C.D. Broad

Ethics

Edited by C. Lewy

Nijhoff International Philosophy Series 20

Martinus Nijhoff Publishers

ETHICS

NIJHOFF INTERNATIONAL PHILOSOPHY SERIES

VOLUME 20

For a list of other volumes in this series see final page of the volume.

C.D. Broad

Ethics

Edited by C. Lewy

1985 **MARTINUS NIJHOFF PUBLISHERS**
a member of the KLUWER ACADEMIC PUBLISHERS GROUP
DORDRECHT / BOSTON / LANCASTER

Distributors

for the United States and Canada: Kluwer Academic Publishers, 190 Old Derby Street, Hingham, MA 02043, USA
for the UK and Ireland: Kluwer Academic Publishers, MTP Press Limited, Falcon House, Queen Square, Lancaster LA1 1RN, UK
for all other countries: Kluwer Academic Publishers Group, Distribution Center, P.O. Box 322, 3300 AH Dordrecht, The Netherlands

Library of Congress Cataloging in Publication Data

Broad, C. D. (Charlie Dunbar), 1887-1971.
Ethics.

(Nijhoff international philosophy series ; v. 20)
1. Ethics--Addresses, essays, lectures. I. Lewy, Casimir. II. Title. III. Series.
BJ1012.B686 1985 170 84-22791

ISBN-13:978-94-010-8739-1 e-ISBN-13:978-94-009-5057-3
DOI:10.1007/978-94-009-5057-3

EDITOR'S PREFACE

This volume contains C.D. Broad's Cambridge lectures on Ethics. Broad gave a course of lectures on the subject, intended primarily for Part I of the Moral Sciences Tripos, every academic year from 1933 – 34 up to and including 1952 – 53 (except that he did not lecture on Ethics in 1935 – 36).

The course however was frequently revised, and the present version is essentially that which he gave in 1952 – 53. Broad always wrote out his lectures fully beforehand, and the manuscript on Ethics, although full of revisions, is in a reasonably good state. But his handwriting is small and close and in places difficult to decipher. I therefore fear that some words may have been misread.

There was an additional complication. In the summer of 1953 Broad revised and enlarged two sections of the course, namely the section on "Moore's theory" and that on "Naturalistic theories" (both sections occur in Chapter 4). The revised version of the section on Moore is undoubtedly superior to the earlier version, and I have therefore included it. But in my opinion this is not true of the new version of the section on naturalistic theories: although more comprehensive than the earlier version, it is not only repetitive in itself, but also repeats, sometimes almost verbatim, passages which occur elsewhere in the lectures. In brief, the new version is not fully integrated with the rest of the course. I have therefore discarded it, and have included instead the earlier version of the section, as it was given in the lectures of 1952 – 53.

I have tried to reproduce the text as far as possible as it is in the manuscript. But I have expanded Broad's abbreviations, and have introduced greater uniformity in punctuation, spelling, and the use of capital letters, italics and quotation marks. I have also added the footnotes, most of which give references to the works discussed in the text.

Some of the material included in these lectures was published by Broad in the form of articles, and I am very grateful to a number of persons and institutions who own the respective copyrights for permission to reproduce it here. The Editor of the Aristotelian Society had kindly allowed me to include the material which Broad published in the *Proceedings* of the Society, vol. 34, 1933 – 34 ("Is 'Goodness' a name of a simple non-natural quality?"); the Editor of *Philosophy* the material published in that journal, vol. 15, 1940 ("Conscience and Conscientious Action"); the Hibbert Trustees the material published in the *Hibbert Journal*, vol. 48, 1950 ("Egoism as a Theory of Human Motives"); the Managers for the Herbert Spencer Lectures, University of Oxford, the material given by Broad as the Herbert Spencer lecture for

1953 ("Self and Others", published in C.D. Broad, *Critical Essays in Moral Philosophy,* ed. by D.R. Cheney, London, 1971); the Editor of *The Journal of Aesthetics and Art Criticism* the material published in that journal, vol. 13, 1954 ("Emotion and Sentiment"); Messrs Norstedt & Söners the material which Broad published in the *Festkrift tillagnad Karl Olivercrona*, 1964 ("Obligations, Ultimate and Derived"); and Professor P.A. Schilpp, as Editor of the Library of Living Philosophers, some portions of the material published in *The Philosophy of G.E. Moore*, 1942 ("Certain Features in Moore's Ethical Doctrines").

I should point out however that the greater part of the present book has not been published before. Moreover, the papers just listed were extracted by Broad from his course on Ethics, and I believe that their full value is best displayed in their original context. The exception is Broad's Inaugural Lecture on "Determinism, Indeterminism, and Libertarianism" (Cambridge University Press, 1934) which was especially written for the occasion and then included in his course of lectures. It is reprinted here, by kind permission of the Syndics, as the first part of Chapter 5. But even in this case the second part of the Chapter, which continues the discussion of the topic, has not been published before.

C. Lewy
Trinity College
Cambridge

TABLE OF CONTENTS

Chapter 1

THE SUBJECT-MATTER OF ETHICS

1. The raw material

Ethics may be described as the theoretical treatment of moral phenomena. I use the phrase "moral phenomena" to cover all those facts, and only those, in describing which we have to use, *in a specifically moral sense*, such words as "ought", "right", "good" and their opposites, or any others which are merely verbal translations of them. (This is not intended as a definition; if it were it, would be circular; for I have had to introduce the phrase "in a specifically moral sense" into my description of moral phenomena.) I have had to do this, because words like "ought", "right", and "good" are also used in various non-moral senses, and then Ethics is not directly concerned with the facts which they describe.

The difference between the moral and the non-moral use of such words can be brought out by the following examples. Consider the three sentences: "You ought to keep your promises"; "It is wrong to cause needless pain intentionally"; and "Nero was a bad man". Contrast them respectively with the three sentences: "You ought to change your clothes as soon as possible if you get wet"; "It is wrong to wear a white tie with a dinner jacket"; and "Nero was a bad actor". In each of the first three sentences we have one of these words used in its specifically moral sense; in each of the second triad we have the corresponding word used in a non-moral sense. At present I shall not attempt to discuss the differences and the relations between the moral and the non-moral uses of these words. We shall have to consider that carefully at a later stage. For the moment all that I want to do is to focus attention on sentences in which these words are used in a specifically moral sense and on the propositions which they ostensibly express. Let us call such sentences "moral indicatives".

Moral indicatives fall into at least two classes, viz. those which contain "ought" or "ought not" or some equivalent, and those which contain "right" or "wrong", "good" or "evil" or some equivalent. The former assert that a person is under a *moral obligation* to act or to refrain from acting in a certain way. E.g. "You ought to keep your promises", "You ought not intentionally to cause needless pain". Let us call these "deontic moral indicatives". The latter assign a certain kind of *value* or *disvalue* to an action, a habit, an experience, or a person. E.g. "It is *wrong* to break a promise",

"Courage is a *good* disposition". "A feeling of jealousy is an *evil* emotion", "Nero was a *bad* man". Let us call these "evaluating moral indicatives". As we have seen, there are also deontic and evaluating indicatives which are not specifically moral, e.g., "You ought not to eat peas with a knife" and "Nero was a bad actor".

The vast majority of sentences in the indicative are neither deontic nor evaluating. It will be useful to take some examples and compare them with deontic or evaluating sentences which are about the same subject. Cf., e.g., "You will often be inclined to break your promises" with "You ought not to break your promises"; "Lying is a habit which we acquire at school" with "Lying is wrong"; and "Nero had his mother drowned" with "Nero was a bad man". In each of these three contrasted pairs of sentences the first is a mere statement of fact, whilst the second is a statement of obligation or of value. Let us call such sentences "purely expository".

It is worth noticing that many words and sentences which seem to be purely expository are really partly expository and partly evaluatory. Take, e.g., the word "lie" and the sentence "That is a lie". If the sentence were interpreted in a purely expository way, it would be equivalent to: "That is a statement intended to produce a false belief about the subject with which it deals". But nearly always it means more than this. What it expresses implicitly would be more explicitly expressed by the following sentence: "That is a statement intended to produce a false belief, and, as such, morally wrong". The sentence "That is a lie" really expresses as a rule a combination of a purely expository proposition and an evaluating proposition based upon the former. I call such words and sentences "amphibious".

They are very numerous and very dangerous. Words like "democratic", "reactionary", "unscientific", and hundreds more, are amphibious. The danger is that they are used in one part of an argument in a purely expository sense and in another part in the mixed sense which involves a pure exposition *plus* an evaluation based upon it. E.g. the purely expository sense of "democratic" is "determined by the votes of a majority of the persons affected"; the amphibious sense is "determined by the votes of a majority of the persons affected, and, as such, politically desirable". It is easy to start by admitting that a measure would be democratic in the purely expository sense and to end by admitting that it is politically desirable, without noticing the suppressed and not very plausible premiss that anything which is determined by the votes of the majority of those affected is as such politically desirable.

We may sum this up by saying that the raw material of Ethics is *moral phenomena*; that moral phenomena are what we refer to when we use deontic and evaluatory sentences in a *specifically moral sense*; and that we can see roughly what this includes and what it exludes by taking examples and contrasting them (a) with sentences which *are* deontic or evaluatory but are *not*

specifically moral, and (b) with sentences which are *purely expository*.

2. Subdivisions

The next question is: What does Ethics do with this raw material?

I think that the topics discussed by writers on Ethics may be subdivided as follows. In the first place, they may be divided into a central part and a peripheral part. The central part consists of properly ethical topics; the peripheral part consists of certain other subjects which have to be discussed because they are so closely connected with Ethics proper. I will now say something about each.

2.1. Central part

The central part has two sub-divisions which we will call "Analytical" and "Synthetical" Ethics.

2.11. Analytical ethics

This is concerned with the *analysis* of moral phenomena. Under this heading come such questions as the following. What is a person really doing when he utters a moral sentence in the indicative, e.g., when he says "*A* ought not to have broken his promise to *B*"? Is the speaker really *asserting an opinion* (correct or incorrect); or is he only expressing a certain kind of emotion which he feels towards the incident? If he is asserting an opinion and not merely expressing his feelings, what kind of opinion is he asserting? Is he asserting merely that he or most people have a certain kind of feeling when they contemplate breaches of promise? Or is he asserting something about *A*'s action toward *B* which is as independent of the feelings of observers as if he had said that *A* is *B*'s second cousin? Suppose that what the speaker is asserting when he says that *A* ought not to have broken his promise to *B* is something quite independent of his own or other men's feelings towards breaches of promise. Then we can raise the question whether such words as "ought" and "right", when used in the moral sense, stand for qualities or relations which are quite unique and peculiar. Or can these notions perhaps be analysed entirely into non-moral terms, e.g. psychological or sociological or biological? Failing this, can some moral notions, e.g. "morally good", be analysed in terms of others, e.g. "ought", together with certain non-moral notions? E.g. "morally good" be defined as "what *ought* to be desired"? These, and many other questions of a similar kind, belong to Analytical Ethics.

2.12. Synthetical ethics

This is concerned with such questions as the following. (i) Are there any synthetic connexions between one moral characteristic, e.g. goodness, and another, e.g. rightness? And, if so, what are they? It might be the case, e.g., that both "rightness" and "goodness" are simple indefinable notions and yet that it is a self-evident synthetic proposition that the right action in any circumstances is that which will produce the most *good* or the least *evil*. (ii) Are there any synthetic connexions between moral characteristics, e.g. rightness, and certain non-moral characteristics? There are alleged to be many such connexions. E.g. it is held that any act of intentionally breaking a promise is as such wrong, and that any act of giving innocent pleasure to another is as such right. Synthetic Ethics has to consider all the more important of these alleged synthetic universal propositions connecting certain moral characteristics with certain non-moral characteristics. It has to consider with regard to each of them whether it is true without exception or only true in most cases. If any of them is true without exception, it will have to consider whether it is a *necessary* truth, like the fact that everything which has shape has extension, or a *contingent* truth, like the fact that all animals which chew the cud have cloven hoofs. (iii) Suppose that there are several non-moral characteristics, e.g. being an intentionally false statement, being an intentional infliction of needless pain, being a breach of promise, and so on, each of which would make any act of the kind wrong. We shall then have to raise the question whether they are just a purely haphazard collection or whether they can be classified under a few general headings? Can they perhaps all be reduced in the end to a single principle, e.g. that an act is right if and only if it produces more pleasure or less pain than any other act open to the agent in the circumstances? These are a fair sample of the questions that fall within Synthetical Ethics.

2.2. The peripheral part

This falls into three subdivisions, viz., *Moral Psychology, Moral Epistemology,* and *Metaphysic of Morals.*

2.21. Moral psychology

We ascribe moral characteristics only to persons who are capable of reflexion and deliberate action; and to the actions, dispositions, habits, emotions, and desires of such persons. It is therefore essential for anyone who is concerned with Ethics to have clear ideas about human psychology. We use terms like "motive" and "intention" very loosely in daily life; and we are far from clear as to what constitutes a person capable of deliberate choice as contrasted with an infant or an animal. Men are in certain respects very like other animals,

and in certain other respects very unlike them, and their moral life is largely concerned with the problems which arise from their mixed nature. It is therefore essential for the moralist to clear his ideas about human psychology in general and particularly about the psychology of emotion, volition, and deliberate action. The following are examples of questions which arise under the head of Moral Psychology. What kinds of motives act on human beings? Are they capable of being reduced to a few fundamental kinds, or possibly to a single kind? Is all deliberate action either explicitly or implicitly egoistic, or is there genuinely non-egoistic action? Does the belief that a certain course of action would be *right* or that it would be *wrong* suffice to constitute a motive for or against doing it, as the case may be? Or must it always be reinforced by some non-moral motive, such as the desire to be thought well of by others or at any rate by oneself?

2.22. Moral epistemology

Under this head come questions about the nature of our moral knowledge and beliefs, the origin of our moral ideas, and the kind of evidence which we have for our moral judgments. Examples of such questions are the following. Are the notions of *right* and *wrong, ought* and *ought not*, etc. wholly acquired by each individual in the course of his life, or are they in some sense innate? If they are wholly acquired, how precisely do we acquire them? If they are in some sense innate, in what precise sense are they so? And by what process do they become clear and explicit as we grow up from infancy? Again, are universal moral propositions, such as "Promise-breaking is as such wrong", empirical generalizations or synthetic *a priori* judgments or irrational prejudices imbibed in infancy?

Moral Epistemology cannot be pursued in isolation. These questions must be considered along with similar questions that can be raised about non-moral notions, such as causation, and non-moral generalizations, such as "Every event has a cause". To separate Moral Epistemology from Epistemology in general is a disadvantage to both. The former tends to become amateurish; and the latter tends to give answers which claim to have general validity but which ignore the special problems raised by *moral* concepts and *moral* judgments.

It must be added that there is a very close connexion between some of the problems of Moral Epistemology and some of those which belong to the Central Part of Ethics. Any view about the correct analysis of moral characteristics will tend to make some theories about the nature and origin of our ideas of those characteristics more probable and others less so. And the converse relation holds also. Suppose, e.g., that rightness is a simple unanalysable characteristic. Then it will be difficult to believe that our idea of rightness is of empirical origin. Conversely, if we are persuaded on

epistemological grounds that all our ideas are of empirical origin, it will be difficult to believe that rightness can be a simple unanalysable characteristic.

2.23. Metaphysic of morals

Ethics leads sooner or later into metaphysical problems, though we can go a considerable way without needing to encounter them. The most obvious point of contact is over the question of Freedom and Determinism. Moral judgments about men's actions plainly assume that there is a sense in which a man who chose alternative *X* could instead have chosen the different alternative *Y*. Now there are certain senses of "could" in which this assumption is plainly true in many cases. But it is doubtful whether any of them are the sense of "could" in which moral judgments about actions presuppose that the agent *could* have chosen a different alternative.

At this stage two problems arise. (i) Can we state clearly what is the sense of "could" in which moral judgments presuppose that an agent could have acted otherwise than he did? (ii) If so, can we admit that any agent could have acted otherwise in this sense of "could"? The first question is hard enough to answer. But the second brings us up against the following problem. Does not the supposition that any choice is free in the required sense conflict with self-evident principles, like the Law of Universal Causation, or with empirical generalizations for which there is overwhelming evidence? Now this last question takes us right out of Ethics into Metaphysics, since it requires us to formulate clearly the Law of Universal Causation and to consider what is the nature of the evidence for it.

I do not think that Ethics can possibly shirk the question of Freedom and Determination. But there is at least one other question which, though not of such great theoretical importance, is of very great practical importance to Ethics. This is the question whether each individual ceases to exist at the death of his present body or whether at least some persons survive the death of their present bodies and continue to live as active intelligent beings under profoundly different conditions. Whether we survive the death of our present bodies or not, we have duties and can act rightly or wrongly. But the *details* or our duties, and the *importance* of acting rightly or wrongly, might be very different according to which alternative is true. On the other hand, the whole notion of duty and of right or wrong conduct ceases to have any application if we are not free to choose between alternatives, in the sense of "freedom" which is presupposed by moral judgment. For these reasons I say that the question of human survival or non-survival of bodily death is of great practical importance to morality, but it is not of the same importance for ethical theory as the question of Freedom or Determinism.

It should be noted that the connexion between Ethics and Metaphysics is two-sided. E.g., one might argue: "We certainly have duties. Therefore we

must be free in the sense required for this. Therefore any metaphysical principle which seems to be incompatible with such freedom must either be false or not really incompatible with it". Or one might argue: "There are self-evident metaphysical principles, e.g. the Law of Universal Causation, which make it impossible that we should be free in the sense required for moral responsibility. Therefore the notions of duty and moral responsibility, and any other notions which involve these, must be delusive".

