and quantification theory. Then it is of no independent interest as a "logic of action".

There is, however, good reason for studying sentences attributing properties to individual actions within a more "refined" calculus than the traditional predicate calculus. I have elsewhere described this more refined calculus and coined for it the name Logic of Predication.¹⁹ Its characteristic feature is that it allows us to make a distinction between denying that an individual has a certain property and affirming that it lacks a property. The distinction, in other words, is between two kinds of negation, an external negation which is of sentences (propositions) and an internal negation which is of properties. To use Aristotle's example,²⁰ between something not being white and something being not-white.

¹⁹ "Remarks on the Logic of Predication", *Ajatus*, 35, 1973.

²⁰ Cf. *Analytica Priora* 52a1-2 and 52a25-. Cf. also my paper "On the Logic of Negation", *Societas Scientiarum Fennica, Commentationes Physico-Mathematicae*, XXII (4), 1959, in which originally I introduced and discussed the distinction between the two types of negation, weak and strong, external and internal—which here I distinguish with the aid of the symbols $\sim$, [] and [-].

What then is the difference between (simply) not having a property and lacking a property? Roughly speaking: the things which lack a given property fall within the "range" of that property: they *could* have the property in question although in fact they have not got it; things outside the range neither have nor lack the property in question.

This is a rough characterization only and its application to specific properties is, often at least, a matter of decision. It is for us to give a meaning to the distinction in question. This, however, can often be done in a way which seems both enlightening and natural. Thus, for example, that an action is not permitted can be taken to mean that it is forbidden, or it can mean that it simply has no deontic status at all. What is forbidden "lacks" permittedness; an action void of deontic status neither has nor lacks permittedness.

If there are several ways of doing an action and the action is performed in one of the ways to the exclusion of the other, then it lacks the characteristic of being an action of the second kind. For example: Let it be that one can open a door either by pressing a button or by turning a key. Then an individual act of opening this door may (accidentally) have the property of being an act of button-pressing and lack the property, which it could have possessed, of being an act of key-turning.

When an action lacks a property which an action performed by that agent on that occasion could have had, it is, normally, "constitutive" (above, p. 116) of an omission. If a child is reading a novel instead of the text he is supposed to be studying (and which he could have read then), his action of reading also constitutes an *omission* of his to study and can therefore be said to lack the property of being a case of studying.

VIII

The same device as before, square brackets [], will enable us to mark the distinction between not having a property and lacking it. Thus $\sim [A]x$ says that x is *not* A , and $[\sim A]x$ says that x is not- A . The axioms of a Logic of Predication are, with minor notational difference, the same as those of our above Logic of Action. One can debate whether a weaker version (cf. above, p. 110) of A4 is valid for predications. Ordinary usage of the negation words is hardly settled, so the answer to the question is a matter of decision. On the whole it seems to me more natural to opt for the strong version. This would mean that a thing is said to lack the conjunction of two properties if, and only if, it belongs in the range of both but has at most one of the two. The axioms are then:

- A1. $[\sim A]x \rightarrow \sim [A]x$
- A2. $[\sim \sim A]x \leftrightarrow [A]x$
- A3. $[A \& B]x \leftrightarrow [A]x \& [B]x$
- A4. $[\sim (A \& B)]x \leftrightarrow [A \& \sim B]x \vee [\sim A \& B]x \vee [\sim A \& \sim B]x$

We can now also define the notion of a "disjunctive property": $[A \vee B]x =_{df} [\sim (\sim A \& \sim B)]x$. By virtue of A4 and A2 it follows that, for example, something has the property "red or round" if it has the one but lacks the other or has both the "simple" properties. But if "red" denotes the colour and "prime" a characteristic of some numbers, then nothing has the property "red or prime". There simply is no such disjunctive property because the range of things which are possibly red and of those which are possibly prime (numbers) have no common member.

The rules of inference are the usual ones of Substitution and Detachment.

For quantified sentences one would have two additional axioms:

A5. $(Ex)([A]x \vee [B]x) \rightarrow (Ex)[A]x \vee (Ex)[B]x$

A6. $\sim (Ex)([A]x \& \sim [A]x)$

and an additional inference rule (Leibniz's Law) to the effect that formulas which are provably equivalent on the basis of A1–A4 are interchangeable *salva veritate* in quantified formulas.

3. DEONTIC LOGIC—A NEW APPROACH

I

Let it be agreed that deontic status can, in the genuine sense, be predicated of *individual* actions. I shall use "*F*" for the property of being forbidden, "*O*" for that of obligatoriness, and "*P*" for permittedness.²¹ " $[F]x$ " may be read " x is forbidden". " $\sim [F]x$ " says that x is not forbidden, and " $[\sim F]x$ " that x is not-forbidden, that it lacks the property of being forbidden. In what way the second is a stronger statement than the first will be discussed presently.

Undeniably, deontic status is often also attributed to generic actions or categories of action. I shall use the letters "*F*" to stand for "forbidden", "*O*" for "obligatory", and "*P*" for "permitted" when this kind of attribution of deontic status is in question. Under this use, the deontic words are not predicates, but operators. About the difference more will have to be said later.

Let " $[A]x$ " say, for example, that x , an individual action, is a case of murder. " $F A$ " then says that murder is forbidden or that it is forbidden to (commit) murder. The *kind of action* called "murder" is forbidden.

The expressions formed by deontic operators followed by a predicate or a molecular compound of predicates denoting generic actions may be used

²¹ This use of the three letters, *O*, *F*, and *P*, is different from the use made of them in the first part of the present study and thereby also different from their established use in writings on deontic logic.

to express *norms* (of action). Norms are given to agents acting on certain occasions. Norms can be either for named individual agents, or for agents of a certain category, or for agents unrestrictedly. Analogously, they can be either for specified individual occasions, or for occasions satisfying certain conditions, or for occasions unrestrictedly (which provide opportunities for doing the actions in question).

II

The attribution of deontic status to individual actions will be called *deontic predication*.

What does it mean that an individual action x is a forbidden action? As was indicated above (pp. 114f), one cannot pronounce an individual action forbidden unless one has first identified it as an action of a certain category or kind. Assume that A is the essential property used for identifying the action. If there is a norm $\mathcal{F} A$ prohibiting actions of this kind, then x is (was) a forbidden action. Let us think, however, that there is no such norm. It does not follow that the action then is not forbidden. For x may possess some other property beside A , say B , such that actions of *that* category are forbidden. Then, obviously, x was a forbidden action (to the agent who on some occasion did it).

We can now define what it is for an individual action x to be a forbidden action, as follows: $[F]x =_{df} (EX)([X]x \ \& \ \mathcal{F} X)$. In short: an action is forbidden if, and only if, it falls under some forbidden category of action. Or, in other words: an agent's action on some occasion is forbidden if, and only if, in performing this action he does something forbidden.

The commission of an individual action is obligatory if, and only if, the action is of a kind such that it is forbidden to omit actions of this kind. Conversely, the omission of an individual action is obligatory if, and only if, the omission is of a kind such that it is forbidden to commit actions of this kind.²²

This interrelatedness between obligation and prohibition in the terms of commission and omission of actions calls for some further comments.

Consider the following example. The agent a enters a garden on an occasion o . The action is thus the entering of the garden by a on o . This action, let us assume, can be performed in three different ways. One can enter the garden either through one of two gates, g_1 and g_2 , or by jumping the fence surrounding it (which is low). It is, however, forbidden to jump the fence. (There are flowers at the foot of the fence.) The agent entered the garden through g_1 . His action of entering the garden was thus also a case of passing through g_1 . If x = the entering of the garden by a on o , and

²² I am indebted to Carlos Alchourrón and Eugenio Bulygin for useful observations relating to these matters.

A is the "property" of being a passing through g_1 , then the action x is A . The action *could* have been a passing through g_2 (B) or a jumping the fence (C). But it *lacked* (p. 118) these two properties. In our symbolism: $[\sim B]x$ and $[\sim C]x$. By virtue of lacking the property C , the action x also constitutes (p. 116) an omission on the part of a to jump the fence on o . If the fact that the action x lacked the property C is considered sufficient ground for saying that a omitted to jump the fence on o , then this omission on a 's part was obligatory. His action, what he *did*, was not, as such, obligatory. But in entering the garden a behaved *in accordance with duty* since he passed through the gate and observed the prohibition to enter by jumping the fence.

Obligatory omissions of action, i.e. the observance of prohibitions, could also be called "*negative*" obligations.

Assume next that our agent had been commanded to enter the garden and to do so through gate g_1 . He was, in other words, not only forbidden to jump the fence but also to pass through g_2 . Then his individual action x was an obligatory action by virtue of the fact that it had the property A , viz. that of being a case of passing through g_1 . Had it lacked this property, it would have been a forbidden action.

Obligatory commissions of action might also be called "*positive*" obligations. Positive obligations often have the character of fulfilling *commands* (orders, imperatives).

As seen the *predicates* "forbidden" and "obligatory" can both be defined in terms of the operator "forbidden". The *operator* "obligatory" again can be defined in terms of the operator "forbidden", thus:

$$OX =_{df} F \sim X.$$

The two operators are interdefinable. This is in agreement with the "traditional" view of the matter.

Neither "in logic" nor "in real life" is there anything to prevent one and the same individual action (or omission) from being both obligatory and forbidden. If Jephthah had sacrificed his daughter, his action would have been obligatory because it was the fulfilment of a promise to the Lord, and forbidden because it was homicide. $[F]x$ & $[O]x$ is not a contradiction. It says that the individual action x is of a kind which is forbidden but also of a kind which is obligatory. It is forbidden by virtue of one of its characteristics and obligatory by virtue of another.

The *predicates* "forbidden" and "obligatory", be it observed, are *not* interdefinable. This is a simple consequence of the fact that individuals cannot be "negated". " $[F] \sim x$ " and " $[O] \sim x$ " are meaningless signs.

III

To deny that an individual action is forbidden is to affirm that it does not

fall under any kind of action which is forbidden or, in other words, that *all* its features signify not-forbidden properties (of individual actions). In symbols:

$$\sim [F]x \leftrightarrow \sim (EX)([X]x \& \mathcal{F}X) \leftrightarrow (X)([X]x \rightarrow \sim \mathcal{F}X).$$

Shall we say that an action which is not forbidden is thereby permitted? This is an aspect of the much debated question whether permission is anything "over and above" the absence of prohibition. I think we are well advised to distinguish between things being permitted in the weak sense of simply not being forbidden and things being permitted in some stronger sense. Exactly in what this stronger sense "consists" may be difficult to tell. That which is in the strong sense permitted is, somehow, expressly permitted, subject to norm—and not just void of deontic status altogether.

The predicate "strongly permitted" we can define as follows: $[P]x =_{df} (X)([X]x \rightarrow \sim \mathcal{F}X) \& (EX)([X]x \& \mathcal{P}X)$. The (in the strong sense) permitted individual action does not fall under any forbidden kind of action but falls under at least one (in the strong sense) permitted one. This definition, of course, does not say anything about the meaning of the (strong) permission-operator.

As easily seen from the above, we have $[P]x \rightarrow \sim [F]x$.