Chapter 2

MORAL PSYCHOLOGY

I shall begin by explaining certain features which are common to all conscious beings, human or animal, as distinct from non-conscious beings such as plants and stones. Then I shall point out the main peculiarities of human minds in contrast to those of any known non-human animal. And I shall gradually work up to those parts of human psychology which are specially relevant to Ethics.

1. General properties of conscious beings

1.1. Dispositions

In Psychology the notion of *dispositions*, innate and acquired, is very important; and I will begin by explaining and illustrating it. Suppose that I am seeing a snake and feeling frightened of it. Then I am having an actual cognitive experience of seeing and an actual emotional experience of fearing. But, even when I am not seeing a snake and not feeling frightened, I may have a permanent disposition to fear snakes. This would be stimulated by the sight of a snake or a snake-like object, and it would then give rise to an actual experience of fear directed towards the object seen. Suppose I were to say "Smith is afraid of snakes". I should probably mean that he has this persistent emotional disposition, and not that he is actually having the experience of seeing and fearing a snake at the moment.

Now many psychological terms are used ambiguously; sometimes in what we will call the "occurrent sense", and sometimes in what we will call the "dispositional sense". When I say that a person "remembers X" or "believes Y", I may mean that he is actually remembering X or actually believing Y. If so, I use the words "remember" and "believe" in the occurrent sense. But much more often I mean that he has a disposition which could be stimulated at any time; and that, if and only if it were suitably stimulated, he would thereupon actually remember X or believe Y. This is the dispositional use of these words. It is evident that at any moment most of our memories, beliefs, knowledge, emotions, and desires exist only in the form of dispositions and not in the form of actual experiences.

It is important to notice that mental dispositions are not open to

introspection like actual experiences. We know a mental disposition only by description, as a hypothetical cause-factor in the total cause of certain recurrent similar experiences, e.g. experiences of remembering a certain event on a number of different occasions.

1.11. Hierarchy of dispositions

Dispositions can be arranged in a kind of hierarchy. The power to talk, e.g., is a disposition. This is not innate, but is acquired in childhood. A baby is not born with the power to talk, any more than a cat is. But a baby *is* born with the power to *acquire* the power to talk, whilst a cat is not. If conditions of the right kind are supplied, the baby will acquire the power to talk; and, according to the nature of the conditions supplied he will begin to talk in English or in French or in German as the case may be. But a kitten will never acquire the power to talk, no matter what conditions may be supplied.

Let us now generalise from this example. We may define a "*first-order*" disposition as follows. It is a disposition, which, when suitably stimulated, leads to a result which is *not* the acquiring or losing or modification of some other disposition. The disposition to blink when anything approaches one's eye, and the disposition to think of a certain person when his name is mentioned, are examples. The former is innate, the latter acquired; and both are of the first order.

A "*second-order*" disposition is a disposition to acquire or to lose a first-order disposition under certain conditions. The power to talk a certain language is a first-order disposition; the power to acquire the power to talk is a second-order disposition.

Dispositions of any order above the second can be defined on the same lines as our definitions of second-order dispositions. Lastly, we may define a "*supreme*" disposition as one which a person has no disposition either to acquire or to lose. I suppose that the power to form associations of ideas is a supreme disposition in human minds. Under certain circumstances we acquire certain special associations, and under other circumstances we may lose certain special associations which we have acquired. But as long as there is anything that could be called a human mind it has the general power to form associations of *some* kind.

1.12. Innate and acquired dispositions

Some dispositions are innate, e.g., the disposition to build nests is innate in most birds, and the disposition to learn to talk is innate in all normal human beings. But many dispositions are acquired in the course of experience and all of them are liable to be modified in the course of experience. The power to talk English is an acquired disposition, depending on the innate disposition to learn to talk and on the particular training which English children get from their mothers and nurses.

In order to account for the experiences or the behaviour of a person at any moment it is usually not enough to refer to the stimuli which are affecting him *at* that moment and to the experiences which he was having *immediately before*. It is very often necessary to refer also to experiences which he had in the remote past. This is perfectly obvious in the case of memory. To account for my now remembering an event which I witnessed last year it is not enough to refer to the present stimulus which acts as a reminder. We have to suppose that the experience which I had a year ago has set up in me a disposition to remember what I then saw, and that the present reminder stimulates this disposition and gives rise to the actual memory-experience.

This property of experiences to set up new dispositions and to modify pre-existing dispositions in the person who has them may be called the *mnemic* property. It is most strikingly illustrated by memory. But it is involved in every case where one is aware of a set of successive events as a series having a certain pattern, e.g. aware of a number of successive sounds as a tune. If there were no retentiveness one might have a *series of experiences*, but one could not possibly have *experience of a series*. It is quite possible to imagine that a very simple kind of creature, e.g. an oyster, has the former but not the latter. Such a creature could not be aware of itself as a *person* who persists throughout a series of changing experiences or of external things as *substances* which persist and have a series of changing states and relations.

1.121. Theory of mental structure and traces

We generally picture to ourselves the facts which I have been describing by means of the following theory. We think of the mind as having from the first a certain kind of "structure" which is partly rigid but is very largely plastic. We think of certain permanent features in this initial structure as corresponding to innate disposition. Then we think each subsequent experience as making a more or less permanent modification in this structure. Such hypothetical persistent modifications set up in the mind by its experiences are called "*traces*".

If we state the facts about the mnemic properties of minds in terms of the trace theory, they appear somewhat as follows. (i) Every experience leaves a more or less persistent modification of the mind or the brain or both, which may be called a *trace*. (ii) Traces do not just coexist passively side by side; they modify each other in accordance with certain laws. Traces left by certain later experiences link up with those left by certain earlier experiences. Instead of the traces t_1, t_2, ..., t_n, which were left respectively by a certain series of experiences e_1, e_2, ..., e_n, just coexisting side by side, you may get a single complex trace $t_{1,2,\ldots,n}$. This happens when we have, not merely a series of experiences, but the experience of a series of events, e.g. the hearing of a tune. In that case a subsequent reminder will tend to excite the whole complex trace

$t_{1,2,\ldots,n}$ and to produce, e.g. a memory of the tune. (iii) The total cause of any experiences contains at least two different factors. One is some present stimulus, external or internal. The other is some of the traces left by our earlier experiences, which are excited by the present stimulus.

The total dispositional pattern of a mind at any moment may be called its "*apperceptive mass*" at that moment. (The phrase is taken from Herbart.) Whether a stimulus will produce a conscious experience at all; and, if so, what the nature of the experience will be, depends very largely on the apperceptive mass which it encounters. This is illustrated by such facts as failing to see things which for some reason one does not want to find, noticing weak points in an opponent's argument and failing to notice similar points in our own, and so on.

1.13. Laws of association and reproduction

We can now raise three questions in terms of the trace-theory. (1) Under what circumstances do a number of simultaneous or successive experiences in a person tend to give rise to a single complex trace, as opposed to a number of isolated traces? (2) When a single complex trace has been formed, under what circumstances does it tend to be excited? (3) When a complex trace is excited what kinds of effect does this have on the actual experiences of the person concerned? The empirical rules which have been discovered on these questions are called the Laws of Association and Reproduction.

(1) (a) When a number of simultaneous and successive experiences are so stimulated that together they constitute a single experience with a single complex object, they leave a single complex trace. An example would be the successive auditory experiences which together constitute the hearing of a certain tune. (b) When two experiences X and Y occur together, and experiences like X and Y are often repeated together and seldom occur separately, the traces left by them tend to combine into a single complex trace.

(2) (a) If experiences like *some* of those which left a single complex trace recur in similar relations to each other, the trace as a whole tends to be excited. An example would be if, after having heard a certain tune played in a certain key on the piano, one were to hear the first few notes of the same tune whistled in a different key. This would probably suffice to excite the trace left by the first hearing of the tune.

(b) When the traces of his experiences X and Y have become associated through frequent repetition of two such experiences together, if an experience like one of them occurs alone it will tend to excite the trace of the other. Suppose, e.g., that one has on many occasions seen lightning and heard thunder in close succession, and that one has rarely seen lightning without hearing thunder soon afterwards and has rarely heard thunder without having seen lightning shortly before. If I now see a flash of lightning it will tend to excite

the whole complex trace left by the two frequently associated kinds of experiences.

(3) The effects of the excitement of a complex trace are very various. They depend on the nature of the exciting stimulus, on the general nature of the person's apperceptive mass, and on his particular interests and activities at the time. Suppose, e.g., that I now hear someone whistle a few notes of a tune which I formerly heard played in a different key on the piano and that this excites the complex trace left by that earlier experience. I may have a memory of the former experience. I may instead have an image of what the rest of the tune would sound like if it were whistled in the present key. Or I may merely have an experience of familiarity, which I might express by saying "I have heard something like that before". And so on.

Before leaving the subject of mental structure and traces I want to issue a warning. It is convenient to state the facts about dispositions, and retentiveness, and the formation of associations, and the reproduction of associated experiences in terms of the theory of mental structure and traces. But it must be remembered that this theory is largely a metaphorical way of describing the facts and not an explanation of them. We know nothing of mental structure except as the hypothetical part-cause of certain introspectable effects. And we know nothing of traces except as hypothetical effects of earlier experiences and hypothetical part-causes of later experiences.

2. Some peculiarities of human minds

I will now mention some features which are peculiar to human minds as contrasted with those of animals.

2.1. Lack of complex first-order dispositions

A baby is born with very few first-order dispositions, and such as he has are very simple in comparison with those of many animals and insects. Birds, e.g., have the innate first-order disposition to build nests at certain seasons of the year, and this comes into action as soon as they have reached sexual maturity. Now nest-building is a fairly long and complex series of coordinated actions which lead up to a very definite kind of result. This is done even on the first occasion without having been learned or deliberately thought out. A baby has no first-order disposition of anything like this complexity.

A baby is born mainly with disposition to acquire dispositions, e.g. with the power of learning to talk, learning to reason, learning to make abstractions, and so on. Under favourable conditions the use of these powers enables him to acquire more specific dispositions, e.g. the power to talk and

understand several languages, the power to make arithmetical calculations, the power to construct and follow arguments, and so on. An animal, on the other hand, is very limited in its innate powers to acquire dispositions. A human being is, and remains for a large part of his life, *teachable* by himself and by others; but the limits within which any animal is teachable are very restricted.

2.2. Intellectual analysis

Human beings have the power of analysing new situations and noticing that they are composed of elements which are already familiar in isolation or in other combinations. We may call this the power of intellectual analysis, comparison, contrast, and abstraction. The results of such analyses are preserved and made available for others in the words and the grammatical forms of language. There seems to be hardly any trace of this power in animals, and indeed one does not see how it could possibly exist without language or how language could possibly arise without it.

2.3. Intellectual synthesis

Using the materials which have been gained by reflection, comparison, and abstraction, and have been stored up in words and grammatical forms, we can construct ideas of things, people, and situations which we never have perceived. We can thus reconstruct in imagination the remote past; we can conceive of things and processes which are not and perhaps could not be present to our senses; we can conjecture the course of future events. Moreover we can make hypothetical conjectures. We can imagine that certain conditions in the past had been different from what they in fact were, and we can conjecture what would have happened if these different conditions had been fulfilled. Similarly we can imagine various alternative future developments of the present situation, and can see that, if a certain event should happen, *one* of these possible future developments will be realised, and that if a certain other event should happen, a *certain other* of them will be realised. In this way we may prepare ourselves beforehand for different possible future eventualities.

2.4. Reasoning

Closely connected with intellectual analysis and synthesis is what may be called the power of rational thinking, deductive and inductive. A person can recognise logical relations of entailment or inconsistency between one or more propositions and another proposition. If he knows or believes certain propositions and sees or thinks he sees that they logically entail a certain other

proposition he will be caused by this to believe the latter and will feel himself to be justified in doing so. If, on the other hand, he sees or thinks he sees that they are logically inconsistent with a certain other proposition he will be caused by this to reject the latter and will feel himself justified in doing so.

Again, one or more propositions may be so related to another proposition that, whilst they neither entail nor exclude it, they make it highly probable or highly improbable. If a person knows or believes the former propositions, and sees or thinks he sees that they stand in this kind of relation to the latter, he will be caused thereby to believe the latter with greater or less confidence, and he will feel justified in doing so.

We may sum this up by saying that some propositions are so related to others that the former constitute *evidence*, either demonstrative or probable, for or against the latter. Human beings are capable of seeing such relations between propositions which they contemplate, and they are capable of extending their knowledge and adjusting their beliefs and disbeliefs in accordance with evidence. This process is entirely different from the reproduction of ideas and the establishment of beliefs by association. The latter plays a very important in human life, as it does in the life of an animal. But an animal has nothing comparable to the power of recognizing evidence and adjusting its ideas and beliefs to the evidence available to it.

An extremely important department of reasoning is the process of discovering and testing general laws by reflecting on the regularities which we find among natural phenomena. Animals and primitive men are led by repetition and association blindly to take for granted the continuance of certain very obvious regularities, e.g. to take for granted that the sun will rise, that water will quench fire, and so on. But the most important regularities are not by any means obvious, and can be found only by deliberate search. And it is a very different thing to have a reasoned conviction of a general law, based on experiment and observation and reasoning, from having a mere blind conviction that the future will resemble the past. Animals in certain respects *act as if* they were aware of certain causal laws. But it is only men who have the explicit notion of particular causal laws and of the general principle that all natural phenomena are subject to such laws. And it is only men who can apply such knowledge to control and modify their environment and themselves.

2.5. Storing and transmission of culture

The following consequence of our powers of intellectual analysis and synthesis and reasoning is very important, and is so familiar that we are liable to overlook it. It is this. The discoveries and beliefs (true or false) of our ancestors and contemporaries are crystallised and embodied in language,

social institutions, buildings, machines, works of art, etc., to say nothing of books which are deliberately written to record them.

In a civilised community almost the whole of one's environment throughout life, from the cradle to the coffin, is man-made. That is true even of the fields and woods in a long-settled country like England. Thus, although a baby does not inherit biologically the dispositions which his ancestors have acquired, the latter are to a large extent embodied in his material and social environment. In this way a community of men with a continuous history bears some resemblance to a single very long-lived individual whose experience is constantly growing and leaving traces which are assimilated into a more and more complex apperceptive mass.

There is no reason to think that there has been any appreciable change, either qualitative or quantitative, in the innate intellectual powers of man in the course of recorded history. But this has been of little importance. All that matters is that each generation should be able to assimilate the crystallised thought of its ancestors, that it should be able to make its own additions and modifications, and that it should be able to hand on the increased and modified stock in a form which the next generation can assimilate.

It is plain that a great many human communities have reached a stage at which each generation makes practically no addition to or modification in the stock of ideas which it absorbs. Presumably most savage communities have reached this stage. If we compare such a community to an individual, it is like a still vigorous man in late middle-life who has become quite impervious to new ideas.

Again, it is clear that there is a very definite limit of achievement which no community can surpass unless the innate intellectual capacities of its members can be improved. Suppose that the stock of ideas to be absorbed by each generation increases rapidly in quantity and changes profoundly in character. Then it may be beyond the capacity of existing human minds to assimilate it and make use of it. The difficulty has so far been met up to a point by specialization and the development of experts. But there is a danger of the various experts and specialists getting very narrow-minded and wholly out of touch with each other and with the ordinary members of the community. If we compare such a community to an individual, it is rather like one who has become dissociated and has developed multiple personality under the stress and complexity of life. It seems to me that modern industrial communities, under the impact of pure and applied sciences, have reached the stage at which they have bitten off more than they can chew, unless the innate mental dispositions of their members can be greatly modified and improved by some process of selective breeding.

There is hardly anything among animals analogous to what I have been describing. The nearest apparent analogy would be an insect-community,

such as a bee-hive or an ant-hill; but the resemblance is quite superficial. In the first place, it could be compared only to a human society which had completely ossified, and not to one which was still growing in knowledge and power. Secondly, insects are typical examples of exactly the opposite mode of procedure to that which is characteristic of human minds. They produce their results by means of extremely elaborate innate first-order dispositions, and they show no trace of intellectual analysis, synthesis, reasoning, and discovering and applying causal laws.

2.6. Reflexive powers

Every human mind has some power of introspecting and thinking about itself, its experiences, actions, and dispositions. I call this *reflexive cognition*. Again a person feels emotions towards himself in respect of his real or supposed qualities, defects, doings, and sufferings. This may be called *reflexive emotion*. Lastly a person has, or seems to have, the power of deliberately altering his own character and dispositions within certain ill-defined limits. This may be called *reflexive action*. These reflexive powers and activities are plainly very important in connexion with morals. E.g. what we call "conscience" is a particular department of them. There is, I think, no trace of them in wild animals; though I think that there may be faint traces of them in certain domesticated animals, such as dogs. The latter have undergone a severe training in cleanliness and table-manners while young, and they sometimes seem to be affected with something that looks like the pangs of a guilty conscience.

2.7. Selfhood and personality

We may suppose that all the experiences which a single animal has at any one moment are to some extent interconnected and form a single total phase of experience. And we may suppose that the successive total phases of experience of a single animal are to some extent interconnected and form a single total strand of experience. Let us call those two kinds of unity respectively *transverse* and *longitudinal* psychic unity.

These two kinds of unity may be present to very different degrees. This can be seen by contrasting one's state of consciousness when fully awake and attentive with what it is when one is drowsy, distracted, or delirious. Presumably in animals it is at best much less intimate than in men in their normal waking state. When the degree of longitudinal and transverse unity among the experiences of an individual reaches a certain level we can say that these experiences all belong to a single *self*. There is no definite lower limit of unity above which one could say unhesitatingly that experiences *do* belong to a self

and below which one could unhesitatingly say that they *do not*. It would be felt to be absurd to call an oyster or a tapeworm a self; one would feel doubtful about an intelligent dog or a young baby; and one would have no hesitation about a normal child of ten years old. The best thing to say is that selfhood has degrees, which depend upon the degree of transverse and longitudinal unity among the experiences of an individual. Unless this unity reaches a certain level we decline to call the creature a self at all; if it only slightly surpasses that level we say that it has a very low degree of selfhood; if the unity among its experiences is very intimate we say that it has a high degree of selfhood.

Now there might be a considerable degree of transverse and longitudinal unity among a set of simultaneous and successive experiences without that set containing any *reflexive* experiences. If so, we might have to say that the experiences all belong to a self, but we should certainly have to deny that this self was *self-conscious*. A set of experiences belongs to a self-conscious self only if it includes in it beliefs or emotions or desires about other experiences in the set or about the self whose experiences they are. We might define a *person* as a self-conscious self. We must remember that such a self need not at all moments be conscious of itself, and that it will never at any moment be aware of more than a small portion of its own experiences and their mutual relations. Self-consciousness, like every other kind of mental state, exists very often only in the dispositional form and not as an actual experience.

Now the presence of reflexive experiences within a group of inter-connected experiences enormously increases the internal unity of the group. E.g. the very fact that one experience in a group is a *memory* of another experience in that group constitutes an important relation between the two. It links together experiences which are widely separated in time. So it is true to say that a person has a much higher degree of selfhood than any self which is not self-conscious and is therefore not a person. Some people might refuse to give the name "self" to any mind which was not capable of reflecting on itself and its own experiences. If the word "self" is used in this restricted sense, it becomes identical in application with the word "person". I prefer to use the word "self" in the wider sense which I have explained, and to say that selfhood is capable of higher and lower degrees. This is merely a question of terminological usage. What is not a mere question of words, but is a matter of fact is this: That a set of successive and simultaneous experiences may be more or less closely interconnected; that it may or may not include reflexive experiences; and that, if it does include reflexive experiences, it will *ipso facto* have a more elaborate kind of structure and a higher degree of internal unity than if it does not.

2.8. Internal conflict

In a person there is the possibility of internal conflicts of a quite unique kind. E.g. he may want two or more alternatives which are incompatible with each other, and he will eventually have to decide to aim at one of them and to give up the possibility of satisfying his desires for the others. Again, he may want a certain end, but may dislike the means without which he cannot possibly reach it. The process by which alone a desired end can be reached may be positively painful, or it may involve strenuous exertion when the agent would prefer to be idle and passive. In such cases the person has to force himself, against many of his wishes, to carry out the process of fulfilling a certain desire.

These are examples of conflicts where the agent is fully conscious of the conflicting factors in himself. But there is good evidence that there are also conflicts where the agent is aware only of one of the conflicting factors in himself and is ignorant of the other factors.

2.9. Specifically moral experiences

Conflicts could occur in a person who had no ideas of right or wrong, moral good or evil, or moral obligation. But human beings do have such ideas, and a special kind of conflict arises in connexion with them. When a normal human being believes a certain course of action to be right this belief constitutes for him a reason or motive for doing it; when he believes it to be wrong it constitutes for him a motive against doing it. This motive very often conflicts with other motives. What we believe to be right may be in other respects highly repulsive to us, and what we believe to be wrong may be in other respects highly attractive to us. In such conflicts between specifically moral motives and others the influence on the will exercised by the belief that an act would be right or that it would be wrong feels very peculiar. We give to this feeling the name "sense of duty" or "feeling of obligation".

2.10. Summary

The nine points which I have enumerated and described above will give a fair idea of the characteristic differences between human minds and the minds of all known animals. I think that in recent years many people have been inclined to insist on the likeness between men and animals, and to try to ignore, minimise, or explain away the specifically human characteristics. This seems to me to be a gross mistake. However the differences between human and animal minds may have originated in the remote past, they are now so great as to be differences in kind and not merely of degrees. The two main causes of

this mistake have been the work of the psycho-analysts and the theory of evolution as applied to human beings. I will therefore say a little about these two points.

(1) The work of psycho-analysts and others has undoubtedly shown that the beliefs, actions, and emotions of ordinary men are much more irrational than is commonly supposed. But this does not in the least diminish the gulf between human and animal minds. Only a reasoning being can reason badly and persuade himself that the products of non-rational processes were reached by pure reason. Animals are not *ir*rational, they are *non*-rational. What can truly be said is this. It is because of the non-rational characteristics which we share with animals that we so often make such an irrational use of the rational characteristics which are peculiar to us.

(2) About evolution it is worth while to make the following remarks. (i) To describe the evolution of a thing is to describe either (a) the successive phases through which that thing itself went before it had become as it now is, or (b) the characteristics of each of a series of its biological ancestors going backwards in time from near ancestors that are very much like it to remote ancestors which are very much unlike it. On either alternative the thing will be said to have evolved if and only if the earlier terms of the series are on the whole simpler and less efficient than the later terms. Now it is plain that this is an account of how the thing *became* as it now is, and is not in any sense a substitute for, or a correction of, an account of *what* it now is. But there is always a strong temptation to forget this elementary fact. We are tempted to think that, if the apparently complex Z evolved by gradual stages from the much simpler A, the characteristic peculiarities which we find in Z are only apparent or are only the properties of A in a disguised form. E.g. if human minds, which have the power of reasoning, evolved from animal ancestors which had only the powers of association and conditioned reflex behaviour, we may be tempted to believe that human minds have not really the power of reasoning or that that power is only association and conditional reflex behaviour in a disguised form. There is really no logical ground for any such conclusion. And, if we try to think what this talk about Z being only A in a disguised form means, we shall find that it means little or nothing. (ii) Let us suppose that the minds of our *remote* ancestors did not differ in any important respects from those of animals, and that the minds of our *immediate* ancestors and our contemporaries do differ fundamentally in certain respects from those of animals. Then there are two possible alternatives. One is that the change took place continuously by insensible degrees; the other is that there were finite jumps at certain stages. I think it is important to insist that the fact that there is a difference of kind and not a mere difference of degree between the earlier and the later members of an evolutionary series does not necessitate that there have been finite jumps at certain points in the series. For

many people object to the notion of such jumps in the course of evolution as "unscientific" and "superstitious". If they think that they would be committed to accepting such jumps by admitting that the human mind differs in kind from any animal mind, they may be inclined to ignore or to try to explain away the fact that it does so differ.

Now the facts seem to me to be as follows. You have may qualitative discontinuity together with quantative continuity. E.g. the powers of intellectual analysis and synthesis and of reflexion are powers which do not exist in animals and do exist in us. Therefore whenever they first arose, and whatever may have been the conditions at the time, there was at that state a *qualitative* discontinuity. But at their first beginning they may have been so limited in range, so feeble in degree, so rarely exercised, and possessed by so few individuals as to have been infinitesimal. So there may have been no finite *quantitative* jump at these stages. We can illustrate this by a physical analogy. Suppose you turn on the current in an electric radiator. The wire gets gradually hotter and hotter and at a certain moment it begins to glow. When it does so there is a qualitative discontinuity, for it was black before that moment and is red afterwards. But there is no finite quantitative jump; for the red colour starts with infinitesimal intensity.

I would add in conclusion that I see no objection myself to the possibility of sudden finite jumps in the course of evolution. It is merely a question of fact whether they do or do not happen. All that I am concerned to assert here is that the existence of finite qualitative differences between the earlier and the later terms of an evolutionary series is quite compatible with the absence of finite quantitative jumps at any stage in the series.

3. Classification of experiences

We cannot define the term "an experience" any more than we can define the term "red colour". But it is quite easy to give examples which everyone can recognize. A person is having an experience whenever he is feeling tired, feeling a sensation of hotness or a twinge of toothache or an emotion of rage. He is having an experience whenever he is perceiving or imagining or remembering any event or thing or person. He is having an experience whenever he is believing or disbelieving or supposing any propositions or making an inference. And he is having an experience whenever he is desiring anything or deliberately striving to get, to keep, or to avoid anything. Each of us knows perfectly well what it is to have experiences because (a) he has them and (b) he has the power of reflexive cognition and can be aware of his own experiences.

3.1. Pure feelings and cognitions

The first division among experiences is made as follows. There are certain experiences which have qualities but do not have objects. These may be called *pure feelings*. The natural question to ask with regard to a feeling is "*How* are you feeling?". And the natural answer is some adjective or adverb, like "Hot" or "Tired" or "Cross". To feel tired is to be feeling in *a certain way*; it is not to be *aware of a certain object*. On the other hand, there are many experiences about which it is natural to ask: "What is the *object* of your experience?" or "What is it *about*?". If a person says that he is having an experience of seeing or hearing or thinking, it is natural to ask: "*What* are you seeing or hearing or thinking about?". And the answer that one expects is some substantive or phrase equivalent to a substantive; e.g. "A red flash", "A squeaky noise", or "The square-root of -1". I shall say that experiences of the latter kind "have an epistemological object" or that they are "epistemologically objective". It is important to notice that an experience may be epistemologically objective even if it be a *delusive* perception or a thought of something which does not and perhaps could not exist. A person who seems to see in a dream a man pointing a revolver at him is having an epistemologically objective experience, although there is no *ontological* object (i.e. no actual man pointing an actual revolver at him) corresponding to it. Similarly, a person who is thinking of a phoenix is having an epistemologically objective experience. He is certainly thinking of *something*. If he were thinking of a dragon instead of a phoenix, he would be thinking of something *different*. And this, in spite of the fact that there are in nature neither phoenixes nor dragons.

So we may begin by dividing experiences into those which have only psychical qualities and do not have epistemological objects and those which have epistemological objects. The former will be called *pure feelings* and the latter *cognitions*. Cognitions may have psychical qualities as well as epistemological objects. Some of them certainly do and perhaps all of them do.

Pure feelings cannot be either veridical or delusive. If a person says that he is feeling tired or feeling cross he is simply saying *how* he is feeling, and the only possibility of mistake is that he may be lying or may be using the words "hot" or "cross" incorrectly. On the other hand any cognition may be either veridical or delusive. It is veridical if there is an ontological object corresponding to its epistemological object. It is totally delusive if there is no ontological object corresponding even remotely to its epistemological object. It is more or less delusive if there is a corresponding ontological object but it differs in certain respects from the epistemological object. A man who is neither lying nor using words incorrectly may say: "I am seeing a pink rat in my bed"; but the experience which he thus truly and correctly describes may be quite delusive.

3.2. Cognition and emotion

Suppose that a person were to say: "I am having an emotion". There are two questions which it would be reasonable to ask: (i) What *kind* of emotion? (ii) Towards what object? The answer that we should expect to the first question would be: "one of hatred", "one of fear", and so on. The answer which we should expect towards the second would be: "Towards Smith", "Towards a ghost", and so on.

All emotions are epistemologically objective experiences, i.e. they are all cognitions, either veridical or wholly or partly delusive. But they are something more than mere cognitions. An emotion is a cognition which has one or more specific forms of a certain generic kind of psychical quality which we will call *emotional tone*. To be fearing a snake is to be cognising something – correctly or incorrectly – as a snake, and for this cognition to be qualified by fearfulness. In general to be fearing X is to be cognising X fearingly; to be admiring X is to be cognising X admiringly; and so on.

3.3. Cognition, emotion, and desire

There is one pair of emotional qualities which stand out from the rest. These are *desire and aversion*. To desire something is to contemplate a possible future state of affairs desiringly. To feel aversion to it is to contemplate it with aversion. The same kind of emotional tone can also qualify our cognition of an actual present state of affairs, though it is unusual to say that we *desire* what actually exists. We should rather say that we "welcome it" or "acquiesce in it".

The peculiarity of desire and aversion is in their effects on action. If I contemplate a possible future state of affairs desiringly I shall be inclined to act in such a way as to bring it about. If I contemplate it with aversion I shall be inclined to act in such a way as to prevent its being realised. If I contemplate the present state of affairs with acquiescence I shall try to keep it unaltered; if I contemplate it with aversion I shall try to alter it.

So it seems to me that desires are cognitions with a special kind of emotional tone directed at special kinds of objects, viz. possible future states of affairs or present actual states of affairs. And they tend to set us acting in certain ways, with a view to bringing about or preventing the realization of such possibilities or to keeping or altering such present actualities.

3.4. Forms of cognition

Ethics is concerned primarily with human action, volition, and emotion. But since all volitions and emotions are also cognitions, and since action is guided

by cognition, it is necessary to say something about the various forms of cognition. I shall be as brief as I can on this topic.

For our purpose the most important division of cognition is into intuitive, perceptual, and conceptual.

3.41. Intuitive cognition

This may be described as direct acquaintance with particulars. So far as we know, a human being is capable of being acquainted with three and only three kinds of particulars, viz. sensa (i.e. colour-expanses, noises, smells, etc.), his own mental images, and his own experiences. The intuitive cognition of these three kinds of particular may be called *sensing, imaging,* and *reflexive acquaintance.*

Whenever one intuits any particular it always manifests itself to one as having certain qualities, e.g. redness, squeakiness, etc. When one intuits several particulars together they often manifest themselves as standing in certain mutual relations, e.g. one may hear two notes as in harmony or in discord with each other. Now we may, if we choose, specially attend to the qualities and relations which are manifested to us by the particulars which we intuit. This kind of attention may be called *inspection.* When it is applied to the qualities and relations of one's own experiences it is generally called *introspection.* Both inspection of senses and introspection of experiences are rather sophisticated kinds of cognition. They are not much used by ordinary men in their daily life, but they are specially developed by artists, psychologists, and philosophers.

3.42. Perceptual cognition

This may be described as cognition of particulars which seems *prima facie* to be purely intuitive but which is found on more careful consideration to be not wholly intuitive. It involves acquaintance with particulars; but it also involves non-inferential beliefs or quasi-beliefs about particulars, which are based on this acquaintance but go beyond the information which it supplies. The three most important instances of perceptual cognition are sense-perception, reminiscences, and self-perception. In sense-perception we base non-inferential beliefs about the existence and qualities of physical things and events upon our acquaintance with visual, tactical, auditory and other sensa. In reminiscences we have non-inferential beliefs about the occurrence and qualities of past events which we have witnessed on acquaintance with present images. In self-perception a person bases non-inferential beliefs about himself and his doings and sufferings on reflexive acquaintance with certain of his experiences.

In each case the presence of intuition, in the form of sensation or imaging or reflexive acquaintance, and the absence of explicit inference, is liable to

make it seem that the cognition is wholly intuitive. Thus one is inclined to take for granted that in sense-perception one is literally *acquainted with* physical things and events; in memory with events that one has witnessed in the past; and in self-perception with one's self. In each case careful consideration shows that this cannot really be true. Neither a physical object nor a past event nor a self is the sort of object with which one could be acquainted, in the sense in which one is acquainted with a sensum or an image or a present experience.

3.43. Conceptual cognition

This includes all those cognitive processes, such as comparison, abstraction, generalization, inference, etc., which operate with general ideas or abstract concepts. By means of conceptual cognition a person can think of things and people and events and situations which he is not acquainted with and is not perceiving or remembering. We do this by thinking of a certain combination of characteristics which constitute the description of a possible object. We then think of the object as "*a* so-and-so" or as "*the* so-and-so" which answers to this description. E.g. one may think of the two properties of being the first Roman Emperor and being an invader of Britain. And we may believe that there was a person answering to the former description and that he also answered to the latter.

We can *imagine* or *suppose* that there is something answering to a certain description without actually *believing* that there is. We can do so while positively disbelieving that there is. This happens, e.g. when we either compose or understand a fictitious narrative such as a novel or a play.

A great deal of cognition which seems at first sight to be purely perceptual turns out on closer inspection to be partly conceptual. It is very doubtful, e.g., whether a person can ever literally *perceive* another person's mind or his experiences, as distinct from his body and his voice and gesture and facial expressions. It seems probable that all my cognition of other selves is really conceptual, though based upon perception of their bodies and their voices. But a great many of one's statements about other selves and their experiences are so expressed as to suggest that one literally perceives them. E.g. "I could see that he was angry".

We share sensation and sense-perception with animals. It is doubtful whether animals have the power of reminiscence, and very unlikely that they have the power of self-perception. But it is the power of conceptual cognition which distinguishes us most sharply, both for good and for ill, from all other animals. It is the basis of all the control that man has gained over nature, and it is at the same time the necessary condition of all the superstitious fears and practices with which men have tortured themselves and each other throughout the ages. Through lack of conceptual cognition animals cannot

design and build reservoirs to hold water in dry seasons, and so they often perish miserably from thirst. On the other hand animals cannot imagine, as some quite intelligent men have done, that the water supply is controlled by the God Moloch, and that the best way to secure a good rainfall is to burn their first-born children alive in an iron idol of the god. Lack of conceptual cognition prevents animals from being either so wise and beneficent or so fantastically foolish and cruel as men.

4. More detailed account of certain kinds of experience

I shall now give a more detailed account of certain kinds of experience which are specially relevant to ethics. Under this head I shall discuss emotion, pleasure, and volition.