IV

Nothing has so far been said to give meaning to the *lack* of the properties F , O , and P . An action x of which it is true that $[F]x$ falls under some norm prohibiting a certain kind of action. An action x for which it is true that $\sim [F]x$ need not fall under any norm at all. But it *may* fall under a permissive or obligating norm. And similarly for the expressions $[O]x$ and $[P]x$ and their negations.

The following suggestions therefore appear natural: That an action *lacks* the property of being forbidden means that it is *not* forbidden but is either obligatory or (in the strong sense) permitted. That an action *lacks* the property of being obligatory means that it is *not* obligatory but is either permitted or forbidden. That an action *lacks* the property of being permitted, finally, means that it is *not* permitted but is either forbidden or obligatory. Thus we have the following three identities:

$$\begin{aligned} [\sim F]x &=_{df} \sim [F]x \& ([O]x \vee [P]x) \\ [\sim O]x &=_{df} \sim [O]x \& ([P]x \vee [F]x) \\ [\sim P]x &=_{df} \sim [P]x \& ([F]x \vee [O]x) \end{aligned}$$

The identities imply equivalences which may then be distributed into conjunctions of (two) implications. Since, in the Logic of Predication, lack of a property in a thing entails that the thing in question has not got this property, we also have the following relations: $[\sim F]x \rightarrow [O]x \vee [P]x$

and $[\sim O]x \rightarrow [P]x \vee [F]x$ and $[\sim P]x \rightarrow [F]x \vee [O]x$. Moreover, since we already proved $[P]x \rightarrow \sim [F]x$ we can now from the first of the above three identities derive the stronger formula $[P]x \rightarrow [\sim F]x$.

V

I have thus taken the view which seems to be reasonable, that when deontic status is predicated of an individual action, this predication is grounded in the deontic status of some category or kind of action under which this individual action falls. As a consequence, we have to regard the deontic predicates as *secondary* to the deontic operators.

By *normative sentences* I shall understand expressions of the forms $\mathcal{O} -$, $\mathcal{F} -$ and $\mathcal{P} -$ and their molecular compounds, the place of “-” being taken by an atomic or molecular predicate (of actions).

Normative sentences will be called *norm-formulations*. A characteristic use of them is for giving (issuing, laying down) norms or rules for human agents. When this use is in question, the normative sentences may be said to *express norms* (cf. above, p. 120).

Normative sentences, however, can also be used for making statements to the effect that there are (have been given or issued) such and such norms or rules. When used in this way, normative sentences express what I propose to call *norm-propositions*.

This ambiguity of usage is a very characteristic and important feature of atomic norm-formulations. Also molecular normative sentences can be used either for expressing norms or for expressing norm-propositions. But their use in the second way seems much more common.

Let there be a prohibition with a disjunctive content (action), $\mathcal{F}(A \vee B)$. Individual actions with the generic characteristic “ $A \vee B$ ” are actions which either have both the characteristics or have the one but lack the other. $[A \vee B]x \leftrightarrow [A \& B]x \vee [A \& \sim B]x \vee [\sim A \& B]x$ is a logical truth in the Logic of Predication. This fact will also be expressed by saying that the predicates $A \vee B$ and $A \& B \vee A \& \sim B \vee \sim A \& B$ are logically equivalent (predicates). We shall lay down the following

Principle of Deontic Equivalence: Logically equivalent predicates are intersubstitutable (*salva veritate*) in norm-sentences (expressing norm-propositions).

By virtue of this principle we may regard $\mathcal{F}(A \vee B) \leftrightarrow \mathcal{F}(A \& B \vee A \& \sim B \vee \sim A \& B)$ as a “truth of deontic logic”. Now consider the following: Any individual action by the performing of which an agent may observe or violate this prohibition is an action which can have or lack the characteristic *A* and can have or lack the characteristic *B*. For this reason it appears natural to say that a prohibition of actions with a disjunctive characteristic is logically equivalent with a conjunction of prohibitions of actions of any

one of the various kinds (the individual members of) which fall under the disjunctive kind. Thus we may regard as a logical truth about norms the formula

$$(1) \quad \mathcal{F}(A \vee B) \leftrightarrow \mathcal{F}(A \& B) \& \mathcal{F}(A \& \sim B) \& \mathcal{F}(\sim A \& B).$$

Let there be a norm to the effect that actions of the category A are forbidden. Then an action x with the property A is a forbidden (individual) action: $\mathcal{F} A \& [A]x \rightarrow [F]x$. This can also be written: $\mathcal{F} A \rightarrow ([A]x \rightarrow [F]x)$. Since this holds for any arbitrary individual action, we also have $\mathcal{F} A \rightarrow (x)([A]x \rightarrow [F]x)$.

The first implication in the formulas, be it observed, does not hold in the reverse direction. It might be the case that all actions of the category A actually are forbidden actions, though not on the ground that they have the property A , but because every one of them happens to belong to *some* (not necessarily the same) forbidden kind of action. We are thus not suggesting, which would be quite wrong, that norms are logically equivalent with general statements about the deontic character of individual actions of certain kinds. (Deontic attributes of individual actions, be it remembered, were defined with the aid of deontic operators.)

From the fact that actions of a certain type are forbidden it thus follows that all individual actions of this type are forbidden; but from the fact that actions of a certain type are permitted it does not follow that all individual actions which are of this type are permitted individual actions. In doing something which is, "as such", permitted an agent may also be doing something which is forbidden. He may, for example, do what he does *in a manner* which is forbidden. Or his action may *lead to* something forbidden.

If a disjunctive action, for example reading or writing, is (generically) permitted, then any individual action which has both the permitted features and no forbidden feature is permitted; and so is also any individual action which has one of the permitted features, is lacking the other, and has no forbidden feature. This seems as good a ground as could possibly be required for holding that

$$(2) \quad \mathcal{P}(A \vee B) \leftrightarrow \mathcal{P}(A \& B) \& \mathcal{P}(A \& \sim B) \& \mathcal{P}(\sim A \& B)$$

is a norm-logical truth.

We already noted that, if an individual action is permitted; then, by definition, it cannot be a forbidden action. (Above p. 122.) On this basis we may hold that a norm permitting actions of the type A excludes a norm prohibiting actions of that same type:

$$(3) \quad \mathcal{P}A \rightarrow \sim \mathcal{F}A.$$

What shall we think about the principle $\sim (\mathcal{F}A \& \mathcal{F}\sim A)$ and its equivalent form $\sim (\mathcal{O}A \& \mathcal{O}\sim A)$?

First we must warn against a misunderstanding. $\mathcal{F}A \& \mathcal{F}\sim A$ would

not have as a consequence that an agent could not perform anything but forbidden actions, so that whatever the poor man does he sins against the law (norms). Because not all individual actions need be such that they either have or lack the feature *A*. They may not fall within the range of actions of this kind at all. (An agent who does *not* do a certain thing omits doing it, we have said (p. 109), only if, on the occasion in question, he *could* have done it.) Why does nevertheless $\mathcal{T} A \ \& \ \mathcal{T} \sim A$ strike us as absurd? Obviously because, for reasons of logic, an agent who is in position to do an action of the type *A* will, whether he does it or not, commit an offence. Is this a *logical* impossibility? Rather than calling it a logical impossibility we should, I think, say that a normative order which happens to contain those two prohibitions is "irrational"—and that therefore the legislator should, "in the name of rationality", lift at least one of the two prohibitions or make them conditional upon different types of situation when they apply.

A "rational" normative order should therefore satisfy the principle

$$(4) \quad \sim (\mathcal{T} A \ \& \ \mathcal{T} \sim A).$$

VI

The formula $\mathcal{T} (A \ \& \ B) \rightarrow \mathcal{T} A$ cannot be accepted as generally valid in a sound logic of norms. From the permittedness, in the strong sense, of the kind *A* & *B* of actions, one cannot conclude to the permittedness, in the strong sense, of the kind *A* of actions. But, as we shall see presently (below p. 127), the permittedness of the kind *A* & *B* is "rationally incompatible" with the forbiddenness of the kind *A* (and the kind *B*). Hence one may from the permittedness, in the *strong* sense, of the kind *A* & *B* of actions conclude to the permittedness, in the *weak* sense, of the kind *A*.

What then of the distribution formula $\mathcal{O} (A \ \& \ B) \leftrightarrow \mathcal{O} A \ \& \ \mathcal{O} B$?

Consider first the implication from left to right. This clearly—although contrary to what might be expected from knowledge of "traditional" deontic logic—cannot be a truth of logic. From the fact that an agent is under an obligation to perform actions which exhibit two characteristics, it does not follow that he is under an obligation to perform actions which have (only) one of the characteristics. From the fact that he has been ordered to enter a garden through a certain gate, it does not follow that he ought to enter the garden regardless of how he does it. We need not think that he is forbidden to jump the fence. But if he finds the gate locked and cannot open it, then he cannot conclude "logically" that, because of the order given, he must now jump the fence—a feat which, let us assume, he could perform.

Consider next the implication from right to left. The agent has two obligations. It is not certain that he can satisfy both by performing just

one individual action. But it may be that, on some occasions, it is possible to satisfy both obligations by doing just one thing *and in no other way*. Then clearly he is obligated to do this conjunctive action. If, however, he can, on that same occasion, do something which satisfies the one and something else which satisfies the other of his two obligations, then there is no need for him to do the *one* action which satisfies both.

Example: An agent has been asked to see to it that the window and the door in a certain room are shut. If he finds the window open, he ought to shut it—and similarly with the door. If he finds both open, he has to shut both. Perhaps he can do this by operating a mechanism, say by pressing a button. If that is the *only* way this can be done, he ought of course to press the button. But if one can also shut the window and the door separately he is not obligated to do this by pressing the button.

Thus there is not a relation of entailment either way between the conjunction $\mathcal{O} A \ \& \ \mathcal{O} B$ of two norms and the conjunctive obligation $\mathcal{O} (A \ \& \ B)$.

VII

Let it be that $\mathcal{F} A$. In the Logic of Predication we prove $(x)([A \ \& \ B]x \rightarrow [A]x)$. This means that if an agent performs an action with the two characteristics A and B he will necessarily disobey the norm $\mathcal{F} A$. Must we not therefore say that if there is a prohibition to the effect that $\mathcal{F} A$ then there is also (implicitly at least) a prohibition to the effect that $\mathcal{F} (A \ \& \ B)$? So that $\mathcal{F} A \rightarrow \mathcal{F} (A \ \& \ B)$ would be a "law of deontic logic"?

I do not think that there is any clearcut answer Yes or No to this question. From the fact that there are individual actions of the kind A it does not follow that there are any of the kind $A \ \& \ B$. Maybe it is quite impossible, either for reasons of logic or for reasons of human ability, to perform actions with these two characteristics, although it is possible and maybe even easy to perform actions with either one of the characteristics. It seems silly then to say that actions of the (empty) kind $A \ \& \ B$ are forbidden, on the grounds that actions of the kind A are forbidden. Maybe the lawgiver would even permit actions of the kind $A \ \& \ B$, if they could be done. (Perhaps possessing the characteristic B would "make good" for the bad which actions with the characteristic A do and which motivated the prohibition $\mathcal{F} A$.)