4.1. Emotion

4.11. Emotions and emotional moods

An emotion, such as anger, always has an object – real or imaginary. If one is angry one is angry *with* someone or something. It is, in fact, a cognition of an object, real or imaginary, qualified by some species of emotional tone. But, corresponding to the various kinds of emotions, there are certain experiences called *emotional moods*. E.g. the emotional mood which corresponds to the emotion of anger is crossness. One may feel cross without being angry with anyone or anything, or alarmed without being frightened at anyone or anything. I think that an emotional mood is either a pure feeling or else a cognition with a very vague indeterminate object. E.g. it might be one's cognition of things in general or of the present total state of affairs. The connexion between an emotional mood and the corresponding emotion is this. The pure feeling or the very vague cognition, which is the emotional mood, has the same kind of emotional tone as the determinate cognition which is the emotion.

4.12. Classification of emotions by their cognitive character

Since all emotions are cognitions, we shall expect to find a division among emotions corresponding to the division of cognitions into intuitive, perceptual, and conceptual. I do not think that most purely intuitive cognitions have any marked emotional tone. But then purely intuitive cognitions are very rare in grown persons; intuitive cognitions occur mainly as constituents in perceptual or conceptual cognitions. Perhaps the primitive fear which all babies feel on hearing any loud sudden noise, such as a clap of thunder, would be an example of an emotion which was purely intuitive on the cognitive side.

Perceptions are often strongly toned with emotional qualities. E.g. one may perceive with fear an object which one takes to be a snake, and so on. Almost any emotional quality which can characterise a perceptual cognition can also qualify a conceptual cognition. Thus a human being can fear things or events which he is not perceiving or remembering but is only expecting or believing or feigning to exist. The result is that the emotions which we share with animals are felt by us towards a much wider range of objects.

There are some kinds of emotion which, from the nature of their objects, can be felt *only* by a being who is capable of conceptual cognition. E.g. hope and anxiety can be felt only be a being who can conceive and expect alternative possible future states of affairs. Religious awe can be felt only by a being who can think of the description of a deity and can believe that there is an object answering to this description. And so on.

4.13. Motived and unmotived emotions

One may feel an emotion towards an object without consciously distinguishing any qualities in it with regard to which one could say: "I feel this emotion towards that objects in *respect of* those qualities". E.g. you may just dislike a person without being able to mention any quality in respect of which you dislike him. But very often one can mention certain qualities, which one believes, rightly or wrongly, to be present in the object, in respect of which one feels the emotion towards it. E.g. you may be able to say "I dislike so-and-so for his ugly voice and his bad manners". To dislike a person in respect of certain qualities, real or imaginary, is a more complex experience than to dislike him for no assignable reason. Presumably all the emotions of animals are of the second kind; whilst many human emotions are certainly of the first kind.

I will now try to analyse these notions rather more carefully. Suppose that a person's emotion E towards an object O *appears to him* to be caused by his knowledge or belief that O has a certain quality Q. Then I shall say that this emotion is *ostensibly motived*. And I shall say that Q is the *ostensible motivating quality*. Suppose that this person's emotion E towards O *really* is caused by his knowledge or belief that O has a certain quality Q. Then I shall say that this emotion is *actually motived*. And I shall say that Q is the *actual motivating quality*. Suppose that this person's emotion E towards O *does not appear* to him to be caused by any knowledge or belief that he has about the qualities of O. Then I shall say that this emotion is *ostensibly unmotived*. Suppose, lastly, that the emotion *really* is not caused by any knowledge or beliefs that this person has about the qualities of O. Then I shall say that the emotion is *actually unmotived*.

We must now notice the following possibilities of mistake.

(1) An ostensibly motived emotion may be really unmotived. E.g. I may

think that my dislike of Smith is caused by my knowledge that he is an atheist. But really it may be caused, not by this or any other knowledge or beliefs that I have about him, but by some peculiarity in his voice or appearance which I have never explicitly noticed but which rouses my dislike through some unpleasant association which it has for me.

(2) An ostensibly motived emotion may be actually motived, but the actual motivating quality may differ from the ostensible motivating quality. E.g. I may think that my dislike of Smith is caused by my knowledge that he is an atheist. But really it may be caused, not by this, but my belief that he is a Communist or my knowledge that he is a Jew.

(3) Even if an ostensibly motived emotion is actually motived, and if the ostensible motivating quality is the same as the actual motivating quality, it may be that the object does not really possess that quality. E.g. I may think that my dislike of Smith is caused by my belief that he is a Communist, and it may in fact be caused by that belief. But the belief may be false; Smith may really be a Conservative.

(4) An ostensibly unmotived emotion may be actually motived. This can happen in two ways. (i) I may have a number of conscious beliefs and bits of knowledge about Smith's qualities and I may think that none of them causes my dislike of him. But I may be mistaken. It may be that one or other of them does cause my dislike of him. (ii) Even if I am correct in this opinion it may be that I have certain unconscious beliefs or bits of knowledge about Smith, i.e. some which exist only in a dispositional form or which for some reason I fail to notice. And it may be that one or other of these is the cause of my dislike of Smith.

An emotion which starts by being actually unmotived will very often generate beliefs about the qualities of its object. It may thus become an ostensibly and even an actually motived emotion. We shall begin to believe that the object has the sort of qualities which generally evoke that kind of emotion. Then we may begin to think that we are caused to feel this emotion by our knowledge that the object has these qualities. And eventually our belief that it has these qualities may become at least a part-cause maintaining and perhaps heightening the emotion which we feel towards it. E.g. one may start with an unmotived emotion of love towards a person. This may in fact be evoked by some very obscure and quite unrecognised bodily or mental qualities in him. We shall then be very liable to believe that he is particularly beautiful or witty or virtuous. We may then think that we love him because we are aware of these properties in him. And eventually our love for him may in fact be maintained partly or wholly by this belief about his properties. I take it that this is at any rate part of what is meant by the word "rationalization".

Beliefs generated in this way are often false, but they are also quite often

true. One may begin with an unmotived distrust of a person. This may generate the belief that he is dishonest, and we may often find in the end that he really is dishonest. On the other hand, an emotional mood, such as crossness, may be due to purely internal causes such as a disordered liver. Once started it is very liable to crystallise into the corresponding emotion, viz. anger, towards the first suitable object which happens to be available. And then it is liable to generate quite false beliefs about that object. Jealousy is the stock example of an emotion which is specially liable to generate false beliefs about its objects and thus provide itself with motives.

It seems to me that, when a belief is generated by an emotion, one has usually a suspicion at the back of one's mind that it will not bear critical inspection. We tend to refuse to inspect such beliefs critically ourselves, and to feel resentment if other people attempt to do so. In fact, beliefs that are generated by emotion are usually themselves emotionally-toned beliefs.

Emotions which are *conceptual* on the cognitive side, i.e. emotionally toned beliefs, expectations, imaginations etc., are, I think, generally motived. If we think of an object which we are not perceiving and perhaps could not perceive, we must do so by thinking of it as the possessor of such and such qualities or as a term standing in such and such relations. And if one's cognition of an object is emotionally toned, the emotion will generally be felt in respect of some of these qualities and relations. Compare, e.g. our emotions towards Charles I with those of a person, like Cromwell, who had actually met him. *We* can cognise Charles I only conceptually, viz. by thinking of him as a person who had such and such qualities and relations and did and suffered such and such things. If we feel emtions towards him, they must be motived by our beliefs about his qualities and relations. But the emotions which Cromwell, who had actually met Charles I, felt towards him might have been evoked by certain peculiarities of his personal appearance or his voice or manner which Cromwell had never explicitly noticed. So, some of Cromwell's emotions towards Charles I might have been unmotived, even if they were ostensibly motived; whilst all our emotions towards Charles I are both ostensibly and actually motived.

4.14. Misplaced emotion

An emotion may be said to be misplaced if either (i) it is felt towards an object which is believed to exist but does not really do so, or (ii) it is felt towards an object which really does exist in respect of qualities which do not really belong to it. In the first case it may be said to be totally misplaced, in the second partially misplaced.

Let us first consider emotions which are perceptions. A perception, or at any rate a *quasi*-perception, may be completely hallucinatory, as e.g. a dream. In a dream I may have an hallucinatory ostensible perception of a

man chasing me with a revolver, and this may be strongly toned with fear. The fear is then totally misplaced.

Again, a perception may not be hallucinatory but it may be largely delusive. There may be a certain physical object corresponding to my perception, but I may misperceiving it to a considerable degree. E.g. I may perceive a certain physical object which is in fact a tree of a curious shape in twilight. I may misperceive this as a man lying in wait for me. My perception will then be toned with fear, but the fear will be misplaced. If I perceived the object correctly as a tree, I should not perceive it with fear. Of course there *are* real qualities in the tree and its surroundings which cause me to mistake it for a man lying in wait. There are certain shapes, certain arrangements of light and shade and so on. These real qualities gave rise to the false belief that it has certain other qualities which it does not in fact have. And it is my false belief that it has these qualities which is the immediate cause of my perceiving it with fear.

Let us next consider emotions which are conceptual. It is evident that these may be completely misplaced; since there may be nothing answering to the description of a certain object which one believes to exist, and yet the belief may have a strong emotional tone. Completely hallucinatory perceptions are very rare in sane waking healthy persons. But beliefs in the existence of objects which do not in fact exist are, and always have been, quite common among sane waking men. Indeed a large part of the life of humanity has been occupied in feeling strong emotions towards beings who do not exist, e.g. the gods Jupiter or Mars; or towards beings who do exist, e.g. Hitler or Stalin, in respect of qualities which they do not possess. We must notice that all emotions which are felt towards other people in respect of their mental or moral qualities must be in part conceptual on the cognitive side. For we cannot literally *perceive* another person's mind or his disposition, or his motives, or his experiences. We can only *conceive* them, and we are very liable to be mistaken in our beliefs about them and thus to have misplaced emotions.

4.141. Appropriate and inappropriate emotion

As we have seen, there are two aspects to any emotion. In its cognitive aspect, it is directed towards a certain object, real or imaginary, which is cognised, correctly or incorrectly, as having certain qualities and standing in certain relationships. In its affective aspect, it has an emotional quality of a certain kind and of a certain degree of intensity. Now some kinds of emotional quality are *fitting* and others are *unfitting* to a given kind of epistemological object. It is appropriate to cognise what one takes to be a *threatening* object with some degree of *fear*. It is *inappropriate* to cognise what we take to be a fellow-man in *undeserved pain or distress* with *satisfaction or with amusement*. Then, again, an emotion, which is fitting in *kind* to its epistemological

object, may be unfitting in *degree*, i.e. inordinate.

A degree of fear which would be appropriate to what one took to be a mad bull would be inappropriate to what one took to be an angry cow. It should be noticed that an emotion which is misplaced may be appropriate to its object as that is misperceived or mistakenly believed to be. If a short-sighted person takes what is in fact a harmless but excited cow for a mad bull, it is appropriate for him to cognise it with a high degree of fear. Conversely, an emotion which is veridical on the cognitive side may be unfitting in kind or inordinate in degree. A woman who panics in the presence of what she correctly takes to be a mouse illustrates this fact. The notion of a certain fittingness or unfittingness, in kind or in degree, between emotional tone and epistemological object, is plainly of the utmost importance to ethics and to aesthetics. I think that it still awaits an adequate analysis.

4.15. First-hand and second-hand emotion

This is an important distinction which arises in connexion with conceptual emotion. Let us take as an example the emotion of religious awe towards God. This would be a *first-hand* emotion if and only if the person who felt it was really thinking at the time of the qualities and relations which constitute a description of God – e.g. a being of infinite power who has created and governs everything – and was really believing that there is something answering to this description.

But most concepts which have been fairly often used have had names attached to them, and it is possible to use the names consistently and correctly without thinking of the characteristics which they connote. Now in many cases a certain name has become associated through early training with a certain kind of emotional mood. If we now hear or see or use that name, the associated emotional mood tends to be excited. We shall then tend to think that we are feeling a certain emotion towards a certain object in respect of certain qualities, when really we are not thinking of the object or its qualities at all. This is what I call "*second-hand*" emotion.

Many words and symbols, particularly those associated with religion, morality, and politics, are almost devoid of cognitive meaning for most people at most times. But they have become extremely powerful stimulants of second-hand emotion. It is obvious that a great deal of the emotions that we feel are second-hand, and there is always a likelihood of emotions which were first-hand becoming second-hand. A typical example is the sorrow felt by a bereaved person. It begins by being first-hand, and in the course of nature it tends to fade away after a while. But often the bereaved person cannot face this fact, and so pumps up a second-hand emotion to replace the vanished first-hand one.

It is important to remember, however, that nearly all second-hand emotion

depends on the existence of a corresponding first-hand emotion in *someone* at *some* time in the past. If no one had ever believed in God with a first-hand emotion of awe, it is unlikely that anyone would now have a second-hand emotion of awe produced by the word "God". But the first-hand ancestor of a second-hand emotion may be a very long way back in the past history of an individual or of a race.

4.16. Pure and mixed emotion

I think we may assume that there is a certain fairly limited number of primary species of emotional tone, just as there is a limited number of primary colours, and that we are born with dispositions corresponding to each of them. Let us call these "primary" emotional dispositions. I should think that the emotional tones of fear and of anger are certainly primary, and that the corresponding emotional dispositions are certainly innate. Probably some innate emotional dispositions do not come into action until certain stages of development, e.g. puberty, have been reached.

Now these primary emotional dispositions are either very specialised or very generalised in respect of the stimuli which originally evoke them. E.g. the disposition to feel fear seems to be excited at first only by sudden loud noises and by the experience of falling. So the original stimulus is here very specialised. The disposition to feel anger, on the other hand, is aroused from the first by the thwarting of *any* impulse. So here the original stimulus is highly generalised. In the course of experience these primary emotional dispositions become generalised or specialised. E.g. we acquire the disposition to fear snakes, to fear policemen, and to fear ghosts, in addition to fearing sudden noises and falls. Or, again, we acquire the disposition to feel angry at injustice done to other people beside being angry at being thwarted ourselves.

I do not think that a given kind of emotional tone remains completely unaltered in quality as the objects of the emotion become extended and made more subtle. No doubt there is a qualitative likeness, e.g., between fearing a sudden noise, fearing an interview with a headmaster, and fearing God. They all resemble each other in a specific way in which, e.g., the experiences of fearing a sudden noise and being angry at a sudden blow do not resemble each other. But there is a difference in emotional quality between these various experiences of fear. This might be compared to differences of *shade* between various instances of the same colour, e.g. scarlet, rose-coloured, pink, etc. I think then that we must say that the various primary kinds of emotional tone become differentiated in shade as the experiences which they qualify become more complex and more abstract.

Now suppose that I perceive or think of an object which has several characteristics. In respect of one of them it may excite one emotional disposition, e.g. that of fear; and in respect of another of them it may excite another

emotional disposition, e.g. that of anger. My perception or thought of the object will then be toned with an emotion which is a *blend* of fear and anger.

I think that the best way to conceive of blended emotions is by analogy with blended colours, such as purple. Any shade of purple resembled pure blue to some extent and pure red to some extent, and there is a continuous series of possible shades of purple stretching from pure blue at one end to pure red at the other. A sensation of purple is produced when the same part of the retina is affected simultaneously by a stimulus which would produce a sensation of pure red and a stimulus which would produce a sensation of pure blue if it acted by itself. In the same way there are many different shades of blended emotional quality, stretching from pure fear without anger to pure anger without fear. The particular shade of blended emotion which is felt on any particular occasion will presumably depend on the relative degree of excitement of the various primary emotional dispositions, e.g. the fear-disposition and the anger-disposition.

The following remarks are worth making about blending. (i) It may be that certain primary emotional dispositions, e.g. those of anger and of fear, are directly linked from the first. Others become linked only indirectly in the course of experience. (ii) Probably a grown man hardly ever has an experience with a *pure* primary emotional tone. The notions of the pure primary emotions, like the notions of the pure primary colours, are ideal limits. (iii) Whilst some of the primary emotional qualities blend readily with each other, as do the colours red and blue; it may be that others will not blend. The latter would have to each other the kind of opposition which there is between complementary colours such as red and green or blue and yellow.

Lastly it is worth while to notice that there are certain emotional adjectives, such as "sad" and "cheerful" which apply to a total phase of experience as a whole rather than to any part of it. They may be compared to adjectives like "bright" and "dull" as applied to the visual field as a whole. We might call such qualities of complex wholes "pattern-qualities". They *depend* on the qualities and relations of the constituent parts of the whole, e.g. on the emotional tones of the various experiences included in the total phase of experience. But they are not *reducible to* these. Very often superficial introspection will catch the emotional pattern-quality of the phase as a whole, and will fail to notice the emotional qualities of the constituent experiences. One may notice that one feels sad or elated without knowing why. More elaborate introspection will reveal the emotional qualities of the constituent experiences, but it may lose sight of the emotional pattern-quality of the whole.

4.17. Sentiments

Supose that a certain object has been repeatedly perceived or thought of by a person. Suppose that it is complex in its nature and structure, and that he has

perceived or thought of it in many different contexts or various occasions. These various cognitions of the object will have produced a very complex trace, i.e. a very complex dispositional idea of the object. Suppose that this trace has become associated with traces of certain names, phrases, or symbols which have often been heard or spoken or seen in intimate connexion with perceiving or thinking of this object. Lastly, let us suppose that on many occasions when the object has been perceived or thought of, strong emotions have been felt towards it. When it was perceived or thought of in certain situations, or when certain aspects of it were attended to, the cognition had the emotional tone *X*. When it was perceived or thought of in certain other situations, or when certain other aspects of it were attended to, the cognitions had the emotional tone *Y*. And so on. The result is that the dispositions corresponding to the emotions *X*, *Y*, etc. will have become associated with the dispositional idea of the object. Henceforth anything that excites the dispositional idea of the object, e.g. perceiving it, thinking of it, or perceiving or thinking of any word or symbol connected with it, will tend to excite all these emotional dispositions. We sum up all this by saying that a "*sentiment* has been formed about the object".

When a sentiment is aroused the emotional tone of the experience will be some shade of a blended tone. The particular shade will vary according to the past conditions under which the sentiment was formed and the present circumstances which are exciting it. It is of course possible that some of the associated emotional dispositions are such that the corresponding emotional qualities will not blend. E.g. it may be that fear and contempt will not blend, and yet that a certain object has come to arouse both of them. In that case we may have the two kinds of emotional tone rapidly alternating with each other. Or we may distinguish certain characteristics in the object, and at the same time feel fear of it in respect of some of them and contempt of it in respect of others.

Sometimes the sentiment gets concentrated on one particular symbol for the object instead of on the object itself. Or it may become concentrated on one particular part of the object instead of the object as a whole. We then say that that symbol or that part of the object has become a *fetish*. Fetishism is a fairly common aberration of sexual emotion.

Presumably there are no *innate* sentiments. But there are certain sentiments which practically every human being will inevitably acquire. One is a sentiment about himself and his own powers, defects, achievements and failures. Another is a sentiment about his parents and parent-substitutes such as nurses, and about the members of his household in general. Another is a sentiment about the social and political groups, other than his household, of which he is a member. Everyone is a self; everyone had parents and started as a helpless infant kept alive and trained by them or by substitutes for them;

and everyone grows up as a member of several social groups. It is therefore inevitable that reflexive, filial, family, and social sentiments should arise in practically everyone.

Certain reflexive emotions, such as remorse, self-approval, and so one are obviously very important to others. It is worth while to notice that we have emotions and sentiments which are not only reflexive but are about our own emotions and sentiments. These may be described as "*second-order reflexive emotions*".

E.g. a person may be ashamed of being afraid, or afraid of being ashamed, or afraid of being afraid, or ashamed of being ashamed, and so on. Or again a person may feel angry with himself in respect of his sentiment of love for a person whom he knows to be worthless and unfaithful to him. This is another example of the extreme complexity of human life and experience as compared with anything that occurs in animals.

There are two points worth noticing about the *names* which are used in ordinary life for various emotions and sentiments. (i) We have an enormous number of such names, e.g. envy, jealousy, contempt, awe, etc. But we must not rashly assume that there are different kinds of emotional quality corresponding to each of these. Names are given to emotions and sentiments partly in respect of their emotional quality and partly in respect of their objects. Two emotions or sentiments which have the same quality may have different names because they have different kinds of objects. E.g. "envy" is the name of a certain kind of emotion called forth by witnessing another person getting what one wants oneself. "Jealousy" seems to be the name of an emotion of the same kind when what one wants and what the other person gets is the affection of some third person. I do not say that there is no shade of difference in the emotional quality in the two cases; but the different names are certainly given in respect of the different kinds of *object*, and not in respect of the difference, if any, in the state of the emotional quality. (ii) Because a certain sentiment is distinguished from others by a certain name, e.g. "love", we must not rashly assume that the blended emotion connected with it contains any emotional constituent that is peculiar to it. It is certain, e.g., that the blended emotion which one feels when one is in love with a person has several factors which occur in other blended emotions. And it is quite possible that there may be no single factor in this blended emotion which does not also occur in some other blended emotion. It may be that what distinguishes this emotion from all others is some pattern-quality due to the particular proportion in which emotional factors, each of which occurs elsewhere, are here combined.

Even when the blended emotion connected with a certain sentiment does have a peculiar emotional constituent, it may be that this by itself is very trivial. Suppose, e.g., that sexual emotion is a peculiar constituent of the

blended emotion connected with the sentiment of sexual love. And suppose that every other constituent of this blended emotion can occur as a constituent of some other blended emotion. It might still be the case that these other emotional factors, though less characteristic of erotic emotion, when taken *severally*, are yet *collectively* essential. Mere sexual emotion, if it should occur unblended with these other constituents, would not constitute that peculiar emotion which is felt when one is in love with a person and when this sentiment is stimulated. People are rather liable to give the same name to (a) the blended emotion connected with a certain sentiment and (b) any emotional quality which is a peculiar constituent of that blended emotion. E.g. the name "erotic emotion" might be given either to the blended emotion which is felt by a person towards another whom he is in love with, or to the purely sexual emotion which is perhaps the only constituent peculiar to that blended emotion. If this happens one is certain to be landed sooner or later in tiresome verbal controversies.

4.2. Pleasure and unpleasure; happiness and unhappiness

I shall now try to clear up these notions and to draw such distinctions as seem to be important. And I shall consider the relations between pleasure and happiness, unpleasure and unhappiness.

4.21. "Unpleasant" and "painful"

The contrary opposite to "pleasant" is not "painful" but "unpleasant". Any experience that is painful, e.g. a twinge of toothache, is unpleasant; but there are plenty of experiences which are unpleasant without being painful. E.g. the experiences of tasting castor-oil or quinine or of smelling sulphuretted hydrogen, are unpleasant but they are not painful. I think that the word "nasty" is commonly used to mean unpleasant but not painful. Suppose we give the name "hedonic tone" to the determinable characteristic of which pleasantness and unpleasantness are the two immediate determinates. Then unpleasantness is itself a determinable, and the two determinates immediately under it are painfulness and nastiness.

At the level of feeling or sensation the experiences which are called "painful" are certain kinds of organic sensation, such as those of burning, of toothache, of rheumatism in muscles or joints, and so on. Each of these has its own characteristic kind of sensible quality, but there is plainly a good deal of likeness between them. All of them are susceptible of such descriptions as "dull", "acute", "throbbing", "stabbing", etc.

Physiologists tell us that there are definite "pain-spots" distributed about the body, i.e. spots which, when stimulated, give rise to sensations of this kind. They have concluded that there is a specific kind of sensation which

they call "pain-sensation". This must be put alongside of other special sensations, such as these of colour, sound, etc. Apparently the physiologists do not think it necessary to postulate a specific kind of sensation of the opposite kind to pain-sensation.

I think that the facts can be stated as follows. There is a certain kind of sensible quality which makes any sensation that has it unpleasant to all normal persons in almost all circumstances and even when it is present in a low degree. Any sensation which has this sensible quality and is thereby made unpleasant is called, not merely "unpleasant" or "nasty", but also "painful". Most sensations are not unpleasant unless their qualities are present in a very intense degree, e.g. a dazzling light or a deafening noise, or in some rather special determinate form, e.g. certain kinds of squeaky noise. The sensible quality of pain, if intense and prolonged enough, is capable of giving to an experience a degree of unpleasantness which no other sensible quality can give.

There are two remarks which seem worth making here. (i) I am inclined to think that there is a kind of sensible quality, analogous but opposite to the pain-quality. This makes sensations which have it pleasant for all normal persons under almost all circumstances and even when it is present to a low degree. The most obvious examples of sensations with this quality are those due to stimulation of the sexual organs. Experiences with this kind of sensible quality stand out from those which are merely "pleasant" or "nice" in much the same way as toothache, etc. stand out from experiences which are merely "unpleasant" or "nasty". I propose to call this peculiar sensible quality the "orgiastic quality". So we can divide pleasant sensations into nice and orgiastic, as we divided unpleasant sensations into nasty and painful. Now pain-spots are distributed all over the body, and therefore almost any part of the body will give rise to painful sensations if it is stimulated by great heat or strong pressure. But the sources of orgiastic sensation are much more definitely localised, and only a few parts of the body will give rise to them however they may be stimulated. (ii) There is a well-known abnormal condition which is called "algolagnia", i.e. taking an intense pleasure in one's own pain. It seems to me that this might arise in two quite different ways. (a) The patient might be hedonically normal but sensitively abnormal. To say that he was "hedonically normal" would mean that he finds painful sensations unpleasant and orgiastic sensations pleasant, just as other men do. To call him "sensitively abnormal" would mean that the kind of stimulus which produces in most men sensations with the pain-quality produce in him sensations with the orgiastic quality. (b) The patient might be sensitively normal and hedonically abnormal. To call him "sensitively normal" would mean that stimuli which produce in us sensations with the pain-quality produce sensations with the pain-quality in him also. To call him "hedonically

abnormal'' would mean that the pain-quality makes a sensation pleasant for him just as the orgiastic quality makes a sensation pleasant for normal people.

Presumably sensitive abnormality would be due to physiological causes; whilst hedonic abnormality might be due to psychological causes. A psychoanalyst might be able to treat successfully an algolagniac patient of the second kind, but he could do nothing for one of the first kind.

So far I have talked only of sensations. I doubt whether any distinction between nice and orgiastic or nasty and painful can be drawn in the case of other kinds of pleasant or unpleasant experiences. We do sometimes call a perception, e.g. hearing an ugly bit of music or seeing a bad picture, ''painful''. But this is only an exaggerated way of saying that it is highly unpleasant. Again, you might say that the memory of a past incident or the expectation of a future interview was ''painful''. But I think it is plain that this means no more than to call it highly unpleasant.

4.22. Dispositional and occurrent senses of the words

All words like ''pleasant'', ''unpleasant'', ''nice'', ''nasty'', ''painful'' etc. are used in two different senses. One is fundamental and may be called the *occurrent* sense; the other is derived from it, and may be called the *dispositional* sense.

We talk of the experience of tasting chocolate as ''pleasant'', and we also talk of chocolate itself as ''pleasant'' or ''nice''. It is plain that the fundamental sense is the former. In that sense it applies to actual experiences and to nothing else. This is the *occurrent* sense. When we say that chocolate is pleasant we mean that any bit of chocolate will give pleasant sensations of taste to most human beings if they eat it. This is the *dispositional* sense.

When we say that chocolate is pleasant and castor-oil unpleasant we are not only using the words in the dispositional sense but are also speaking elliptically. We mean that chocolate is pleasant to *taste*, and that castor-oil is unpleasant to *taste* and to *smell*. Chocolate is rather unpleasant to see or to touch if it is warm and sticky, and castor-oil is not at all unpleasant to see.

We have seen that nothing but experiences can be occurrently pleasant or unpleasant. Can an experience be dispositionally pleasant or unpleasant? There is a sense in which this is possible. Suppose I have had an experience and that it has left a trace, so that I can remember it or reflect on it on many future occasions. The memory or thought of this experience may be itself a pleasant or unpleasant experience. Suppose that on all or most occasions when I remember a certain past experience the memory is a pleasant experience. Then we could say that the original experience is ''dispositionally pleasant''. To speak more accurately we should say that it is dispositionally pleasant to *dwell upon*. A similar account could be given of dispositionally

unpleasant experiences. It may be remarked that an experience which was occurrently pleasant may be dispositionally unpleasant or vice versa. What was pleasant to experience may be embarrassing to remember and reflect upon.

4.23. Pleasant-making and unpleasant-making characteristics

The hedonic tone of an experience is not an independent property of it. It is always sensible to ask: "What *makes* it pleasant, or what *makes* it unpleasant?" The answer will always be: "Because it has such and such a non-hedonic characteristic", e.g. "Because it has the toothachy quality; because it is the fulfillment of a pre-existing desire", and so on. Those non-hedonic characteristics of an experience which make it pleasant or unpleasant, in the primary occurrent sense, may be called its "*pleasant-making*" or "*unpleasant-making*" characteristics. If a characteristic is either pleasant-making or unpleasant-making we will call it an "hedonifying" characteristic.

Now two experiences which were precisely alike in their non-hedonic characteristics might not be precisely alike in their hedonic characteristics if they occurred in different persons or in the same person at different times in his life. A non-hedonic characteristic of an experience is not *absolutely* pleasant-making or unpleasant-making; it is so relatively to what we may call the "*pleasure-taking*" and "*unpleasure-taking*" disposition of the person who has the experience. E.g. in a young child the rattling quality of the noise made by a rattle is pleasant-making and the taste and smell of a lighted cigar would be unpleasant-making. In the child's father, or the child himself when he has grown up, the auditory sensations produced by rattles are unpleasant and the characteristic taste and smell of a lighted cigar are pleasant. This is presumably because the pleasure-taking dispositions of the child and his father, or the individual as a child and as a grown man, in respect of these sensible qualities, are different.

It should be noticed that when we say that two persons have different tastes, or that a person's tastes have changed in course of time, there are two alternative explanations possible. (1) It might be that the pleasure-taking dispositions are the same in both cases but that similar stimuli produces sensations with different sensible qualities. E.g. a person might enjoy the taste of cream when he was well and find it very nasty when he was bilious. The probability here is that the cream actually produces different sensations of taste under the two conditions. (2) It might be that similar stimuli produce similar sensations in both cases, but that the pleasure-taking dispositions are different. This is probably the case in the example of losing a taste for rattles and acquiring a taste for cigars. Of course both factors may enter. It may be that similar stimuli produce somewhat dissimilar sensations, and also that the pleasure-taking dispositions are somewhat different. It seems likely that both these factors enter when a person's tastes change as he gets older.

It is important to notice that there is no *logical* necessity for the pain-quality to be unpleasant-making and the orgiastic quality to be pleasant-making. It is simply an empirical generalization about human pleasure-taking and unpleasure-taking dispositions that the vast majority of the human race find toothache extremely unpleasant and being tickled extremely pleasant. There might be a whole race of creatures for whom the "achiness" of the sensation which we call toothache made those sensations intensely pleasant.

4.24. Classification of pleasures and unpleasures

(1) We can begin by dividing hedonically toned experiences into those which are and those which are not *purely sensuous*. A purely sensuous pleasure or unpleasure is a pure feeling or a sensation which derives its hedonic tone entirely from its sensible qualities, e.g. its throbbingness, its sweetness, its squeakiness, and so on. What are commonly called "bodily" pleasures and pains are a certain sub-class of purely sensuous pleasures and unpleasures. They are those feelings or sensations which are made pleasant by their orgiastic quality or unpleasant by their pain-quality.

(2) Sometimes a complex sensation, e.g. of sight or hearing, is made pleasant or unpleasant by certain relations between the various component sensa which together make up its complex sensum. An example would be the unpleasant experience of hearing together or in close succession two sounds which are out of tune. Another would be the pleasant experience of sensing by means of a kaleidoscope certain patterns of adjoined visual sensa which harmonise and contrast in colour. Here the pleasantness or unpleasantness depends partly on the qualities of the component sensa and partly on their mutual relations. Such pleasures and unpleasures are *sensuous*, but are *not purely* sensuous. I will call them *sensuously aesthetic*.

(3) We now come to the level of *sense-perception* as contrasted with mere sensation. At this level we do not dwell on the qualities and relations of the sensa which we sense. They function mainly as signs on which we base non-inferential judgments or quasi-judgments about physical things and events, their qualities, relations, etc. Now the perception of physical things as having certain qualities and relations is pleasant, and the perception of them as having certain others in unpleasant, quite regardless of their probable effects for good or ill on oneself or those in whom one is interested. I shall call such pleasures and unpleasures *perceptually aesthetic*.

The essential difference between sensuously and perceptually aesthetic pleasure and unpleasure is this. Are the hedonifying characteristics of the experiences the qualities and relations of the sensa themselves? Or are they the qualities and relations of the physical things or events which the sensa signify? Probably the pleasures and unpleasures of listening to music are predominantly of the sensuously aesthetic kind. But those of visual experiences

must always contain a large proportion of the perceptually aesthetic kind; for it is almost impossible to prevent visual sensa from being made instruments of perceptual judgment.

(4) There is a certain amount of aesthetically pleasant and unpleasant experience at the conceptual level. To contemplate a geometrical argument in which very far-reaching and unexpected conclusions are shown to follow from very simple premisses is a pleasant experience. To contemplate a clumsy or confused argument is an unpleasant experience. A great deal of the pleasure which certain people get from listening to certain kinds of music is conceptually aesthetic.

(5) Next we come to the pleasures and unpleasures of memory and expectation. These may be called *reflexive*, since they arise in contemplating one's own past and future experiences. They may also be called of *the second order*, in so far as what makes a memory or an expectation pleasant or unpleasant is the pleasantness or unpleasantness of the experience which one remembers or expects.

The expectation of a future pleasant experience is pleasant, and the expectation of a future unpleasant experience is unpleasant. The rules about memory are not so simple. The memory of a past pleasant experience tends as such to be pleasant. But it may easily be a component of an unpleasant experience through the contrast between the happy past and the less happy present. The memory of a past unpleasant experience tends as such to be mildly unpleasant. But it is often a component of a pleasant experience of relief and contrast.

(6) We come next to beliefs about other persons and their experiences. Suppose that there is no special emotional relationship between two persons A and B. Then A's belief that B is happy will tend to be a mildly pleasant experience, and his belief that B is unhappy will tend to be a mildly unpleasant experience. But certain pre-existing sentiments in A towards B will increase, diminish, or reverse this. Suppose that A loves B. Then his pleasure in believing that B is happy and his unpleasure in believing that B is unhappy will be greatly intensified. Suppose that A hates B or is jealous of him or is afraid of him. Then his belief that B is happy may be unpleasant, and his belief that B is unhappy may be pleasant. The pleasures and unpleasures which come under the present head are of the *second order*, since they consist in taking pleasure or unpleasure in pleasure or unpleasure. But they are *not reflexive*, since the experiences in which A takes pleasure or unpleasure are not in himself but in B.