Assume now that actions of the kind $A \ \& \ B$ actually are permitted. We thus have a norm $\mathcal{P} (A \ \& \ B)$. Must the lawgiver then repeal the norm $\mathcal{F} A$ if there was one before? If there are the two norms $\mathcal{P} (A \ \& \ B)$ and $\mathcal{F} A$, then no agent could avail himself of the permission without breaking the prohibition. This is logically true. And this fact would make it, if not "illogical" at least "irrational" to let permission $\mathcal{P} (A \ \& \ B)$ and prohibition $\mathcal{F} A$ co-exist within the same code of norms. A rational code should

therefore satisfy the principle

$$(5) \quad \mathcal{P} A \ \& \ (x)([A]x \rightarrow [B]x) \rightarrow \sim \mathcal{F} B.$$

Since $(x)([A]x \rightarrow [A]x)$ is logically true, it follows from (5) that $\mathcal{P} A \rightarrow \sim \mathcal{F} A$ or that what is (generically) permitted cannot be (generically) forbidden.

Thus the principle (3) which was already accepted as a "truth of deontic logic" is seen to be a consequence of a more general deontic principle to the effect, loosely speaking, that "what follows from the permitted cannot be forbidden".

In a similar manner it may be shown that the principle (4), one of the traditional corner stones of a deontic logic, is but a special case of a more general principle which, in the name of "rationality" if not in that of "logic", seems acceptable, viz.

$$(6) \quad \mathcal{F} \sim A \ \& \ (x)([A]x \rightarrow [B]x) \rightarrow \sim \mathcal{F} B.$$

This principle says that if all (individual) actions with an obligatory characteristic also have another characteristic, then this other characteristic cannot be (generically) prohibited. Let it be observed, however, that this other characteristic need not itself be (generically) obligatory; it might even be the case that all actions which are B without also being A are forbidden actions.

If in (6) we for " B " substitute " A " we obtain $\mathcal{F} \sim A \ \& \ (x)([A]x \rightarrow [A]x) \rightarrow \sim \mathcal{F} A$ which reduces to $\mathcal{F} \sim A \rightarrow \sim \mathcal{F} A$ which is the same as $\sim (\mathcal{F} A \ \& \ \mathcal{F} \sim A)$ or $\sim (\emptyset A \ \& \ \emptyset \sim A)$.

VIII

The facts upon which I have here based my arguments for accepting and for not accepting certain statements about the relations between norms are *logical truths*. They are derived from the definitions of the deontic predicates with the aid of principles of the Logic of Predication. The accepted statements themselves we might call truths of (a) Deontic Logic. But I feel a certain hesitation calling them "logical truths" at all. The reason for this is that it seems to be a matter of extra-logical decision when we shall say that "there are" or "are not" such and such norms. Shall we, for example, say that "there is" a $\mathcal{F}$ -norm with a disjunctive norm-content, if there are (have been given, issued) norms concerning all the various ways in which this disjunctive norm-content may be realized through action? Perhaps the norm with the disjunctive norm-content was never formulated or even thought of. Yet it was there "implicitly" one could say. Had the norm been given in the disjunctive formulation, it would have imposed exactly the same demands and granted exactly the same freedom to agents as the norms about the disjuncts would have done jointly.

The distribution principles (1) and (2) are conceptually on a somewhat different footing from the principles (3) and (4) and the more general principles (5) and (6) from which (3) and (4) may be derived. (1) and (2) are in a sense "linguistic", concern the way "and" and "or" are used when speaking of prohibitions and permissions. The principles which say that the permitted cannot also be forbidden or that prohibitions (obligations) with contradictory contents cannot co-exist are more of the nature of requirements of rational legislation than of strictly logical thinking.

Thanks to the distribution principles any molecular compound of norm-formulations can be split up into atomic constituents consisting of the letters " $\mathcal{T}$ " and " $\mathcal{P}$ " followed by atomic predicates or conjunctions of predicates and their negations. Over these constituents we can distribute truth-values subject to the two restrictions imposed by the principles (5) and (6). For example: If there are two constituents $\mathcal{T}A$ and $\mathcal{P}(A \& B)$ and the first is given the value "true", then the second must be given the value "false". If, under all permissible distributions of the truth-values, the molecular formula assumes the value "true", it might be called a "deontic tautology".

IX

The deontic operators which we have been studying so far are prefixed to names of action-categories. The "deontic logic" which emerged from this study, could be called a logic of what one ought to, may or must not *do*. The things which tell what ought to, may, or must not be done, we called *norms*.

Deontic operators, however, can also be prefixed to action-sentences—or to sentences generally. For example: it ought to be the case that *a* on *o* does *p*.

In order to avoid confusion I shall introduce the symbol N_d for obligation ("deontic necessity") and M_d for permission ("deontic possibility").²³ No special symbol for prohibition will be needed now. For sentences I shall employ symbols *s*, *t*, *u*, . . . from the end of the alphabet.

One can build a deontic logic for sentences of the form " $N_d s$ ", " $M_d s$ " and their molecular compounds. Such sentences say that a certain thing, e.g. that so and so does that and that on such and such an occasion, ought to or may or must not *be*.

This deontic logic may rightly be regarded as an off-shoot of modal logic. Just as there are several modal logics, one may also construct several systems of such a deontic logic. But I see no particular reason why it should be constructed in a manner which deviates considerably from the well-known modal logics—except in that it rejects the formula $s \rightarrow M_d s$. I

²³ These symbols N_d and M_d correspond to the symbols *O* and *P* respectively of traditional deontic logic.

shall not here inquire into the interest of preserving within such a deontic logic the distinction between strong and weak permission and of having a permission operator which distributes conjunctively over disjunctions. Perhaps there is no good motivation for these peculiarities at all.

The more interesting variations of this type of deontic logic arise, I think, when instead of the variables s , t , etc. we employ action-sentences $[p]$ (a , o), etc. and their compounds. Then the basis on which the deontic logic stands is not propositional logic alone but also a Logic of Action of the type sketched in Part II of this paper.

Of particular interest will now be problems connected with quantification.

There can be no obstacles of a logical nature to applying deontic operators also to quantified action sentences. For example: $N_d(a)(o) [p]$ (a , o) says that such a state of affairs is obligatory that everybody on all occasions does p .

The logical situation changes radically when we consider quantification *into* deontically qualified contexts. Consider, for example, the expression $(a)(o)N_d[p]$ (a , o). Here " $[p]$ (a , o)" is *not* a sentence which expresses a true or false proposition but a *generic* sentence (open sentence, propositional function).

What could $(a)(o)N_d[p]$ (a , o) possibly mean? It might be an attempt to say that a norm which makes obligatory actions of a certain *kind* is addressed to all agents on all occasions. Then we are not concerned with the obligatory character of a certain *state of affairs* but with a norm obliging *agents*. Instead of $N_d[p]$ (a , o) we have a norm θA which is being given to every agent and is for all occasions. How this universal character of the norm shall be properly expressed in a symbolism is not easy to tell. Perhaps we should introduce a *normative relation*, R , which holds between an agent, an occasion, and a norm, $R(a, o, \theta A)$. This relational sentence could then be quantified, e.g. so as to become $(a)(o)R(a, o, \theta A)$.

In which relation do the expressions $(a)(o)R(a, o, \theta A)$ and $N_d(a)(o)[p]$ (a , o) stand to one another? We shall assume that the action which is the doing of p by a on o is an action of the category A . The answer to the question, I suggest, is as follows: The two types of deontic expression are not, "by themselves", related to one another in any way whatsoever. If it is a deontic necessity that all agents always do a certain thing it does not follow that there is any norm addressed to all agents ordaining them always to do this thing. But a lawgiver who is anxious to see to it that, if possible, all agents on all occasions behave in this way may issue a norm to the said effect. He may, however, issue such a norm also *without* deeming it necessary for any particular end of his that all agents always behave as he has ordained.²⁴

²⁴ The vistas which here open are discussed more fully in the concluding part of the next essay. (1983)

Norms, Truth, and Logic

1. DEONTIC LOGIC AS A LOGIC OF NORM-SATISFACTION

I

A philosophic difficulty connected with the idea of a "logic of norms" has been the following:

A "logic", presumably, is concerned with such relations as entailment (logical consequence) or compatibility and incompatibility between the entities which it studies. The most natural way of explaining what is meant by those relations makes use of the notion of truth and of various truth-functional notions. For example: that something follows logically from something else seems to "mean" (something like) that if the second is *true* the first must be true, too.

It is, however, a common, even if not universally shared, opinion among philosophers that norms have no truth-value, are neither true nor false. Therefore it is at least doubtful whether norms can have a "logic"—whether for example one norm can be said to follow logically from another norm.

Ernst Mally,¹ who was the first to try to develop in a systematic fashion a logic of norms, seems not to have been troubled by such doubts. The first to articulate them were some philosophers and philosophical logicians of the logical positivist trend in the 1930s.² Normative discourse was by those thinkers labelled "atheoretical" and sometimes even said to be "meaningless". Norms were assimilated to imperatives and value-judgements to exclamations, i.e. to entities which quite obviously cannot be said to have truth-value.

Like Mally, I too was not troubled by the problem of truth when in 1951 I designed my first system of deontic logic.³ This is perhaps surprising, considering that I then was, and still am, firmly of the opinion that (genuine) norms lack truth-value. At first, I did not "combine" this opinion of mine with my work in logic. But soon I noticed that I had been overlooking a problem. My first reaction to it was to think that logic "has a wider reach than truth". "Deontic logic," I wrote in the Preface to my book *Logical Studies* (1957), "gets part of its philosophic significance

¹ *Grundgesetze des Sollens, Elemente der Logik des Willens*, Leuschner & Lubensky, Graz, 1926.

² Notably Dubislav, Jørgensen, and Alf Ross. Cf. W. Dubislav, "Zur Unbegründbarkeit der Forderungssätze", *Theoria*, 3, 1937; J. Jørgensen, "Imperatives and Logic", *Erkenntnis*, 7, 1937–38; and Alf Ross, "Imperatives and Logic", *Theoria*, 7, 1941.

³ In the paper "Deontic Logic", *Mind*, 60, 1951.

from the fact that norms and valuations, though removed from the realm of truth, yet are subject to logical law."

I know that some people have hailed this extension of the province of logic beyond the confines of truth as a satisfactory answer to the problem. It seems also to agree with the view implicitly taken by the vast majority of writers on deontic logic since the time of the subject's (re-)birth in the early 1950s. Few only have bothered to question its superficiality.