(7) Certain emotions, such as fear, are unpleasant experiences if they are at all strong. Others, such as hope, are definitely pleasant. The following remarks are worth making in this connexion: (i) All strong emotions are accompanied by and fused with certain characteristic organic sensations, e.g.

heart-beat, nausea, hotness or coldness, sweating, etc. These are generally sensuous pleasures or unpleasures. (ii) A total experience, which contains an unpleasant emotion as a constituent, may be pleasant in spite of, or even because of, the unpleasantness of the constituent. The pleasantness of any dangerous sport, such as rock-climbing, depends in part on the presence in the background of a slightly unpleasant component of fear. This is even more obvious in the pleasant experience of reading a really terrifying ghost-story. (iii) The occurrence of a certain emotion which is itself unpleasant may make other simultaneous experiences pleasant or vice versa. E.g. to feel jealous is in itself an unpleasant experience. But if *A* believes *B* to be unhappy at the time when he is feeling jealous of *B*, his unpleasant emotion of jealousy may make his belief a pleasant experience.

(8) The fulfilment of any desire is, as such and for the moment at least, a pleasant experience. The intensity of the pleasure of fulfilment increases with the intensity of the desire. The prolonged frustration of any desire is unpleasant; and the disappointment of a desire is unpleasant, especially if it is unexpected and is believed to be final. Here again the intensity of the unpleasure of frustration or disappointment increases with the intensity of the desire.

Of course later experience or reflection may make one on the whole sorry that a certain desire was fulfilled or glad that it was frustrated. Alas we often find that the desired object, when gained, is not so satisfactory as we expected it to be. The mere fact that we had to wait for it gave it a fictitious value. So the pleasure of fulfilment is often followed immediately by the unpleasant experience of disappointed hopeful expectation.

(9) The state of desiring is not in itself an unpleasant experience, provided that what is desired is believed to be capable of attainment. No doubt it implies that one is dissatisfied in a certain respect with the present state of affairs and wants to effect a certain change in it. But that dissatisfaction may be only partial. And the total experience, which includes pleasant hopeful expectation and the initiation and carrying-on of an active process, is often, I think, predominantly pleasant.

(10) One of the most important kinds of pleasant and unpleasant experience is that associated with active process, whether bodily or mental. The experience of efficiently performing any action is predominantly pleasant; the experience of doing it inefficiently is predominantly unpleasant. If there are obstacles and difficulties which we gradually overcome, and we feel ourselves to be moving in spite of them towards a desired end, the experience may become extremely pleasant. If, on the other hand, the obstacles occur too often or are found to be insuperable, the experience is extremely unpleasant. No doubt the appearance of each obstacle is greeted with a temporary feeling of annoyance which is itself unpleasant. But the satisfaction of overcoming

an obstacle is proportional to its apparent magnitude. These unpleasant components enhance the pleasantness of the total experience by preventing it from becoming insipid. They act like pepper and other hot or sharp condiments in cookery.

Corresponding to any given degree of skill and efficiency in a person is a certain optimum amount of obstacles. If less than this is provided the activity becomes too easy and the experience is boring. If more than this is provided the activity is too much checked and frustrated, the agent begins to lose hope and interest, and the experience is on the whole unpleasant.

There are three remarks to be made about the pleasure of successful activity in face of obstacles. (i) I think that they are probably much the most important that human beings enjoy. They are indeed less intense than certain purely sensuous pleasures. But they last much longer; they are much more capable of variation and development; they do not lead to satiety and reaction; and they are felt to be more "worth while", i.e. to be a source of legitimate self-satisfaction. (ii) The high value which we set on them is shown by the fact that men have always invented and practised games and contests of bodily skill and endurance, puzzles, and games of mental skill like chess. In a game like football we deliberately set an end before ourselves and deliberately arrange obstacles in the form of an opposing team and restrictive rules. When we begin to play the end does not seem very important, and the main pleasure is in the activity itself. But the activity and the opposition soon make us, for the time being, attach great importance to gaining the end. The result is that, on the conclusion of the game, we may have a strongly pleasant experience of fulfilled desire or a strongly unpleasant experience of frustrated desire, according to whether our own side wins or loses. This is an instance of a general rule. The process of actively pursuing an end against obstacles makes one desire the end more strongly. And this in turn makes the experience of activity more intense and in general more pleasant. (iii) The pleasures and unpleasures of activity in presence of obstacles are involved in all the more elaborate kinds of pleasant or unpleasant aesthetic experiences. Suppose that the harmonies of a piece of music are very obvious, or that the charm of a picture or poem "hits one in the eye". Then the experience of listening to it or seeing it or reading it becomes insipid even though it may have a good deal of sensuous or perceptual aesthetic pleasantness. The activity of noticing harmonies which are just not discords, and of discovering subtle points of beauty in pictures or poems, adds greatly to the pleasantness of the experience provided that the difficulties and obscurities are not too severe. Here, again, there is a certain optimum amount of difficulty for any person at a given stage of his aesthetic development. Music or painting or poetry which present much less difficulty than this is for him insipid or "chocolate-boxy"; that which presents much more difficulty is for him a mere cacophonous set of noises or

a meaningless collection of coloured patches or words.

4.241. Summary of the classification

I think that the list of ten kinds of pleasant and unpleasant experiences which I have given covers all the most important cases. We can extract from it the following classification. (1) The first division is into those experiences which are made pleasant or unpleasant by being the *fulfilment or the frustration of a pre-existing desire*, and those which are made pleasant or unpleasant in some other way. We will call the former pleasures and unpleasures of *fulfilment or frustration*. (2) The second class can then be subdivided into the pleasures and unpleasures of *action* and the *passive* pleasures and unpleasures. (3) The latter can be subdivided into the pleasures and unpleasures of *emotion*, of *contemplation*, and of *sensation*. (4) The pleasures and unpleasures of contemplation can be divided into those of the *first-order*, viz. the perceptually and conceptually aesthetic pleasures and unpleasures, and those of the *second-order*. Those of the second-order can be divided into the reflexive, viz. the pleasures and unpleasures of memory and expectation, and the *non-reflexive*, viz. the pleasures and unpleasures of contemplating the happiness or misery of others. These latter, as we have seen, are very much affected by the existence and the excitement of certain pre-existing emotional dispositions or sentiments. (5) The pleasures and unpleasures of sensation can be subdivided into the *sensuously aesthetic* and the *purely sensuous*. Finally, the purely sensuous pleasures can be divided into experiences which are *nice*, e.g. the sensation of smelling a rose or tasting chocolate, and those which are *orgiastic*. And the purely sensuous unpleasures can be divided into experiences which are *nasty*, e.g. the sensation of tasting castor-oil, and those which are *painful*, e.g. the sensation of being scalded.

Of course these various kinds of pleasure and unpleasure are inextricably mixed up with each other in our total experiences. E.g. desire, action, emotion, and organic sensation are intimately linked. Action is generally started and kept up by desire or emotion. It is accompanied by emotion, expectation, and characteristic organic sensation. It tends to intensify the desire which started it, and it ends in total or partial fulfilment or frustration. The hedonic tone of one's total experience at any time depends on the hedonic tones of its various constituents, but it does not depend on these in any simple way. E.g. you would not necessarily increase the pleasantness of a pleasant total phase of experience by removing all its unpleasant constituents. To do this, if it were possible, might merely render it insipid and boring.

4.25. Conditions of pleasure and unpleasure

(1) Some pleasures depend, not on any intrinsic pleasant-making quality in the experience, but on the hedonic difference between the present state and an

immediately previous state. Suppose that one has been having an unpleasant experience and that this is immediately followed by one which is in itself less unpleasant or neutral or positively pleasant. Then the transition from more to less pleasant, or from unpleasant to neutral or to pleasant, is felt as pleasant. Similarly a transition from a neutral to a pleasant state, or from a less to a more pleasant state, is felt as pleasant. And in such cases, so long as the later state is contrasted in memory with the earlier and hedonically inferior state one has a pleasant experience of contrast. Obvious examples of this are the pleasure experienced by a patient who is recovering from an illness or by a person who is resting after violent exertion.

In the same way a transition from a pleasant to a less pleasant or a neutral state, or from a neutral to an unpleasant state, or from a less to a more unpleasant state, is felt as unpleasant. And as long as the later state is contrasted in memory with the earlier and hedonically superior state one has an unpleasant experience of contrast. We might call these the pleasures and unpleasures of *hedonic transition* and *hedonic contrast*. I think it is true to say that one's memory both of the pleasantness and the unpleasantness of past experiences fades very quickly. E.g. a patient on recovering soon ceases to be able to remember in detail what it felt like when he was seriously ill. So the pleasures and unpleasures of hedonic contrast tend to be evanescent.

(2) Some purely sensuous pleasures arise only through a process of putting into the body substances of which it has become depleted. The most obvious examples are the pleasures of eating when hungry or drinking when thirsty. The sensations accompanying such processes of restoration are at first extremely pleasant; but their pleasantness quickly diminishes as the bodily depletion is progressively made up. And this kind of pleasure cannot again be enjoyed until the body is again depleted. We may call these *pleasures of restoration*.

The bodily states of depletion, e.g. lack of food or of drink, produces characteristic sensations and desires. These are not unpleasant in moderation, though they can become intensely painful if the depletion continues beyond a certain point. The purely sensuous pleasures of restoration will always be accompanied by the pleasure of fulfilling a desire, e.g. the desire to eat or to drink. And, if the sensations of hunger or thirst have become unpleasantly intense, there will also be the pleasures of hedonic transition and of contrast as the process of restoration goes on.

(3) Some purely sensuous pleasures arise only through a process of discharging from the body substances with which it has become charged. These substances may be either waste products of processes, such as digestion, or special secretions such as seminal fluid. The sensations accompanying the evacuation of these substances are generally pleasant and often intensely so. But such pleasures are from the nature of the case very short-lived, and they

cannot be enjoyed again until the body has again become replete with the relevant kind of stuff. We may call them *pleasures of evacuation.*

The bodily states of repletion in respect of these substances produce characteristic sensations and desires. The sensations are generally unpleasant, even when not intense, and they become intensely painful if the repletion passes beyond a certain point. The purely sensuous pleasure of evacuation will always be accompanied by the pleasure of fulfilling a desire, which may be very intense even when the sensations arising from repletion are not at all unpleasant.

(4) Suppose that a certain stimulus produces a pleasant sensation in a person. Suppose that the same stimulus continues to act on him without any change in character or in intensity. Then a point will come at which the sensation reaches its maximum pleasantness. If the stimulus be continued after this, the sensation will become progressively less pleasant. And if it be continued beyond a certain period the sensation may become definitely unpleasant. A simple example of this would be a note sounded on some such instrument as a flute. It may well be that the sensible quality of the sensation changes in course of time through the relevant part of the nervous system becoming temporarily exhausted. But, even if it did not, the mere continuance of the same kind of sensation would become first boring and at length positively unpleasant.

Again if a stimulus of a low degree produces a pleasant sensation, the same kind of stimulus in a slightly higher degree will generally produce a slightly more pleasant sensation. But there is an optimum point for a given person and a given kind of stimulus. If the stimulus is made more than this it will produce a less pleasant sensation. And by increasing the intensity of the stimulus enough it will produce a sensation which is positively unpleasant.

The rules about stimuli which produce unpleasant sensations are different. If a stimulus produces an unpleasant sensation, the unpleasantness will generally increase as the same stimulus continues to act for longer and longer. The only exception to this is that the person may become inured to the unpleasant stimulus if it is not too intense.

Again, if a stimulus of low degree produces an unpleasant sensation in a person, a similar stimulus of higher degree will produce a more unpleasant sensation. There is no limit to this except that the person may eventually lose consciousness.

(5) Suppose that a stimulus which would produce a pleasant or an unpleasant sensation initially begins at t_1 and lasts till $t_1 + \tau$. Suppose that it then ceases altogether and starts again after an interval with the same intensity at t_2. Suppose that there is a considerable interval between $t_1 + \tau$, the end of the first period, and t_2, the beginning of the second. Then, in general, the degree of purely sensuous pleasantness or unpleasantness will be roughly the same at

corresponding moments in the two periods. And the same will hold on each subsequent repetition after a sufficient interval. But of course other hedonifying factors may come in beside the quality of the sensation. The first experience will have the characteristic of novelty, which may make it either pleasantly interesting or unpleasantly alarming. Later repetitions will lack this feature. And, if it has been repeated on many occasions, it may acquire the unpleasant-making characteristic of staleness.

(6) It is worth noting what an important part *boredom* plays as a source of unpleasantness in the experience of civilised human beings. Animals, and probably primitive men, have the power of going to sleep quite easily by day as well as by night; and so presumably they do not suffer from boredom. But the ordinary civilised person has neither the power nor the opportunity to go to sleep at will; and, when he is not working or playing, he is liable to be bored and to be unable to escape boredom by sleep.

4.26. The nature of hedonic tone

Under this head I shall discuss three points which are of some interest.

(1) J.S. Mill and some other writers have said that pleasure and unpleasure can vary in *quality*, as well as in intensity. This doctrine is ambiguous. In one sense it is obviously true and in the other sense it is very doubtful. (a) The sense in which it is obviously true is this. Every experience which is pleasant or unpleasant derives its hedonic tone from some non-hedonic hedonifying quality. When we say, e.g. that pleasures differ in quality we might mean simply that two experiences which are equally pleasant may derive their pleasantness from very different non-hedonic qualities. This is a mere truism. Obviously a pleasant intellectual activity, such as solving a cross-word puzzle, and a pleasant organic sensation, such as being tickled, differ profoundly in the non-hedonic qualities which make them respectively pleasant. (b) The other interpretation is as follows. Pleasantness might be like redness, which can be present in various *shades* as well as in various intensities, and not like temperature which can be present only in various intensities. If so, you could compare and contrast two pleasant experiences in *three* respects, viz. their pleasant-making qualities, the intensity of their pleasantness, and the shade of their pleasantness. Everyone admits that they can be compared and contrasted in the first two respects; the peculiarity of Mills' doctrine, on this interpretation, would be that they can also be compared and contrasted in the third respect. Similar remarks would apply to unpleasantness.

I see no reason to accept this theory. I do not know of any facts which it is needed to explain. And I think that, if it were true, it would be almost impossible to establish it by introspection for the following reason. If two experiences did differ in the shade of their pleasantness there would almost certainly be a difference in their non-hedonic pleasant-making characteristics also. It

would be very difficult to be sure that there was a difference in shade of pleasantness over and above this difference in non-hedonic qualities.

(2) The next question is whether hedonic tone is a quality, like sensible colour or temperature, or whether it is a relational property, such as being liked or disliked by so-and-so. I am inclined to think that it is not a quality but a relational property. Suppose I say that a certain experience which I am now having is pleasant. It seems to me that what I mean is that I *like* it *for its intrinsic qualities* as an experience; and that, in so far as I pay attention only to these, I want it to continue. When I say that a certain experience which I am now having is unpleasant I mean that I *dislike* it *for its intrinsic qualities* as an experience; and that, in so far as I pay attention only to these, I want it to stop. Of course any experience has always plenty of other characteristics beside its intrinsic quality as an experience. It stands in certain relations to my other experiences, to my general scheme of life, and so on. It is likely to lead to certain consequences in myself or others. It may be morally innocent or morally wrong, and so on. Now an experience which I like for its intrinsic qualities may be liable to lead to consequences which I should dislike. It may fit in very badly to my general scheme of life and I may morally disapprove of it. For these *extrinsic* reasons I may want it to stop. Similarly, an experience which I dislike for its intrinsic qualities may be a necessary condition of consequences which I should like. It may be an essential factor in my general scheme of life and I may morally approve of it. For these extrinsic reasons I may want it to go on. Thus it is quite compatible with my view that one may *on the whole* desire the cessation of a pleasant experience or the continuance of an unpleasant one.

On this view what I have called "pleasant-making" characteristics of an experience are those intrinsic properties of it which make one like it and want it to go on. Unpleasant-making characteristics of an experience are those intrinsic properties of it which make one dislike it and want it to stop. Of course one may not at the time distinguish and take note of the qualities of an experience which are making him like it or dislike it and want it to go on or to cease. I do not see any reason to postulate, in addition to the pleasant-making and unpleasant-making qualities of an experience and the attitude of liking or disliking, a special kind of *quality* called hedonic tone.

(3) We can consider next the opposition and the blending of pleasantness and unpleasantness. If I am right this will depend upon the opposition and the blending of our likings and dislikings for our experiences in respect of their various intrinsic qualities. An experience may have simultaneously several characteristics. Some may be such that if they were there alone one would like the experience, others may be such that if they were there alone one would dislike the experience. Since both are there together, one's attitude will be one of blended like and dislike; which may be predominantly liking or pre-

dominantly disliking. In such a blend we must not suppose that if, e.g., the dislike for one factor is weaker than the liking for another, the dislike just neutralises so much of the liking and leaves a feeling of pure liking of diminished intensity. The opposition of pleasure and unpleasure is not like that of, say, positive and negative electricity put into the same conducting body. Both the liking and the disliking exist as factors in the total reaction, whether that be predominantly one of liking or one of disliking. Suppose, e.g., that one is on the whole indifferent to an experience, i.e. one neither likes it nor dislikes it on the whole. There is a great difference between what may be called "balanced" indifference and "uninterested" indifference. The former arises when the experience has *both* pleasant-making and unpleasant-making features, and the liking evoked by the former just balances the disliking evoked by the latter. The latter arises when the experience has *neither* pleasant-making *nor* unpleasant-making features. The two kinds of experience feel very different.