II

In *Norm and Action* (1963) I made a tripartite distinction which I think is useful between norms, norm-formulations, and norm-propositions. Norm-formulations have a characteristic "ambiguity": one and the same form of words may be used both *prescriptively* to enunciate a norm or a rule of conduct and *descriptively* for stating that there is (exists) such and such a norm or rule. Accordingly, I distinguished between a prescriptive and a descriptive interpretation also of the formalized counterparts of norm-formulations constructed by means of the deontic operators *O* and *P* and constants and variables belonging to a Logic of Action, of Change, and of Propositions. My view was then that Deontic Logic is a logic of descriptively interpreted formalized norm-formulations. This made the application of truth-connectives and of such meta-logical notions as entailment, consistency, and contradiction uncontroversial. But something remained problematic about the whole undertaking. As said in the book (p. 134): "Deontic Logic is a logic of descriptively interpreted expressions. But the laws (principles, rules) which are peculiar to this logic, concern logical properties of the *norms* themselves, which are then reflected in logical properties of norm-propositions. Thus, in a sense, the 'basis' of Deontic Logic is a logical theory of prescriptively interpreted *O*- and *P*-expressions."

The deontic logic developed in *Norm and Action* was thus a sort of conflation or mixture of a "logic of norms" and a "logic of norm-propositions". The basis of the logical principles *at the descriptive level* governing the norm-propositions were definitions *at the prescriptive level* of norms of the notions of self-consistency (of a norm), consistency (of sets of norms), and entailment (between norms). I was aware of the fact that these notions were not "purely logical" but related to ideas about *rationality* ("rational willing").⁴ This, I think, was a basically sound insight. I did not then draw from it what now seems to me the appropriate conclusions.

For the sake of arriving at clarity in these matters it is necessary to make a *sharper* separation than was done in *Norm and Action* between the descriptive and the prescriptive interpretation of the deontic formulas. On the descriptive interpretation they express propositions to the effect that

⁴ Cf. *Norm and Action*, pp. 151f.

such and such norms exist. These propositions are true or false and obey the laws of "ordinary" logic. If, in addition, they obey special principles of a logically necessary character this must be due to conceptual peculiarities connected with the notion of the *existence* of a norm. We need not doubt that there are such peculiarities. In *Norm and Action* I said something about them⁵—but I shall not here repeat or defend my past position.

On the prescriptive interpretation deontic formulas have a "prescriptive meaning" and do not express true or false propositions. It makes no sense to speak of relations of contradiction or entailment between the formulas when thus interpreted. The positivistic sceptics who, like Alf Ross, doubted the possibility of a deontic logic were in an important sense right in maintaining that *norms have no logic* or that normative discourse is "alogical". But norm-giving activity, and also the norms themselves, can be judged under various aspects and standards of *rationality*. Some such aspects, moreover, can be associated with considerations of a strictly logical character—and in the logical patterns which emerge from such considerations we can recognize the structures which deontic logicians (in the now traditional sense of the term) have studied and tried to systematize. How this happens I hope to be able to show in the sequel.

III

I shall in the following construct more than one formal system of deontic logic. The calculus or system constructed in this first part of the present paper answers, relatively closely, to what may be called the *standard* systems of deontic logic.⁶ The calculus has the following vocabulary:

- (a) Variables p , q , etc. They represent sentences describing states of affairs or processes which may or may not obtain (go on) at a given time or on a given occasion. Such sentences are *open*, i.e. the propositions which they express are not "in themselves" true or false, but acquire a truth-value when associated with an individuating, spatial and/or temporal determination. Such propositions will also be called *generic*.
- (b) Sentential connectives $\sim$, $\&$, $\vee$, $\rightarrow$, $\leftrightarrow$ for forming molecular compounds of sentences.
- (c) The deontic operators, O and P , symbolizing the deontic status of obligatoriness and permittedness respectively.
- (d) Brackets ().

⁵ *Ibid.*, especially Ch. VII, pp. 107–28.

⁶ The term "standard system" seems to have been coined by Bengt Hansson (see above, p. 103). I here use it in a somewhat more restricted sense than the one proposed by Hansson in that "mixed formulas" are not regarded as norm-formulations.

The well-formed formulas of the calculus are schematic norm-formulations. They are defined as follows:

- (i) A variable, or molecular compound of variables, preceded by a deontic operator, is an atomic norm-formulation of the first order.
- (ii) A molecular compound of norm-formulations is a norm-formulation.
- (iii) A deontic operator followed by a norm-formulation or by a molecular compound of variables and norm-formulations is a(n atomic) norm-formulation of higher order.

Rules for the use of brackets need not be stated here; they should be obvious from the contexts of formulas appearing below.

Examples. $O \sim p$ is an atomic norm-formulation. The formula should be read "it ought to be the case that it is not the case that p " or, shorter, "it ought to be that not p ". $Pp \vee P \sim p$ is a molecular norm-formulation (of the first order). POp is an atomic norm-formulation of the second order. $POp \vee Oq$ is a norm-formulation. $p \rightarrow Oq$ is *not* by our definition a norm-formulation—but $O(p \rightarrow Oq)$ is a norm-formulation.

Expressions occurring within the scope of a deontic operator will also be referred to as (expressions for) *norm-contents*.

IV

A problem which calls for special attention has to do with the application of sentential connectives to norm-formulations.⁷

Consider first negation. How shall we understand, for example, $\sim Op$? "It is not the case that it ought to be the case that p ." This form of words could be used for stating that there *is* not (does not exist) a norm to the effect that it ought to be the case that p . But then the sentence expresses a norm-proposition. It is descriptive, not prescriptive. Has the sentence no prescriptive use then? I think it cannot be rightly denied that it has one. But what does the sentence *mean* when used prescriptively? As far as I can see, the only plausible meaning which can then be associated with it is that of the sentence "it is permitted that it is not the case that p ". But for that we also have the symbolic expression $P \sim p$.

Similarly, the form of words "it is not permitted that it is the case that p " or, which says the same, "it is not the case that it may be the case that p " can be understood either descriptively or prescriptively. Descriptively understood, the words mean that there is no permission to the effect that it may be the case that p . No such permission has been given, issued, or, if once given, has later been withdrawn or cancelled ("derogated"). Prescriptively understood, the words state a prohibition, mean that it must not be

⁷ This problem was mentioned, and discussed rather extensively, in *Norm and Action*, pp. 130ff.

the case that p , i.e. ought to be the case that not- p . For this last we also have the symbolic expression $O \sim p$.

I shall, following the terminology adopted in *Norm and Action*, say that Op and $P \sim p$ and, similarly, Pp and $O \sim p$, when used prescriptively, express the *negation-norms* of one another.

That "and" has a genuine prescriptive use for joining norm-formulations seems beyond doubt. "It ought to be the case that p and may be the case that q " could be a genuine prescription.

The notion of disjunction, i.e. the particle "or", is more problematic. The same problem attaches to the word "not" when prefixed to a string of norm-formulations joined by "and".

"It ought to be the case that p or it may be the case that q " can, of course, be used for stating that either there is (exists) a certain obligation or a certain permission. But could the same form of words be used for giving a prescription? The question is confusing. If someone is told that it ought to be the case that p or may be the case that q he would not be told anything definite about the normative (deontic) status of those two states of affairs. But if he himself is in a position to *make* things obligatory or permitted, then being told this might be an *injunction* (command, norm) to him either to make a certain thing obligatory or a certain other thing permitted. This would surely be a genuinely prescriptive use of the form of words in question. A superior norm-authority might use them in addressing a subordinate norm-authority.

Similar observations apply to the form of words "It is not the case that it ought to be the case that p and may be the case that q " or, in symbols, $\sim (Op \& Pq)$. Accepting the usual transformations of expressions formed by means of sentential connectives, this last formula could also be written $\sim Op \vee \sim Pq$. And accepting what was just said about negation and norms, this again could be transformed into $P \sim p \vee O \sim q$. "It may be the case that not p or ought to be the case that not q ". A prescription thus worded would, I think, most naturally be understood as an injunction to someone either to permit something or to make obligatory something else.

It thus seems that disjunctions of norm-formulations and negations of conjunctions of norm-formulations can be prescriptively understood. On what seems to me to be the most natural way of thus understanding them, the molecular constructs in question do not prescribe the deontic status of the states of affairs themselves involved in the atomic components, but prescribe which deontic status some states of affairs ought to or may be given.

Consider an arbitrary formula F of our calculus. By the *normative negation* (-formula) of F I shall understand the formula which is obtained from $\sim F$ by pushing the negation-sign $\sim$ as far "inside" as allowed by the rules of propositional logic (PL) and making the shift from $\sim O$ to $P \sim$ and from $\sim P$ to $O \sim$.

Examples. The normative negation(-formula) of $Pp \vee O \sim q$ is $O \sim p \& Pq$, that of POp is $OP \sim p$, and that of $O(p \rightarrow Oq)$ is $P(p \& P \sim q)$.

V

Norms of higher order have been something of a neglected problem-child of traditional deontic logic. In a system of deontic logic in which the "contents" of norms are thought of as actions, norm-formulations of higher order are not even well-formed. This is the case, for example with my first or 1951 system.

In a system of deontic logic in which the contents are thought of as states of affairs, expressions of higher order may be regarded as well-formed. But their interpretation is open to question.

Consider for example the expression POp or "it may be (the case) that it ought to be (the case) that p ". If we stick to the view, taken here, that the content of a norm is a state of affairs then not only the variable " p " in the expression but also the compound " Op " must represent a state of affairs. The state of affairs represented by " Op " is that a certain thing (state of affairs) is obligatory. " Op " is thus the schematic form of an expression to the effect that a certain norm exists. In other words, " Op " must be interpreted descriptively. As norm-formulation it expresses a norm-proposition and not a norm.

The higher-order expression POp , however, may be interpreted either descriptively or prescriptively. Descriptively interpreted it states that there is permission to the effect that it may be the case that it is the case that it is obligatory that p . In other words it says that the state of affairs which obtains when there is a norm which makes it obligatory that p is itself a permitted state. Prescriptively interpreted the expression permits, i.e. gives permission to create, this state. It might, for example, be permission which a higher authority gives to a subordinate one to make certain things obligatory under norm.

As seen, norm-formulations of higher order can be given a prescriptive interpretation. But any norm-formulation which occurs itself within the scope of a deontic operator has to be interpreted descriptively.⁸ This necessity of combining a prescriptive with a descriptive interpretation of expressions occurring in one and the same formula may strike one as surprising. But once the necessity is clearly seen it ought not to strike us as problematic.

In view of the fact that a norm-formulation which is part of the descrip-

⁸ This observation has also been made independently by C. Alchourrón and E. Bulygin in an essay "von Wright on Deontic Logic and the Philosophy of Law", in a volume *The Philosophy of Georg Henrik von Wright* for the series Library of Living Philosophers.

tion of the content of a higher-order norm has to be interpreted descriptively one can raise the question why, in the formation of negation-norms of a given (higher-order) norm, the shifts from $\sim O$ to $P \sim$ and from $\sim P$ to $O \sim$ should be allowed (cf. above, p. 134). Why should we say, for example, that the negation-norm of POp is $OP \sim p$ rather than saying that it is $O \sim Pp$? I shall try to answer this question later when discussing what I propose to call the Transmission-of-Will Principle for norm-giving activity (see below, pp. 147 and 163).

VI

The status of permissions is traditionally a much debated question in the philosophy of norms, and particularly in legal philosophy. Is permission an independent category, or can permission be defined in terms of obligation? That obligation and prohibition are interdefinable has seldom, if ever, been doubted. The question of the independence of permission is usually raised in the following form: Is permission anything "over and above" the absence of prohibition?