4.27. Pleasure and happiness

When we distinguish happiness from pleasure and unhappiness from unpleasure what we generally have in mind is this. We think of happiness or unhappiness as a characteristic of a person's life as a whole or of a considerable stretch of it. And we think of pleasures and unpleasures as particular strands of experiences which may have lasted for longer or shorter periods and may have succeeded each other or gone on side by side or have partially overlapped in time. We talk of a happy *person* and not only of a happy *life*.

The happiness or unhappiness of a person depends, not only on the pleasantness or unpleasantness of his various experiences taken severally, but also on the order in which they occur and their interconnexion with each other, and on his memories and anticipations of them. This is because we are at almost every moment of our waking lives looking backward on the experiences which we have had and forward to what we expect to do and to experience. An important factor in happiness is the consciousness that our life is going according to plan; that present experiences are fulfilments of past desires; that we are overcoming obstacles and exercising our various capacities efficiently; that we are acquiring new powers and improving our old ones; and so on. Again, among a person's various activities he regards some as more serious and "worth-while" than others. A person tends to feel dissatisfied with his life if much of it is occupied in purely sensuous or passive pleasures at the expense of activities of various kinds. And he tends to feel somewhat dissatisfied if much of it is occupied with activities which he regards as frivolous at the expense of others which he considers to be more serious and worth-while. Purely sensuous and passive pleasures and relatively non-serious activities do play an essential part in a happy life. The point is

that they contribute most to happiness when they are enjoyed as incidental rewards and relaxations in a life which is mainly occupied with activities which are felt to be serious.

It is of some interest to consider the following imaginary case. Suppose that at every moment of his life a man's first-order experiences, e.g. his sensations, were predominantly unpleasant. But suppose that, in spite of this, he had at every moment a delusive memory that all his past first-order experiences had been highly pleasant and a delusive expectation that all his future first-order experiences would be highly pleasant. Consider what would be the hedonic state of such a person at each moment. On the unpleasant side there would be two factors, viz. his present unpleasant sensations and other first-order experiences, and the unpleasant second-order experience of disappointed hopeful expectation. For he had expected all his future first-order experiences to be pleasant and those which he now has turn out to be unpleasant. On the pleasant side would be his false belief that all his past experiences have been highly pleasant and his false expectation that all his future experiences will be highly pleasant. It might well be that the pleasantness of his false memory and his false expectation would outweigh the unpleasantness of his present first-order experiences and his present disappointment at every moment of his life. In that case I think we should have to say that the life of such a person was a happy one and that he was a happy man.

The facts which I have described about the conditions of purely sensuous pleasures suffice to show that a happy life cannot be constructed out of them alone. The pleasures which depend on restoring a state of bodily depletion, or discharging from the body substances which have been stored up in it must, from the nature of the case, be short in duration and separated by considerable intervals. And successive experiences of a pleasure of this kind do not constitute successive stages in the progressive realization of any scheme of life. Then, again, pleasures of restoration are necessarily followed by feelings of temporary repletion, which may be unpleasant; and pleasures of evacuation by feelings of temporary exhaustion which may also be unpleasant. Lastly, purely sensuous pleasures which are not of these kinds suffer from the fact that the pleasantness of the experience tends to decrease after a time as the experience continues unless the stimulus increases in intensity; whilst, if the stimulus is increased beyond a certain point the sensation ceases to be pleasant.

On the other hand, I think that purely sensuous unpleasure, particularly if it takes the form of continuous bodily pain, will suffice to make any life unhappy whatever other factors there may be in it. It does so both directly and indirectly. Unlike purely sensuous pleasure, purely sensuous unpleasure and pain can be continuous. The unpleasantness can increase by the mere continuance of the same stimulus without any need for the stimulus to be inten-

sified. And there is no limit, short of the point at which one loses consciousness, to the intensity of pain which a person can experience at any moment. This is the direct contribution which it can make to unhappiness. The indirect contribution is that intense and continuous pain is extremely distracting. It enforces attention on itself and prevents a person from performing either the less serious or the more serious activities from which so much of our happiness is derived.

A very important condition for securing happiness, or at least avoiding unhappiness, is the ability to adjust oneself first to growing-up and then to growing old. Certain kinds of pleasant experiences and activities are possible and appropriate at certain stages of life, and are impossible or obtainable only with increasing difficulty and diminishing satisfaction at later stages. On the other hand up to a certain age, which varies very much from one person to another, new possibilities of pleasant experience and activity present themselves. In order to secure happiness and avoid unhappiness it is important that at each stage one's main desires and interests should be directed to the kinds of experiences and activities which are appropriate to that stage. If they remain centred on those of an earlier stage one is bound to feel constant unavailing regret, to embark on actions which can no longer bring satisfaction, and to miss opportunities of enjoying the pleasant experiences and activities which are appropriate to the present period of one's life. Typical examples of this are the person who has never grown up, and continues to react to the situations of adult life as if he were still in the nursery; and the elderly man or woman who makes himself or herself miserable and ridiculous by erotic interests and activities which would have been highly pleasant and quite appropriate at an earlier stage.

The last lap of a long life can scarcely be very happy except for a person who is effectively convinced that he will survive bodily death and that there is a good chance that he will pass into a state of enlarged opportunities for pleasant experience and enjoyable activity when his present body is destroyed. Lacking any such belief the concluding years of an old person can hardly fail to be made somewhat melancholy by the consciousness of decaying powers, the loss of contemporary friends and relatives, and the difficulty of adjusting himself to changes in his material and social environment. Beside this there is always growing bodily discomfort and weakness and sometimes almost constant pain, and the humilating awareness that one is becoming a useless burden to others and is being treated more and more as a child by them.

It is very important for everybody, and especially so for persons who have passed middle life and do not expect to survive bodily death, to take an active interest in persons and institutions which will outlive them. How much happiness depends on this can be seen in the part which the doings and sufferings of

grown-up children and grandchildren play in the lives of old persons. It can also be seen in the pleasure which a man takes in his work for a College or club or society which will go on after him and, as he fondly imagines, will continue to remember and be grateful for his services. Of course one's relationships to such persons and institutions can also be a source of great unhappiness. One's grandchildren may go to the bad in one's lifetime and one's College may be bombed to powder or ruined financially. I think that there are two kinds of person in this matter. One of them tries to avoid possible causes of unhappiness by withdrawing himself as much as possible from emotional commitments to other persons or institutions. The other takes the risk of such unhappiness by actively interesting himself in other persons or institutions, but he may secure much greater positive happiness in this way than is possible by the policy of self-insulation.

Lastly it may be remarked that in order to obtain happiness it is not desirable to aim too directly at it or to keep that object too constantly before one's mind. It is important to have reflected at some time on the general conditions of happiness and unhappiness, and on one's own special powers and opportunities and limitations for living a happy life. And it *is* important from time to time to review one's life from this point of view; to consider what mistakes one has made, what new facts one has learnt about oneself and other people and things; and in the light of this to make such changes as seem desirable. But for the greater part of one's life one will be most likely to secure happiness if one thinks mainly about other things, and carries on one's ordinary work, and play, and hobbies, and social relationships without explicitly considering the pleasure or unpleasure which one is getting from them.

4.3. Action and other notions involved in it

We can now consider the notion of action and various other notions, such as intention, motive, means-and-ends, etc., which are connected with it.

4.31. Different kinds of action

The actions of animate beings, i.e. creatures which are both living and conscious, may be classified as follows. (1) Those which they perform merely as organisms and not as conscious beings; and (2) those in which the fact that they are conscious as well as living is an essential factor. The former may be called physiological reflex actions. The latter may be divided in accordance with the kind of cognition which is predominant in them. They may be called *sensori-motor, perceptual,* and *conceptual actions*. From the point of view of Ethics only conceptual actions are directly of importance, and I shall consider them in detail. But it will be necessary first to say something about the other kinds for two reasons. In the first place we shall understand the peculiarities

of conceptual action better if we compare and contrast it with the other kinds of action. And secondly conceptual action involves the other kinds, although it cannot be reduced to them.

4.311. Physiological reflex actions

Here a certain kind of stimulus affects a certain nerve-coding and evokes a certain kind of bodily movement in a purely mechanical way owing to there being a connexion between the sensory nerve which carries the effect of the stimulus inwards and a motor nerve which goes outwards to a certain muscle. Examples are blinking when anything rapidly approaches the eye, sneezing when one sniffs pepper, and so on.

The process *may* be accompanied by a characteristic sensation, but it need not. And, when it is, neither the quality of the sensation nor its meaning in terms of physical things and events is a factor in causing the reaction.

The stimulus to a physiological reflex may come from within the body. One organ, e.g. a certain gland, may produce an internal secretion. This may be conveyed in the blood to another organ, and it may constitute the natural stimulus for its characteristic kind of reflex action. Such internal secretions are called *hormones*. They and the glands which secrete them play a most important part in preserving the balance of one's body and its processes and in determining our temperaments and our sanity or insanity.

Some physiological reflex actions can be controlled or inhibited deliberately if the stimulus is not too strong or too long continued. E.g. one may be able to prevent oneself from sneezing if it is important to do so and if the stimulus is only moderate. Such control is an instance of a *conceptual* action. Many reflex actions are altogether outside the voluntary control of all normal people. It is said that there are methods of training in use by Yogis which enable them to gain voluntary control over reflexes, such as the beating of the heart, which are quite out of the control of most people.

Some physiological reflexes can be *conditioned*. This means that they can be made to respond to a new stimulus through frequent association of this new stimulus with their original normal stimulus. The rules governing such conditioning in the case of certain reflexes of certain animals have been carefully investigated by Pavlov and his pupils.

4.312. Sensori-motor actions

These resemble physiological reflexes in their mechanical and unintelligent character. The difference is that the *intrinsic quality* of the sensation, though not its meaning in terms of physical things and events, is an essential factor in causing the action. It is not a mere idle accompaniment as in the physiological reflex. An example is dropping a hot plate. One drops the plate because of the intolerably unpleasant character of the tactual sensation. If one were

anaesthetised, so as not to have the sensation, one would not drop the plate and might be badly burned. This sometimes happens with unconscious patients who fail to withdraw their feet from uncovered hot-water bottles put into their beds by careless or incompetent nurses.

Images may act in somewhat the same mechanical way as sensations. A vivid image of the taste and smell of some very nauseous stuff, e.g. castor-oil, might make one feel sick and even be sick if one were in a susceptible state. This is called *ideo-motor* action.

A very complicated series of actions, which looks superficially as if it were guided by intelligent insight and foresight, may really be sensori-motor or reflex. This can arise in the following way. Suppose we imagine a creature, e.g. an insect, provided with a very complex innate disposition consisting of the factors *abc...lmn*, elaborately interconnected. Suppose that *a* is stimulated from outside by the appropriate natural stimulus and produces the action α. It may be that α itself, or some part of the results of α, is the natural stimulus to *b*. So *b* will now be stimulated and will produce the action β. Similarly β itself, or some part of its results, may be the natural stimulus for *c*, which would produce the action γ when stimulated. And so on. Thus we should have a chain of inter-connected actions $\alpha\beta\gamma...\lambda\mu\nu$ started by the initial external stimulus which set off *a*. These might be so interrelated as to lead up to a result which is important to the agent itself or to its offspring. E.g. they might lead up to the deposition of an egg as a certain part of a caterpillar which had previously been stung in just the right places to paralyse it without killing it. This would look like an extremely intelligent series of actions by the insect to secure that the larvae would have fresh meat when they came to be hatched. But, if we varied the conditions, we should find that the action still went on in a blind mechanical way to a result which was futile or positively detrimental.

Probably a great deal of the apparently intelligent behaviour of insects is of this kind. Mammals in general, and men in particular, seem to have no trace of these elaborately organised innate dispositions to reflex or sensori-motor actions. But they have the power of building up such dispositions by training and practice, and then using them in perceptual or conceptual actions. E.g. the habitual movements of the tongue, lips, and throat in speaking our native language must depend on a very complex set of dispositions which we acquired in childhood. One does not know precisely what one is doing with one's tongue, etc. when one speaks each word; and the actual mechanism of speech must consist of elaborately organised dispositions to reflex or sensori-motor actions.

4.313. Perceptual actions

In perceptual action our behaviour is guided, not by the qualities and rela-

tions of our sensations as such, but by the *meanings* which these have for us in terms of physical things and events. A very good and nearly unmixed example of a perceptual action is a man playing a vigorous game of skill – such as tennis, or a cat hunting a mouse. The player's action is guided at every moment by his visual sensations; but he is guided by their qualities and relations only in so far as he takes them as indicating the position, velocity, and direction of the ball, the position of his opponent, and so on. On the other hand, there is no question of making calculations and inferences based upon general laws. That would be conceptual action; and the game is too fast to allow time for it.

There are two characteristic features of perceptual actions which distinguish them from complicated sensori-motor or reflex actions. They are what Stout calls "the *prospective attitude*" and "*persistence with varied effort*".[1] I will now say a little about each. (1) What one perceives at any moment is not literally instantaneous, it stretches a little way back into the past. This is shown by the fact that we can literally *perceive things as changing*, if they change quickly enough, as opposed to *merely inferring that they have changed*. E.g. we perceive the second-hand of a watch as jumping, a flame as flickering, and so on. So what we perceive at any moment must be represented, not by a mere dot, but by a short line stretching back from that moment. This short stretch of past time is called the *specious present*. Suppose we divide what is perceived within a single specious present into successive thin slices. Then we must say that, although it is all being perceived, it is not all being perceived with the same *degree of presentedness*. At the earlier end the degree of presentedness is at a minimum; for there what is perceived fades away with what has just ceased to be perceived and just begun to be remembered. At the later end the degree of presentedness is at a maximum. For intermediate positions in the specious present the degree of presentedness has intermediate values. Our experience consists of a continuous series of specious presents. These are not just adjoined end to end, so that the content of one has nothing in common with that of the next. (If they were our experience would not be continuous but jumpy.) Specious presents which are near enough in time partially overlap each other. Part of the content of the earlier is also part of the content of the later. But it will have a lower degree of presentedness in the later than in the earlier of the two. (If we want to represent both the variation in degree of presentedness from one end to the other of a single specious present and the partial overlapping of successive specious presents we can do so as follows. We can represent each specious present by a little wedge. And we can represent the course of our experience by an échelon of partially overlapping wedges. Since the course of experience is continuous, we must conceive something which we cannot draw, viz. that between any two wedges, however near together, there is a third wedge.)

1. G.F. Stout, *A Manual of Psychology* (4th edition, London, 1932), p. 337.

I shall call the fact that what is perceived at any moment stretches back a little way into the past from that moment "*short-range* retrocognition". *Long-range* retrocognition is memory and inferential knowledge or belief about events in the remoter past. Presumably animals have short-range retrocognition; it is more doubtful whether they have long-range retrocognition also.

Now there is also something which may be called "*short-range* expectation". By this I mean non-conceptual cognition of a very thin slice of the immediate future. It is obvious that at every moment one is expecting to have in the immediate future further experiences which will join up with one's present experiences in the way in which they joined up with one's immediately past experiences. This expectation is never completely determinate and never completely indeterminate. The details of one's short-range expectations at any moment are determined within certain limits by the character of the contents of one's specious present at that moment. The proof that one has short-range expectations is the fact that one is capable of feeling *surprise* at the immediate developments of the present situation. The criteria by which to judge what has and what has not been expected is to notice what developments would, and what would not, cause surprise to an individual.

Suppose, e.g., that a cat has been catching a mouse for a time. He will not be surprised if, when what is now immediately future shall have become present, the mouse shall have remained still or shall be jumping in any direction. But he will be extremely surprised if the mouse shall have exploded or suddenly vanished, or if there should be a sudden change in the environment such as the fall of a picture. We may say, then, that the cat has a short-range expectation that the general environment will remain unchanged in the immediate future and that the mouse will do one or other of a limited range of alternatives. But there is no particular one of these alternatives which he expects the mouse to do rather than another. He watches with insense interest to see *which* of these alternatives will be realised, and he automatically adjusts his body so as to be ready to act appropriately with the least possible delay at the first sign of the mouse adopting a certain one of them.

We can now define the "prospective attitude". It consists in the following things. (i) In taking for granted that the general environment will remain practically unchanged in the immediate future, and that a certain perceived object in it will actualise one or other of a certain limited range of possible alternatives. (ii) In watching this particular object with special attention in order to see which alternative it is going to realise. (iii) In holding one's mind and body in readiness to act appropriately with the least possible delay at the first sign of a certain alternative being about to be realised. I think that Stout is right in counting this as a characteristic feature of perceptual action in con-

trast with reflex or sensori-motor action. It is a mark of action which is intelligent without being intellectual.

(2) The other characteristic feature is *persistence with appropriately varied effort*. There is plenty of persistence at the purely reflex level. We can see this when a moth repeatedly dashes itself against a window or an electric-light bulb under the stimulus of light. But there is no sign of appropriate variation in the action in order to circumvent an obstacle or to overcome a failure. Action at the perceptual level is continually and delicately modified to meet each relevant change in the perceived situation. If it is checked by some unexpected obstacle, it is not just blindly repeated. It is varied within limits in such a way as to get over or round the difficulty.