It should be noted in passing that if permission can be defined in terms of prohibition (obligation), then also the reverse is true. The above question could therefore also be put: Is prohibition anything "over and above" absence of permission? It is worth reflecting why the question is seldom, if ever, put in that way. Prohibitions and obligations are somehow ontologically more "basic" or "real", it would seem, than permissions. This presumably is connected with the fact that neglecting obligations and breaking prohibitions is normally connected with "sanctions" of one form or another such as legal punishment or moral reprobation.

In legal philosophy, the question of the status of permissions is related to the problem of "gaps" in the law. If things which are not prohibited are *ipso facto* permitted, then any action or state has a legal status: it either is prohibited (and its "opposite" obligatory) or else it is permitted. The legal order is of necessity in this sense *closed*. It has no "gaps". A forceful proponent of this opinion was Hans Kelsen. Other legal philosophers, perhaps even the majority, do not find this position satisfying. They would argue that beside things which are permitted and prohibited there are also a good many things which are not regulated by norm at all and therefore lack deontic or legal status. They are not prohibited, but they are not (thereby) permitted either.

I think that this dispute can be settled once and for all by strictly observing the distinction between descriptive and prescriptive discourse.

"Absence" is a descriptive term. "Absence of prohibition" means that there *is* no prohibiting norm, that no norm to the effect that a certain thing is prohibited *exists*. To maintain that absence of prohibition is tantamount

to a permission is to make a statement at the level of *norm-propositions*; and I do not think that this statement could be proved true by logical argument. The question of "gaps" in the law is not a question of logic. It is rather a question of accepting or not accepting certain "meta-norms".

One could, for example, argue that it is part of the idea of a *Rechtsstaat* or "reign of law" in a society that only deeds which are prohibited by an already existing law may be punished as criminal offences. *Nulla poena sine lege*, or *nullum crimen sine lege*, as the sayings go. In a society where this meta-norm is actually observed anyone can do what is not forbidden without having to fear sanction. Then one could also say that in this society everything which is not legally forbidden is legally permissible. But this argument concerns an "ideal" legal order and is not an argument in logic.

There are, moreover, systems of normative relationship in which the principle that anything which is not forbidden is permitted may appear unreasonable. The relations between parents and their children (before a certain age) might serve as example. There are things which have never been prohibited to the child—but for the doing of which the child may yet be reprimanded or punished. It must not do these things without first, as we say, having *asked permission*. The parents will then consider which "deontic status" these actions should have.

At the prescriptive level things are different. If in saying "this is not prohibited" I *give a prescription* and do not just speak about what has and what has not been subjected to norm, then surely what I do is that I permit this thing to be done or to exist. And vice-versa, if I say "this is not permitted" and "mean" this as a *prescription*, then surely I thereby prohibit a certain thing.

Thus in prescriptive language "not permitted" means "forbidden", and "not forbidden" means "permitted". Similarly, "not obligatory that" means "permitted that not" and "not permitted that not" means "obligatory that". In the prescriptive language of norms we can therefore safely assume " P " = " $\sim O \sim$ " or " O " = " $\sim P \sim$ " and, if we wish, dispense altogether with one of the two operators. This is also what is usually done in known systems of deontic logic. But not always for the right reasons. And in a logic of norm-propositions, i.e. of descriptively interpreted norm-formulations, we have no right to assume the above identities and to dispense with one of the operators. Here both operators are needed.

VII

Most norms come into existence and pass out of existence in the course of time. They may come into existence through an act of issuing or promulgation by some "norm-authority", and pass out of existence through an act of cancellation or derogation. This is normally the case with legal norms;

but it need not always be the case. Some legal norms pass out of existence by what is called *desuetudo*, and some come into existence from customs which in due course acquire the force of legal obligations or rights.

The existential status of moral norms is more complicated.⁹ Moral codes and various moral taboos have a connection with custom and with religious beliefs and teaching. But rules such as that promises ought to be kept, or that it is immoral to tell a lie, seem rather different. I do not know what to think about their status in time. Their existence cannot truly be said to depend on historical contingencies such as the enacting of a law or even the gradual formation of a custom. They are rather meant to be valid *semper et ubique*, like the laws of nature. On the other hand, the rule that promises ought to be kept presupposes the existence of the *institution* of promising. This institution need not be universal. It may exist in some societies but not in others. Something similar holds true of moral rules condemning theft. The institutions themselves of contract and property are, however, also legal institutions. Custom, law, and morality are thus inseparably interwoven.

Norms with a history I shall call *positive* norms. The clearest example of positive norms is statutory law. I shall not here inquire whether there are norms of an essentially unhistorical nature and whether the logical study of such norms would have peculiarities which distinguishes it from a "normal" deontic logic.

VIII

A norm to the effect that it ought to be the case that *p* is *satisfied*, I shall say, if, and only if, at *all* times in the history of this norm, it is the case that *p*. And it is *dissatisfied* if, and only if, at *some* time in its history it is *not* the case that *p*. For norms enjoining or prohibiting *actions* the situation is a little more complex. The norm to the effect that *X* ought to be done is satisfied, I shall say, if, and only if, *X* is done on *all* occasions when there is an opportunity for doing it—and dissatisfied if, and only if, *X* is not done on *some* occasion when there is an opportunity (cf. below, p. 197).

Whenever *X* is done by a norm-subject, the subject obeys the norm; whenever *X* is not done by a subject when there is an opportunity for him to do *X*, the norm is broken or disobeyed. Satisfaction, as here defined, thus means exceptionless obedience to the norm.

If a lawgiver prohibits something or makes it obligatory, he will have to count with the possibility that the norm will not always be observed. But

⁹ Cf. above, pp. 75f.

he may be said normally to want or desire, to "will", that what he has enjoined should without exception be the case. He wants the norms to be satisfied. If for some reason or other it would be *impossible* that the obligatory states always (in the history of the norm) obtain, we might say that his wish (will) is not "rational", since it *cannot* be fulfilled. The irrationality would be particularly glaring, if an obligatory state could never obtain and a prohibited state consequently always had to obtain. This would be the case, for example, if the lawgiver had enjoined a contradiction, $p \ \& \sim p$, to be the case.

If a lawgiver had made obligatory both of two mutually contradictory states, then one of the obligations may be satisfied but only at the expense of the other necessarily remaining dissatisfied. If one of the two states obtains at some time, and the other at some other time, neither obligation is satisfied. Since it is impossible that both be satisfied, it was surely irrational, foolish, on the part of the lawgiver, to enact both norms. Doing this may also diminish his prestige in the eyes of his subjects.

But could he not issue contradictory orders merely with a view to placing the subject or subjects in an "unsolvable predicament"?¹⁰ So that he can punish them whatever they do. Surely this is possible. Must it be "irrational"? Not if the purpose is to perplex or punish people. It is irrational only relative to the *normal* purpose which the institution of giving orders serves; *viz.* to achieve the states of affairs consequent upon fulfilment of the orders. Perhaps one should in the imagined case call the order "perverse" rather than "irrational".

Permissions are in a somewhat different position. Normally, a permitted state of affairs is not also obligatory, but its contradictory is permitted too. A lawgiver may perhaps hope or wish that his subjects never avail themselves of a permission he has granted, *i.e.* hope that a permitted state of affairs should never obtain. (Perhaps he was forced to grant the permission under duress.) But if it were quite impossible for this state to obtain, its permittedness would only be a "joke". It would be a "mock-permission". It is therefore a thoroughly reasonable and rational demand—on the part of the norm-subjects if not of the norm-authority—that permitted states of affairs can sometimes obtain, although the demand need not be that they can obtain always. If a state and its contradictory are both permitted, they cannot both *always* obtain—yet there is nothing irrational about both being allowed to obtain.

I shall say that a permissive norm is satisfiable if, and only if, it is possible that the permitted state of affairs obtains at *some* time in the history of the norm. And it is satisfied if, and only if, at some time in its history that which it permits actually is also the case.

¹⁰ I owe this example to Professor Elizabeth Anscombe.

A norm (obligation or permission) which is satisfiable will also be said to be *normatively (self-)consistent*.

IX

For a finite set of co-existing norms I shall introduce the technical term a *corpus*.¹¹

If the norms of a corpus are "historical", their histories may vary in length. By *the history of the corpus* I shall understand the overlapping portion of the histories of the norms which are members of the corpus. A long-lasting norm may thus be a member of a good many different corpora. The totality of norms which make up a legal order or system is normally a succession (family) of sets which are here called corpora.¹²

Since a corpus is a finite set of norms we can express its total content in the form of a conjunction of all the norm-formulations in the corpus. We thus form of the norm-formulations *one* (complex) norm-formulation.

A corpus of norms is satisfiable if, and only if, it is possible that all states which the norms of the corpus make obligatory obtain throughout the history of the corpus and all states which the norms permit obtain some time in this history.

A satisfiable corpus will be called *normatively consistent*; a corpus which is not satisfiable again normatively inconsistent.

I shall, furthermore, regard satisfiability as a criterion of *rationality* of a corpus of norms. A lawgiver may be said to envisage an ideal state which agrees with his wish or will, and in which everything obligatory is the case as long as the obligations exist and everything permitted is the case at some time or other in the history of the permissions. The lawgiver can be said to "will" that all obligations be satisfied or *fulfilled* and to "tolerate" that all permissions be satisfied or *used* at some time or other. For the sake of simplicity I shall in the sequel speak of his will or toleration as simply his "will".

If a norm, or corpus, is not satisfiable (normatively consistent) the lawgiver or supreme (sovereign) norm-authority cannot *rationally will* the ideal state envisaged in his norm-giving activity or legislation.

The condition of rationality which I have stated may be regarded as a necessary minimum condition. Further conditions may be laid down, for example that the obtaining of the various states which the norms make obligatory or permitted should also be physically possible, or humanly

¹¹ *Norm and Action*, p. 151.

¹² The conception of normative systems as a temporal succession of sets of norms has been emphasized and elaborated by Alchourrón and Bulgin. Cf. their book *Sobre la existencia de las normas jurídicas*, Oficina latinoamericana de investigaciones jurídicas y sociales, Valencia (Venezuela), 1979, Ch. XI.

possible, to achieve through action. But we shall not here discuss such additional requirements. Perhaps one could say that the condition as laid down above is the only one which is of relevance to a "logic" of norms.

Should it emerge, either from "theoretical deliberations" or from "legal practice", that a corpus does not satisfy the minimum condition of rationality, the lawgiver is likely to take various steps in order to "restore" rationality. He may, for example, *derogate* some norms, remove them from the corpus, or he may restrict their scope through some process of *conditionalization*. Both derogation and conditionalization of norms are interesting from a logical point of view. Derogation will not be further discussed in this essay;¹³ but on the logical form of conditional norms remarks will be made in Part 2.

X

If the negation norm of a given norm is *not* satisfiable the norm itself will be called (*normatively*) *tautologous* or said to be a tautologous norm.

Consider, for example, the norm $O(p \vee \sim p)$. Its negation norm is $P(p \& \sim p)$. This is not satisfiable. At no time in the history of this "permission" could, for reasons of logic, its content be realized. Giving such a permission is not, on our adopted standard, rational.

That $O(p \vee \sim p)$ is a normative tautology does not *mean* that it is logically true, "a truth of deontic logic", that it is obligatory that p or not p . What it means is, simply, that one cannot "rationally will" the normative negation of $O(p \vee \sim p)$. And whether such willing is or is not possible depends upon what standards of rationality one adopts for norm-giving activity. Our adopted standards are not the only possible ones—and perhaps in *some* normative contexts not even the most reasonable ones.¹⁴

Consider next a corpus consisting of (or containing) the two norms Op and $O \sim p$. This corpus cannot be satisfied. The conjunction $Op \& O \sim p$ is normatively inconsistent. The normative negation of it is $Pp \vee P \sim p$. We shall call this too a normative tautology. But again it must be remembered that the tautological character of $Pp \vee P \sim p$ does not mean that it is "a truth of deontic logic" that either any given state of affairs is itself permitted or its negation is permitted. What it means is, strictly, that one cannot—on the adopted standards—"rationally will" that a certain state be and that it, at the same time, not be.

Let us assume that Op is a member of a normatively consistent corpus

¹³ It is a merit of Alchourrón and Rulygin to have noted the peculiar logical problems connected with the concept of derogation (see above, pp. 137ff).

¹⁴ These observations should remove the qualms which I have long felt about accepting, as a "truth" of a logic of norms, that tautologous states of affairs are "obligatory" (see above, p. 105).

of norms. If to this corpus we add $O \sim p$ it becomes normatively inconsistent. This we may regard as a reason for saying that the negation norm of $O \sim p$ which is Pp is *implicit* in the corpus, or that the conjunction of norms in the corpus *normatively entails* the norm Pp .

Op may be the only norm in the corpus. Then, by the above, Op entails Pp —assuming that Op itself is normatively (self)-consistent.

Op & $O \sim p$ is the normative negation of $Op \rightarrow Pp$. The fact that Op & $O \sim p$ is normatively inconsistent is our ground for calling $Op \rightarrow Pp$ a normative tautology. Does this tautology “mean” or “show” that an obligation *entails* a corresponding permission or that it is a truth of deontic logic that what is obligatory is also permitted? Saying this would be obscuring, not to say false. The right thing to say is this: If something has been ordered, it would be irrational also to forbid it. Therefore, if one has made something obligatory, one is “tacitly committed” to permit this thing too. But this commitment is not “logical”. It is a commitment which one has only in so far as one aspires to be rational.

We could lay down the following *Rule of Rational Commitment* as I shall call it: *If the conjunction corresponding to a given corpus of norms is normatively consistent as it stands but becomes inconsistent when a new norm is added to the corpus, then the negation-norm of the added norm was “implicit” in the corpus.*

XI

We can apply this rule for obtaining equivalents of many of the well-known theorems of deontic logic:

(i) $O(p \& q) \& P \sim p$ is normatively inconsistent. Hence a corpus of norms containing $O(p \& q)$ also contains implicitly, in the sense of our Rule of Commitment, the negation-norm of $P \sim p$, i.e. Op —and, of course, Oq too.

(ii) $Op \& Oq \& P(\sim p \vee \sim q)$ is likewise normatively inconsistent. Hence, if a corpus contains the two norms Op and Oq it also implicitly contains $O(p \& q)$.

(i) and (ii) jointly constitute the formula $O(p \& q) \leftrightarrow Op \& Oq$ of traditional deontic logic. This formula is thus a normative tautology.¹⁵

(iii) $O(p \vee q) \& O \sim p \& O \sim q$ is another normative inconsistency. Thus a corpus is irrational in which there is a norm which makes it obligatory that the one or the other of two states of affairs obtain, but also norms prohibiting the obtaining of both states individually. If the lawgiver

¹⁵ There is no conflict between this and the non-validity of the “corresponding” equivalence formula in more finely structured systems of deontic logic (cf. above, pp. 125f, and below, p. 190).

wishes to retain the two prohibitions he is, in the name of rationality, committed to permit the obtaining of the conjunction of the negation of both the prohibited states too ($P(p \& q)$). If again he is keen on the disjunctive obligation he will have to cancel at least one of the two prohibitions.

This last may be said to be a reflection of the deontic formula $Op \& O(p \rightarrow q) \rightarrow Pq$, sometimes rendered in words by saying that what is obligatory can commit only to something which is permitted. Since, in fact, also $O(p \vee q) \& O \sim p \& P \sim q$ is inconsistent, we can strengthen this to a principle which is sometimes expressed in words by saying that the obligatory commits only to things which are themselves obligatory. This is reflected in the formula $Op \& O(p \rightarrow q) \rightarrow Oq$ of deontic logic.

(iv) $P(p \vee q) \& O \sim p \& O \sim q$ is a normative inconsistency. What deontic commitments would follow from this?

One commitment is obvious: an authority which prohibits both of two states cannot rationally allow that one of them sometimes obtains. But assume that he is keen on allowing that either of them sometimes obtains. Then he cannot rationally prohibit them both. He must allow at least one of them. But which one? This is open to his choice. Shall we therefore say that, if a corpus contains a disjunctive permission, it implicitly contains also a disjunction of two permissions? This sounds strange. A corpus contains norms. Is a disjunction of norms a norm? The question can also be put as follows: Has "or" a *prescriptive* use for joining norm-formulations? We have already raised the questions once (above, p. 134). I feel tempted to answer No to the first and Yes to the second—but I think we should be liberal here. The agent who is being told that it is permitted that p or permitted that q may feel at a loss and complain that he has not been given any rule for his acting (for what he may do). The norm-authority, on the other hand, could say to himself "I cannot rationally prohibit both states, so I must allow at least one of them"—and this soliloquy could perhaps be called "prescriptive".

When we abandon talk about norms "entailing" each other, the puzzle dissolves in a natural manner. The situation with regard to the "normative tautology" $P(p \vee q) \leftrightarrow Pp \vee Pq$ is simply this: if the norm-authority permits a disjunctive state of affairs, he is "rationally committed" to permit at least one of the disjuncts. If he prohibits both, his norm-giving activity is "irrational". And this is all we have to say.

It is worth making here the following observation on the *prescriptive* use of *permissive* norm-formulations: A prescription permitting a disjunctive state of affairs to obtain is normally understood to permit every one of the disjuncts individually. This does not make $P(p \vee q) \rightarrow Pp \& Pq$ a "normative tautology" as defined by us. But it may be taken to suggest that one should exclude, as being "incorrect" or "ungrammatical", a combined use of a prescriptive "it is permitted" and a prescriptive "it is forbidden" for permitting a disjunction of some states and, at the same time, prohibiting one or several of the disjuncts. "It is permitted that p or

q but forbidden that p " may be a true statement about the deontic status of some states of affairs. But as a prescription this would either be unintelligible or understood as an "odd" way of permitting that q while prohibiting that p . (Cf. above, pp. 104f and p. 124 on the notion of "free choice permission".)

XII

There have been endless discussions among deontic logicians of some "paradoxes". In elementary deontic logic they are essentially variations of one and the same, viz. the so-called Ross Paradox. It arises, as well known, from the fact that the norm Op is alleged to "entail" the norm $O(p \vee q)$.

Let us see how this is reflected at the level of rational norm-giving activity. $Op \ \& \ P(\sim p \ \& \ \sim q)$ is normatively inconsistent. Hence, if a corpus contains a norm to the effect that it ought to be the case that p , it cannot "rationally" also contain the negation-norm of the norm that it ought to be the case that p or q . The negation-norm would permit that neither of these two states obtains and *a fortiori* that the first of them does not obtain either. Hence what the Ross Paradox amounts to is simply that a lawgiver cannot rationally *both* make a state obligatory *and* at the same time permit that this state, together with some other arbitrary state, does not obtain. This is obvious and not in the least "paradoxical". It appears paradoxical only when one speaks about it as an "entailment" between a norm Op and a norm $O(p \vee q)$. There *is* no such thing—and the Ross Paradox makes us aware of the absurdity of talking as though it existed. Let it be said here in tribute to the memory of the eminent Danish jurist that Alf Ross was right when he offered his paradox as a proof that, properly understood, there could not be such a thing as a logic of norms. But then one must also lay special emphasis on the words "properly understood".

XIII

When considering norms of the first order we need not pay attention to possible restrictions on the time during which the norms exist. It suffices to assume that the norms of a corpus are co-existent. The norms may even be "eternal" ("ahistorical"), existing "from the dawn of creation" to "the end of time".

When we proceed to consider norms of higher order the situation is different.

Consider, for example, the expression POp , "it may be the case that it ought to be the case that p ". When would this form of words be used? Perhaps for saying that a superior legal (or moral) authority permits a subordinate norm-giving authority to issue a norm to the effect that a

certain thing ought to be the case. This would be a descriptive use for stating the norm-proposition that a certain permissive norm has been issued, exists. This norm permits the state of affairs that a certain thing is obligatory. Whether this state exists or not depends upon whether the subordinate authority has availed himself of the permission or not.

Assume that the sub-authority avails himself of the permission. This he need not do "at once", as soon as the permission is given. He may do it when the permission has already existed, been "in force", some time. The permission may then be withdrawn or otherwise cease to exist. Does it follow that the obligation too then ceases? Not necessarily. It may have acquired an existence "independent" of the permission. But this part of its life, if there is such a part, will not interest us here.

It may also happen that the sub-authority avails himself of the permission and makes a certain thing obligatory (for some norm-subjects) but then withdraws the norm while the permission to issue it is still in force. He may do so repeatedly. This is not at all unrealistic. A sub-authority may, for example, be permitted (entitled, have a right) to order a curfew in the case of emergency. From time to time he may avail himself of this right.

The satisfaction of a given norm may thus require the coming into existence of other (subordinate) norms. The norms thus coming into existence may outlast the life-span (history) of the superior norm, but such parts of their life are immaterial to the satisfaction of the superior norm.

Let the norm be *OOp*. If this norm is satisfied then, throughout its duration, there exists a norm—issued maybe by a sub-authority—making it obligatory that *p*. Assume that this is the case but that the state of affairs that *p* does not obtain throughout the duration of the norm or norms making it obligatory. Then some norm of lower order is *not* satisfied. Shall we say that the higher-order norm *is* satisfied, or not?

The answer is not immediately clear. We could answer the question by Yes or by No. If we give the first answer then the higher authority "does not care", one could say, whether what the lower authority makes obligatory is or is not. His "will" is only that there always be such an obligating norm. But why should he "will" this unless he is anxious that the state of affairs itself, which is made obligatory, obtains? At least in normal cases, an authority who orders that something be made obligatory wants the obligation satisfied. He, as it were, "transmits" his will through the intermediary of a lower authority. Therefore his will is not fulfilled unless the norms which are its immediate objects are themselves satisfied.

If we do not add (some) such Principle of Transmission of Will, as I shall call it, to the notion of rational willing (norm-giving) the theory of higher-order norms becomes "uninteresting". That is: it would not then contain any deontic tautologies over and above those already familiar from first-order deontic logic.

Accepting this principle, an obligation norm of higher order is satisfied if, and only if, the state of affairs which it makes obligatory exists throughout its history *and* all norms the existence of which is a constituent part of this state of affairs are themselves satisfied.

If instead of "throughout its history" we say "some time in the course of its history" we obtain the criterion of satisfaction for permissive norms of higher order.

With this extension of the notions of satisfaction and satisfiability, or of our criterion of rational norm-giving, one can also apply to norm-formulations of higher order and their molecular compounds our previously defined notions of normative consistency, tautology, and entailment.

Consider $Op \rightarrow OOp$. Its normative negation is $Op \& PP \sim p$. Why cannot an authority rationally issue two norms, one ordering it to be the case that p and another permitting it to be permitted that it is not the case that p ? The answer is not obvious. The two norms rationally conflict only if we assume that an authority who permits (to some sub-authority to give) a permission that it be not the case that p is bound in consistency to tolerate that it is not the case that p . Only if we assume this is there a conflict with the norm, issued by that authority, that it ought to be the case that p . But it is at least highly natural, even if not compelling, to accept this as a criterion of rational norm-giving.

Assume that Op is normatively consistent. Since adding $PP \sim p$ to a consistent corpus containing Op would then make the corpus inconsistent, we can also say that the higher-order norm OOp is implicit in this corpus. But no such norm need have been given. Its "implicit existence" consists only in the fact that issuing its negation-norm would be an act of "irrational legislation". The legislator could not rationally permit anybody (himself included) to permit that it be the case that $\sim p$ since he has made it obligatory that p .

This is the sense in which $Op \rightarrow OOp$ may be regarded as a "law of deontic logic".

A formula which has been much discussed as a plausible candidate for a higher-order truth of "deontic logic" is the formula $O(Op \rightarrow p)$.¹⁶ "It ought to be the case that if it ought to be the case that p then it is the case that p ." This sounds intuitively plausible. That which ought to be the case is, as everyone knows, not necessarily the case—but it ought to be the case, since it is obligatory. The formula $O(Op \rightarrow p)$ is indeed a normative tautology. This means that its negation norm $P(\sim p \& Op)$, if ever given or issued, could not possibly be satisfied. A lawgiver who enacted such a norm would surely undermine his own authority. He could also be said to

¹⁶ Introduced into the literature by Arthur N. Prior in his book *Formal Logic*, 2nd edn, Oxford University Press, London, 1962.

be irrational, or not fully to understand what obligations and permissions are.

Assume that a lawgiver actually ordained that it ought to be the case that that which (according to his will) ought to be the case also is the case. What would we think of such a "law"? Perhaps not that to enact it is "irrational"—but maybe a little ridiculous. Perhaps its enactment would be a way of emphasizing that the lawgiver wants his (other) laws to be taken seriously. Like his saying, with emphasis, "if something *ought to be* the case it ought to *be the case*". Be it observed in passing, that ordaining $O(Op \rightarrow p)$ does not "entail" that Op had been ordained. As regards the state of affairs that p the norm is conditional. It only says that should this state be made obligatory, now or in some future time, then it ought to obtain.

The question may be raised how we should understand the idea of the sovereign authority "transmitting his will" to sub-authorities if we had opted for a weaker conception of the notion of a "negation-norm" not allowing the shift from $\sim O$ to $P \sim$ or from $\sim P$ to $O \sim$ within the scope of a higher-order deontic operator (cf. above, p. 136). Assume, for example, that the normative negation of OOp were $O \sim Pp$ (and not $OO \sim p$). The sovereign wants that the sub-authority does *not* permit the state that p . Does he then want this state to be prohibited, i.e. does he want the norm-subjects to behave in such a way that the state that p does not obtain? Possibly this is what he wants, but not necessarily, it seems. But he may be said to want to "reserve room", logically speaking, for such a prohibition. If the prohibition were issued it must also be possible that it is observed, i.e. it must be (logically) possible that the state of affairs that $\sim p$ never obtains within the existence-span of this prohibition. But this is precisely what ought also to be possible if he had wanted a sub-authority to issue the prohibition. Hence from the point of view of a Transmission-of-Will Principle the two norms $O \sim Pp$ and $OO \sim p$ have the same conditions of satisfiability. Whether one could define satisfiability in such a way that the difference between the two norms becomes perceptible (at the level of satisfiability) is a question into which I shall not here inquire. Be it observed, however, that I do not claim that the two norms are "identical". They obviously are not.

XIV

Under the name "deontic logic" one may understand an axiomatized logical calculus the theorems of which are all the normative tautologies. What counts as a normative tautology, however, depends upon how the notion of satisfiability of norms is defined—and this again depends upon our choice of criteria for rational norm-giving and/or rationality in norms.

The system of deontic logic for norms of the first order has the following axioms:

- A0. All tautologies of PL when for the variables are substituted norm-formulations.
- A1. $P(p \vee q) \leftrightarrow Pp \vee Pq$
- A2. $P(p \vee \sim p)$
- A3. $O(p \vee \sim p)$
- A4. $Op \leftrightarrow \sim P \sim p$.

The rules of inference are:

- R1. For a variable may be substituted another variable or a molecular compound of variables.
- R2. The usual Rule of Detachment.
- R3. Formulas which are provably equivalent in PL are intersubstitutable in DL.

If we extend R1 by allowing substitution of formulas of DL for variables, and extend R3 to formulas provably equivalent in DL, we obtain a system of deontic logic under a restricted definition of the notion of satisfiability.

If we extend the notion of satisfiability in accordance with the "Transmission of Will"-Principle we may add to the system the axioms

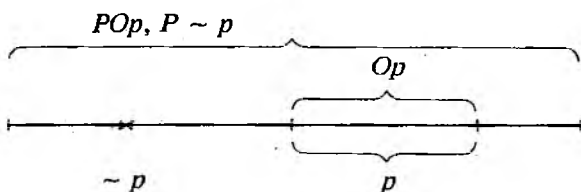
- A5. $Op \rightarrow OOp$

and

- A6. $O(Op \rightarrow p)$

The system of deontic logic thus obtained may be said to be S4-like.

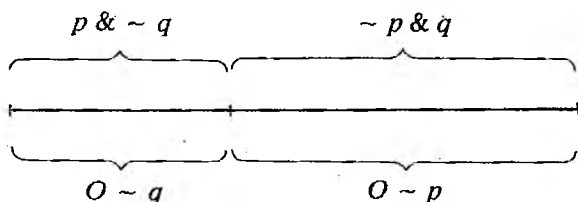
If the system admitted the principle $POp \rightarrow Op$ it would be S5-like. This formula says that if it is permitted that something is obligatory then this thing is obligatory. In our view of rational norm-giving this is not a deontic tautology. Its normative negation is $POp \& P \sim p$. The sovereign permits that it is not the case that p but also that a sub-authority forbids this. This corpus would be satisfied if, for some time, the sub-authority does not avail himself of his permission and during that time, or a part of that time, it is not the case that p . The following illustration shows that the negation of $POp \rightarrow Op$ is satisfiable:



The satisfiability of the corpus consisting of POp and $P \sim p$ essentially depends upon the fact that the sub-norm Op may have a shorter existence-span (history) than the corpus itself. If, however, norms were not historical

entities but had infinite duration (were eternal) then the sub-norm would have to be coexisting with the superior norms of the corpus. And then no occurrence of $\sim p$ could be allowed. The addition of $P \sim p$ to POp would result in an inconsistency and the principle $POp \rightarrow Op$ would be a "normative tautology". Hence we may conclude that the deontic logic of eternal norms is S5-like.

It is also of some interest to note that (our) deontic logic is not even S4.3-like. This is shown by the fact that $Pp \& Pq \rightarrow P(p \& q) \vee P(p \& Pq) \vee P(q \& Pp)$ is not a deontic tautology. Its normative negation is $Pp \& Pq \& O(\sim p \vee \sim q) \& O(\sim p \vee O \sim q) \& O(\sim q \vee O \sim p)$. The following illustration shows that this conjunction is satisfiable:



If things are as in the picture then, in the history of this corpus of five norms, it is sometimes the case that p and sometimes the case that q , always the case that it is either the case that not p or that not q (in other words, never the case that both states obtain), always the case that not q when it is the case that p and, finally, always the case that not p when it is the case that q . This means that the five norms are all satisfied.

XV

It is one of the traditional views of the nature of norms that norms are expressions of a norm-giving authority's will. What is said here about the laws of deontic logic as "principles of rational norm-giving" must not be taken as an endorsement of the said view. It is nevertheless a useful fiction, it seems to me, to think of a set of norms belonging to the same code, or normative order, or system as emanating from a "will" that certain things be the case and certain other things tolerated. It is natural to make it a criterion of the *rationality* of such a "will" that the "willed" total state of affairs be (at least) logically possible. On the basis of this criterion one can then determine the analogical meanings of contradiction and entailment also for norms, although norms have no truth-value.

If there existed no source, *like* a will, for the unity of a set of norms there would be no reason, it seems, why the contents of norms should be mutually consistent rather than contradict each other.

One could, however, drop the fiction of a "will" behind the norms and speak *only* of the *ideal* state of things envisaged in a normative order. It is

a natural requirement that an ideal should picture a logically possible state of affairs. Ideals which do not satisfy this requirement are like crying for the moon; to entertain such ideals is "irrational" just as it is "irrational" to demand (to "will") that the world should live up to them.

2. CONDITIONAL NORMS AND HYPOTHETICAL IMPERATIVES

I

A good many, perhaps a majority, of (genuine) norms which actually are issued are *conditional* or *hypothetical*. The standard way of formulating them in language is to say that *if* something or other is the case (or is done), *then* something else ought (may) also (not to) be (or be done).

As we shall see this standard way of formulating conditional norms is not altogether happy and can easily lead to (serious) confusion.

Let us ask: How shall conditional norms be adequately expressed in the symbolic language of a "deontic logic"? The question has aroused much discussion and many different answers have been suggested.

In my first paper on deontic logic (1951) I did not discuss explicitly conditional norms but made some remarks on the related notion of (a) *commitment*. I used for this the symbol $O(p \rightarrow q)$. It may be read "it ought to be the case that, if p , then q ", or, more accurately, "it ought to be the case that either not- p or q ".

Against this it has been objected that the suggested formalization runs into the difficulties caused by the so-called Paradoxes of Implication.¹⁷ If it is prohibited that it be the case that p , then *if*—notwithstanding the prohibition—it *is* the case that p , there is a "commitment" that it be the case that q too. This is thought to be so because in a deontic logic of the traditional type $O \sim p$ "entails" $O(p \rightarrow q)$. Similarly, if it is obligatory that q then whatever is the case "commits" to its being the case that q . Oq too "entails" $O(p \rightarrow q)$.

As easily seen, these "paradoxes" are but variations of the Ross Paradox. By substituting " $\sim p$ " for " p " and replacing " $p \rightarrow q$ " by " $\sim p \vee q$ " in $O \sim p \rightarrow O(p \rightarrow q)$ and cancelling double negations we obtain the Ross Paradox formula $Op \rightarrow O(p \vee q)$. Similarly, $Oq \rightarrow O(p \rightarrow q)$ is but an equivalent form of $Op \rightarrow O(p \vee q)$.

It was, among other things, in order to avoid such "paradoxes" that I set out to develop a theory of *dyadic* modalities in the early 1950s.¹⁸ The

¹⁷ The first to point out this was A. N. Prior in a note "The Paradoxes of Derived Obligation" in *Mind*, 63, 1954.

¹⁸ First in a congress report "A New System of Modal Logic" in *Proceedings of the 11th International Congress of Philosophy*, Vol. 5, Amsterdam-Louvain, 1953; and then, in

idea was to let a dyadic functor $N(p/q)$ symbolize the *relative necessity* of one proposition given another. Similarly, one could have a symbol $M(p/q)$ for relative possibility, connected with the symbol for necessity through the (defining) equivalences $M(p/q) \leftrightarrow \sim N(\sim p/q)$ or $N(p/q) \leftrightarrow \sim M(\sim p/q)$.

In a note in *Mind*¹⁹ I made a first attempt to apply these ideas about dyadic modalities to deontic concepts and propositions. The attempts were continued, with variations in several directions, in my writings on the logic of norms through the 1960s. The fullest statements of them were in *Norm and Action* (1963) and in *An Essay in Deontic Logic and the General Theory of Action* (1968).

It is not clear to me yet what the interest and merits are of a theory of dyadic modalities in general and of dyadic deontic operators in particular. But as I see things now, the logical problems connected with the formulation of conditional norms can be satisfactorily coped with *without* resort to a dyadic theory. This I would regard as an advantage.

Consider the form $p \rightarrow Oq$. It has been one of the candidates for the logical form of a conditional norm. We shall then have to think of " p " as representing a closed sentence expressing a (true or false) individual proposition. " q " may represent a closed or open sentence. The expression Oq , we know, has two interpretations, viz. a prescriptive and a descriptive (existential) one. How do these interpretations "match" the form $p \rightarrow Oq$?

On the descriptive interpretation of Oq , the implication formula says that it is either the case that it is not the case that p or it is the case that (there is a norm to the effect that) it ought to be the case that q . So therefore, if it is the case that p , it is also the case that there is a norm to the said effect. This is a factual statement, true or false as the case may be. It would be wrong or at least highly misleading to call this the expression of a conditional norm.

On the interpretation of Oq as a prescription (norm), the expression $p \rightarrow Oq$ would consist of a descriptive and a prescriptive ingredient joined by a sentential connective. Does such an expression "make sense"? The question is not really easy to answer. It seems to me clear, however, that if the sentential connective in question is a *truth*-connective, then this "linguistic hybrid" is a monster with no place in meaningful discourse. Since I wish to stick to the truth-functional understanding of the implication sign (" $\rightarrow$ " meaning "*not* - *or* -"), my answer to the above question is in the negative.

One could, however, give to $p \rightarrow Oq$ a prescriptive interpretation by regarding it as an elliptic formulation with a concealed or suppressed

much greater detail, in an essay with the same title, published in my book *Logical Studies* (1957).

¹⁹ "A Note on Deontic Logic and Derived Obligation", *Mind*, 65, 1956.

second-order deontic operator. The words "if p then it ought to be that q " might very well be used for giving someone to understand that this is how things ought to (or may) be. A higher authority could thus address a subordinate authority. The full expression would then be $O(p \rightarrow Oq)$ where the second-order O is prescriptive and the first-order O is descriptive (cf. above, p. 135).

To say that $O(p \rightarrow Oq)$, when prescriptively interpreted, expresses a conditional norm seems to me correct. And the same is true of the first-order expression $O(p \rightarrow q)$. Indeed this last is what I should regard myself as the *prototype* logical form of a conditional (obligation) norm.

But how then shall we deal with the "paradoxes"? Does not $O \sim p$ "entail" $O(p \rightarrow q)$ and does this not mean that doing the forbidden "commits" the agent to doing just anything? The fault here is with the use of the words "entailment" and "commit". The fact that $O \sim p \rightarrow O(p \rightarrow q)$ is what I have called a "deontic tautology" only means that in a normative code in which it is (categorically) forbidden that it is the case that p it cannot "consistently" be permitted that that p is the case in conjunction with something else. This, moreover, is "consistency" in the sense of "rational norm-giving", not in the sense of "formal logic".

Similarly, there is no question of a categorical obligation "entailing" a(n infinity of) hypothetical norm(s) to the effect that something or other by being the case would "commit" one to doing the obligatory thing. The fact that $Oq \rightarrow O(p \rightarrow q)$ is a "deontic tautology" only means that, if something is obligatory, then it cannot consistently with this (in the sense of "rational norm-giving") be the case that the negation of this thing, under some circumstances, is permitted.

The Paradoxes of Commitment and of Derived Obligation are thus harmless. The air of paradox originates through an unwarranted *reading* of such formulas as $O \sim p \rightarrow O(p \rightarrow q)$ or $Oq \rightarrow O(p \rightarrow q)$, which are valid in any sound system of deontic logic. These formulas constitute as such no obstacle to regarding the form $O(p \rightarrow q)$ as the prototype form of a conditional norm.

Let there be a conditional norm $O(p \rightarrow q)$. Assume now that, on some particular occasion o , it is the case that p . Is it then not absolutely natural to say that, on this occasion, it also ought to be the case that q ?

This looks like an inference, *modo ponente*, from $O(p \rightarrow q)$ and p_o to Oq_o . This, of course, is not in ordinary logic a valid schema. Formally valid would be an inference from $p_o \rightarrow Oq_o$ and p_o to Oq_o or from $(o)(p_o \rightarrow Oq_o)$ and p_o to Oq_o . Shall we conclude then that, after all, the correct expression for the conditional norm is $p \rightarrow Oq$ and not $O(p \rightarrow q)$? I think the solution to the puzzle is as follows:

When from the norm to the effect that it ought to be the case that, if p then q , in combination with the fact that it is the case that p we "infer" that it ought to be the case that q , we are actually using "ought" here in

two senses. The first is the ought of a (genuine) norm or a "deontic Ought" as I shall also call it. The second is what I propose to call the ought of a "practical necessity" or a "technical Ought". For the second, "must" is perhaps a better word than "ought".²⁰

The technical Ought is not properly a normative concept. It expresses that something or other has to be (done) in order that something else is (attained or secured). Unless the first is, the second will not be. In the case of our hypothetical norm $O(p \rightarrow q)$ above, Oq_o says that unless in a situation o when it is the case that p it is not also the case that q , the norm will not be satisfied. Something it prohibits will be the case.

Since the two oughts are (very) different, we should distinguish them also in the symbolism—for example as O_d and O_t .

A technical ought-statement is, nearly always, *elliptic* in the sense that it contains implicit reference to something which will not be if that which, in the technical sense, ought to be is not. There are a great many things which can be referred to thus. A common characteristic of them seems to be that they are actual or potential *ends* of human action, something we may intend, or want, or wish to attain. The thing of which a technical ought is predicated usually has the character of a *means* to this end.

When a technical ought-statement is supplemented with mention of the (actual or potential) end, we have a statement of the type often called a *hypothetical* norm or imperative. In ordinary language such norms are formulated, sometimes as if-then-statements ("If you want — —, you must (ought to, have to) — —"), sometimes as in-order-to-statements ("In order to attain — it is necessary that —"), sometimes as "Unless"-statements ("Unless — —, — — will (not) be").

Hypothetical norms in this "technical" sense must not be confused with conditional norms in the "deontic" sense—even though the latter too may, not inappropriately, be called "hypothetical", and the former "conditional". I myself prefer to call the former *technical norms*.²¹

As seen, one can from the deontic Ought $O_d(p \rightarrow q)$ "extract" a conditional technical Ought $p \rightarrow O_t q$ where the latter formula is an elliptic way of saying that, if it is the case that p then it must also be the case that q in order that the norm $O_d(p \rightarrow q)$ be satisfied, the requirement which it imposes fulfilled. But we could also "extract" from it a conditional $\sim q \rightarrow O_t \sim p$ which says that if it is not the case that q then, in order that the norm $O_d(p \rightarrow q)$ be satisfied, it must also not be the case that p . Or we could co-ordinate to $O_d(p \rightarrow q)$ a technical Ought $O_t(p \rightarrow q)$ which then says, elliptically, i.e. without explicit mention of the end in view, that, in

²⁰ Cf. the discussion of practical necessity, the "must", and what is here called "technical norms" in the first essay in this volume, above, pp. 1–17. On the two senses of Ought, see also above p. 74 and 96.

²¹ Cf. *Norm and Action*, p. 9f.

order that $O_d(p \rightarrow q)$ be satisfied, it must be the case either that not p or that q .²²

The agent who is under the obligation $O_d(p \rightarrow q)$, or another agent who is anxious to "enforce" this obligation, may from the technical $p \rightarrow O_d q$ (and correspondingly from $\sim q \rightarrow O_d \sim p$) "extract" a new norm or order in the particular situation when the state that p (or that $\sim q$) happens to obtain. This would be a norm $O_d q$ (or $O_d \sim p$), the existence-span of which ranges over this or that particular situation o when it is the case that p (or $\sim q$) (cf. below, p. 165f).

II

What is the logic of the technical Ought, O_t ? Answer: it is the same logical structure as the logic of the deontic Ought, O_d . This is a reason, I presume, why we naturally use the same word "Ought" in both cases—but also a reason why one is inclined to confuse the two.

The identity of logical structure referred to is not a triviality. It is a fact which emerges from what has been termed "a reduction of deontic logic to alethic modal logic".

The technical Ought, as we know, is "elliptic". It refers to a necessary condition of some state of affairs—for example the attainment of some good (end) or the avoidance of some evil. Call this state S . The technical Ought, in relation to this state, can then be defined as follows: $O_t p =_{df} N(S \rightarrow p)$, where " N " is the conventional symbol for necessity.

We now stipulate that the state that S is possible (its contradictory state not necessary). If then the definition of the technical Ought is used for translating the axioms A1–A4 of the standard system of first-order deontic logic we obtain a fragment of a system of alethic (ordinary) modal logic known as the System M or T . This is the (non-trivial) sense in which deontic logic may be said to be reducible to standard modal logic and also the sense in which the logic of the technical and the deontic Oughts are the same formal structure.

The *locus classicus* of this reduction is Alan Anderson's note in *Mind*, 67, 1958 entitled "A Reduction of Deontic Logic to Alethic Modal Logic".²³ Later, the idea was taken up by myself and further developed in

²² Cf. the discussion of "derived obligations" in the essay "Practical Inference", above pp. 14ff. See also above p. 70f on the foundation of normative statements in technical rules.

²³ The idea is already found in an earlier, mimeographed paper by Anderson, "The Formal Analysis of Normative Systems", Technical Report No. 2, Office of Naval Research, Group Psychology Branch, New Haven, 1956. Essentially the same "reductivist" idea also occurs in a mimeographed paper by Stig Kanger, "New Foundations for Ethical Theory", Stockholm, 1957, reprinted in *Deontic Logic: Introductory and Systematic Readings*, ed. by Risto Hilpinen; D. Reidel, Dordrecht, 1971.