Now, if we are to talk of "persistence with varied effort" we must have some criterion of which constitutes a repetition of the same action as distinct from the starting of an entirely new action. In order to do this we must introduce the notion of an action "expressing a *need*". When a cat is hunting a mouse its actions, whether successful or unsuccessful, express a certain need, viz. to hunt and catch a certain object. So long as that need is predominant that object, viz. the mouse and its movements, remains at the centre of the cat's interest and attention. Finally one of two things happens. Either the cat catches the mouse and eats it or the mouse escapes. In the first case the cat will lose interest for the time in hunting and will probably go to sleep. We say then that the action has *fulfilled* the need which it expresses. In the second case the cat will for some time show signs of unrest, such as prowling about and looking into corners. After a time these will subside if nothing like a mouse appears and the cat is not desperately hungry. We say then that the action expressed the same need, but *failed to fulfill* it.

It is important to notice that the final stage of catching and eating the mouse is not by itself the fulfillment of the need. If you simply put a live mouse into a cat's mouth for it, the need to hunt will not have been fulfilled and the cat will merely be bored. In fact a need is always for attaining the result in a certain kind of way. If the process fails to attain the result, or if the result is attained without the process, the need equally fails to be fulfilled. We will call these two indispensable factors the "needed process" and the "needed result".

When a need is excited a process of the needed kind tends to start and to continue. During this period anything perceived which tends either to further or to hinder the needed process in reaching the needed result will be of special interest. Anything perceived which neither helps nor hinders will tend to be ignored. The needed process comes to its natural end if it brings about the needed result. The need then ceases to be active and the centre of interest changes. Of course a similar need may recur after an interval. Most of our natural needs, like the need for eating, for sleeping, and so on, are recurrent.

While the needed result is still unattained the needed process is varied to fit with the changing relevant details in the environment and in order to overcome obstacles. If, in spite of all such variations, the needed result fails to be attained, the process may eventually die away although the need remains unfulfilled. But it will be likely to die away by fits and starts. And the unfulfilled need will be likely to perturb one's other experiences and actions.

This brings us to the distinction between *needs* and *wants*. It is one thing to have a need; it is another thing to know *that* one needs; and it is yet another thing to know *what* one needs. Now wanting is a special kind of experience closely connected with needing. When there is an unsatisfied need there will generally be an experience of wanting; and when there is an experience of wanting there is always an unsatisfied need at the basis of it. Suppose that a certain unsatisfied need gives rise to a certain experience of wanting. What is wanted may be different from what is needed either in detail or in principle. The wanting may be no more than a vague emotion of discontent with the present situation, leading to an aimless restlessness. But suppose that it is a definite experience of wanting so-and-so, and leads to a certain course of action. It may still be the case that this action is not really the kind of action needed, and that the result is not the needed result. If so, it will be found that, when the action has been done and the wanted result has been gained, the feeling of discontent does not vanish.

Now, as a perceptual action goes on, one experiences a continuous series of short-range wantings. Our cognition at each stage looks only one step ahead, and our wanting is at each moment this short-range expectation qualified by a characteristic kind of emotional tone. The fact that we have short-range wantings is proved by the fact that we feel disappointed if the situation develops in certain of the expected ways and elated if it develops in certain others of them. Wanting is a *prospective* emotion felt towards the *expected* future. Disappointment and elation are *retrospective* emotions, felt towards the present situation in respect to its fulfilling or frustrating the want that we felt in the previous situation. The test of what we were wanting in the immediate past is what disappoints or elates us in the present.

Now this binds the successive phases of experience which we have in connexion with any long perceptual action with a characteristic unity. Each short phase is bound to its immediate predecessor by fulfilling or frustrating the wants which were factors in that predecessor, and being toned with retrospective elation or disappointment in consequence. And each short phase is bound to its immediate successor by containing wants which are directed forward to it. These conative and emotional links bind the successive short phases into a single outstanding strand of experiences.

Now, whenever we are performing any bodily action, there will be a series of sensations due to the adjustments which we are making in our bodies and

to the reactions on us of other bodies which are cooperating with or resisting our efforts. (The case of a man swimming against a stream or sawing up a log is a good example.) This series of bodily sensations becomes fused with the strand of experiences which I have been describing and it is an essential part of the total experience of acting at the perceptual level. If it were not for these bodily sensations and their constant variation we should not have the clues needed for continually readjusting our bodies. Probably the action could not go on at all; and, even if it could, the experience would be quite strange and unfamiliar. Suppose, on the other hand, that a precisely similar series of bodily sensations could be produced artificially instead of arising in connexion with fulfilling a need and with the experiences of short-range expectation, wanting, disappointment, and elation. Then they would not suffice to constitute the experience of actings. They would be a mere series of bodily sensations with no special internal unity.

I will now sum up this account of perceptual action and the experiences which are characteristic of it. From time to time an animated organism has a certain need. When it has a need an internal process starts in it and certain perceived objects become invested with a special interest. This process will be of a special kind, according to the nature of the need, and it will propagate itself in a special direction. It will be a process which tends to bring about a certain result, viz. the needed result. But this tendency can be realised only by constant interaction with other agents, which have laws, properties, and tendencies of their own. The action will be successful in so far as the process within the agent brings about a suitable co-operation between the other agents and between them and himself. At each moment the other agents will be reacting on him and he will be perceiving their present and immediately past reactions. This perception will determine, in a non-inferential non-intellectual way, his present expectations of their further actions and reactions in the immediate future. His present short-range expectations will determine his next bodily and mental adjustments. And these internal changes in the agent will be a factor in coordinating the other agents to each other and to himself in the next phase of the total process directed towards fulfilling the need. In each phase of the process his experiences will be toned with elation or disappointment, according as that phase does or does not fulfill the short-range wantings of the immediately previous phase. And, again, each phase will include as a factor short-range wantings directed towards the immediately subsequent stage. In this way the successive phases are hooked together with a single outstanding strand of experience. The adjustments of the agent's body and the reactions of other agents upon it keep up a continual stream of bodily sensations, and these fuse with the outstanding strand of experience just described. The whole thus formed is what we call an "experience of acting" or a "conative experience" at the perceptual level.

It is plain that such strands of conative experience can and do occur in a creature which has no power of reflexive cognition and therefore cannot *think of itself* as an agent interacting with other agents in a process which is adapted to fulfil a certain need. But they also occur in beings like ourselves who have the power of reflexive cognition and of intellectual analysis and synthesis. And in such a being, this kind of action and the experiences which accompany it are the data from which he derives his notions of agents and activity, cause-factors and total causes, and so on.

4.314. Conceptual actions

The essential difference on the cognitive side between conceptual and perceptual action is this. We now have long-range expectations and long-range beliefs about the past. We contemplate and compare alternative possibilities, and weigh up the *pros* and *cons* of each, and make a decision on this basis as to which we will try to realise. We analyse situations and we make inferences based on our knowledge or belief about the laws and properties of agents. The main outlines of an action may now be thought out and decided upon beforehand. We may also have decided beforehand what alternative kinds of action we will take according as future events turn out in one or another of various alternative ways. We may then leave the details to be filled in under the guidance of perception when the time comes for beginning to act. It is only at this level that such notions as *intention, motive, means-and-end*, and so on, begin to apply.

The experience which accompanies conceptual action differs in certain characteristic ways from that which accompanies perceptual action. In perceptual action each phase is linked only to its *immediate* successor and its *immediate* predecessor by short-range expectations and wants and by the elation or disappointment which is the sign of their fulfilment or frustration. In a conceptual action this kind of unity will be present, but there will also be a further kind of unity which binds together phases that are not adjacent in time. In such action the process does not merely *in fact* tend towards a certain end. It has been contemplated in outline by the agent beforehand and devised in order to reach a certain end which he has set before himself. What is wanted at each stage is, not merely a certain immediate development of the present situation, but a certain course of development leading up to a certain proposed end. At each stage one will feel elation or disappointment in so far as one thinks that what has just happened is in accordance with one's plan and is leading on to the desired end in the desired way, or is a hindrance diverting one from reaching the desired end in the desired way.

I do not suggest that at every stage we are explicitly contemplating all the future stages and the proposed result. But at any rate we *have* done so, and the trace of this cognitive experience is continously excited and qualifies every

phase of the experience. And at certain stages we do explicitly take stock of the situation and contemplate what is yet to be done and what result we hope to reach. Owing to this kind of unity, depending on memory and long-range expectation, there can be temporal gaps in a conceptual process of action, and yet the phases on each side of the gaps are united with each other into a single strand of conative experience. A single conceptual action, e.g. studying for an examination, may go on for days or months, broken by intervals of sleeping and doing other things.

4.32. Relations between the four kinds of action

Presumably a very simple organism, like an oyster, is capable only of reflex or sensori-motor action. A higher animal, such as a cat, is capable of both this and perceptual action. A human being is capable of all four kinds of action.

Now we must not assume that these various types of action are mutually exclusive. In creatures which are capable of several of them they form a kind of hierarchy. Perceptual action may be said to be built upon reflex and sensori-motor action, and conceptual action upon perceptual action. I will now try to explain this statement.

Let us begin with perceptual action. This may be built upon sensori-motor and reflex action in at least the two following ways. (i) In a total course of action which was predominantly perceptual there might be occasional stretches which were reflex or sensori-motor. (ii) A course of action which, taken as a whole, is perceptual and intelligent, may be divisible into a series of short successive phases, each of which, taken separately, is reflex or sensori-motor. Playing a difficult piece of music on the piano is obviously a perceptual process taken as a whole. But the striking of each note is probably an acquired reflex or sensori-motor action. The intelligence which is displayed in a course of perceptual action, taken as a whole, consists in the appropriate order and combination in which various dispositions to reflex or sensori-motor action are stimulated and inhibited under the guidance of perception. In fact the word "intelligent", as applied to a perceptual action, is the name of a pattern-quality. It applies to the process as a whole in virtue of the order and interconnexion of the phases; and it need not belong to the phases themselves.

Next let us consider conceptual action. An action which is conceptual as a whole may, and generally does, consists of phases which are perceptual. Suppose, e.g., that one plans and constructs a bit of machinery, or that one writes an essay. These actions, taken as a whole, are plainly conceptual, i.e. deliberately undertaken for a purpose, guided and controlled by thought. But each phase of using hammers, files, lathes, etc. is a perceptual process. And the process of writing is a set of perceptual processes which are themselves composed of sets of acquired reflexes. What makes the action as a whole con-

ceptual is the fact that all these perceptual processes are started, stopped, co-ordinated, etc., under the guidance of a plan which has been worked out in thought in its main outlines.

A conceptual action is a complex pattern whose immediate components are perceptual actions interrelated in a characteristic way under the guidance of thought. Each of these immediate components is itself a complex pattern whose immediate components are sensori-motor and reflex actions interrelated in a different characteristic way under the guidance of perception. We might compare a conceptual action to an intelligible complex sentence; its perceptual components to the clauses in this sentence; and the sensori-motor and reflex components of each perceptual component to the words in a clause.

4.33. Merits and defects of the four kinds of action

Conceptual action is most suitable when a new problem presents itself, when one is placed in a new situation, and when one wants to gain knowledge of or control over one's environment or oneself. The defect is that it is necessarily slow and that it is tiring. It is therefore unsuited to situations which are changing quickly, and when an immediate response is needed to each change in the situation.

Provided that the main outline of the situation remains fixed and familiar and only the details change quickly, as e.g. in playing a game like tennis, perceptual action is the only satisfactory kind to use. It is quick enough and it is not nearly so tiring.

A conceptual action which is often repeated tends to set up a disposition which will enable the action to be repeated in future with less and less thought. Eventually this kind of action may become completely perceptual. An obvious example is learning to play a game of skill, to drive a car, to use a lathe or typewriter, and so on. While the action is still being performed conceptually it is done inefficiently and slowly. But this is an essential step in building up the disposition by which it will eventually be performed non-conceptually and efficiently. When this is acquired our very limited power for conceptual action will be set free for tackling new problems. This is an extremely characteristic feature of human life. We are continually using conceptual action to give ourselves powers of non-conceptual action which will enable us to dispense with conceptual action in a certain department. As we do so our action in that department becomes more efficient, and our power of conceptual action is set free to tackle something else.

A similar result is often achieved in another way. We may use our powers of conceptual action to devise and construct machines which will do certain things more efficiently than they can be done either conceptually or perceptually. In this way we set free both the power of perceptual action and that of conceptual action.

Here the advantage, though often very great, is much more mixed. To acquire new powers of mind or body is an addition to one's personality and to one's sources of interest; for what makes life worth living is exercising one's powers in doing efficiently a variety of things that seem interesting and important. It is therefore an unmitigated gain when conceptual action is used to give oneself new powers of non-conceptual action. But the delegation of skilled action to machinery, though it eliminates much drudgery, may leave the majority of human beings with very little to do which they can feel to be interesting and important, and with very few powers except that of pressing buttons and pulling levers. Personality is thus impoverished and life becomes boring and all that one can do seems trivial. Thus, at any stage of human development there is probably a certain optimum amount of mechanical invention. Mechanical invention now increases much faster than new sources of human activity and interest are developed. It seems likely that this optimum has been passed, and that human life in highly industrialised communities is undergoing an increasing impoverishment.

Just as certain actions which were conceptual tend to become more and more completely perceptual with practice, so certain actions which were perceptual tend to become more and more completely habitual. Now an habitual action is a kind of acquired reflex or sensori-motor action. When one is learning to ride a horse the action is at first conceptual. When one has gained a certain amount of skill it becomes perceptual, and one can think of other things though one still cannot let one's attention wander very far. Finally, it becomes habitual, and one can go to sleep on the horse's back provided it is quiet and does not stumble. Since habitual action demands little or no attention, it is less tiring than either perceptual or conceptual action. It is suitable when both the outline and the details of the situation are familiar and practically constant. In such cases the action may become perceptual or even conceptual again if the circumstances change, e.g. if the horse stumbles or one sees a traction-engine approaching.

Now both conceptual and perceptual action have intrinsic value. To be able to perform many such actions well is a source of interest and self-satisfaction to a person. But habitual action has little, if any, *intrinsic* value. Such value as it has is almost wholly instrumental, viz. that it releases energy which may be used for perceptual and conceptual action. But it may easily happen that the energy which is set free through certain conceptual and perceptual actions sinking to the habitual level or by the use of machinery is not in fact used to increase the variety and delicacy of one's powers of conceptual and perceptual action. A person may have no opportunities to do this, or he may fail to use such opportunities as he has. In such cases the process which I have been describing has a twofold disadvantage. In the first place, one's life and personality are impoverished and one sinks to the level of an animated

automatic machine. Secondly, the energy set free and not utilised will certainly cause restlessness and discontent. And it will probably find an outlet in destructive emotions and actions, such as violent crime, political revolution, or patriotic nonsense leading to war.

4.34. Notions connected with conceptual action

I shall now consider certain notions which are associated with action at the conceptual level. The most important of these are intention; motive; action and consequences; means and end.

4.341. Intention

Any action can be considered under at least three headings viz. (i) the *hedonic tone* of the experiences which accompany it, i.e. whether it will be pleasant or unpleasant in itself; (ii) its *non-causal* relational properties, e.g. whether it will be an act of promise-keeping, of ingratitude, and so on; (iii) its *causal* relational properties, i.e. the effects which it will have on the agent himself or on others.

Now a person who does an act may either have considered it beforehand in respect of some of its properties or he may do it altogether without previous consideration. An example of an *unconsidered* act would be when a person quite suddenly sneezes without knowing that he is going to do so. If an act is considered beforehand by the agent it may be either done for a reason or not done for a reason. In the latter case we will say that it is *considered but unmotived*, and in the former that it is *motived*. The following would be an example of an act that is considered beforehand but unmotived. Shortly before sneezing during a sermon in church a person may feel the impulse to sneeze and he may consider the act and know that it would have the pleasant consequence of clearing his head and the unpleasant consequences of disturbing the sermon and attracting attention to himself. But these considerations may be no part of the cause of his sneezing. He may just sneeze involuntarily, and not as we should say "with a view" to clearing his head. Suppose, on the other hand, that he could have avoided sneezing but decided to let it take its course, because he had decided that it was worth while to put up with interrupting the preacher and attracting attention to himself in order to clear his head. Then his act of sneezing would be, not only considered, but *motived*. Now an act that is motived may have been *chosen* out of several possible alternatives which the agent seriously considered and balanced against each other. If so we will say that it was *deliberately chosen*. Or it may have been done without any appreciable consideration of other possible alternatives. E.g. suppose a person asks one a legitimate question on a matter which does not involve anything confidential. One generally returns what one believes to be a true answer without considering various alternatives lies

which one might tell instead. Here the answer is motived, but not deliberately chosen. Suppose, however, that a true answer would hurt the questioner's feelings, or betray a confidence, or expose oneself to humiliation. Then one seriously considers various alternatives to giving a true answer, and weighs up the *pros* and *cons* of telling the truth and telling one or other of these lies. If in the end one decides to give a true answer, e.g., the act is not only motived but *deliberately chosen*. It is obvious that the feature of acting for a reason or with a motive comes out most clearly in the case of actions which are deliberately chosen. For here there has been a process of deliberation, in which the various reasons for and against the action which is eventually chosen have been made explicit in order to compare them with the reasons for and against other alternatives.

To sum this up. An act may be either unconsidered or considered before being done. If it has been considered it may either be motived or unmotived; i.e. the agent's beliefs, at the time when he considered it, that it would have such and such properties, may or may not have been for him *reasons* for doing it. If it is motived, it may either have been deliberately chosen out of a number of possible alternatives or have been done without serious consideration of other alternatives. I shall call an act *intentional* if and only if it is both *considered and motived*, but without regard to whether it is deliberately chosen or not. On this definition a reflex jerk of the knee is unintentional because unconsidered. When Oedipus married Jocasta his act was intentional in so far as it was an act of *marriage*. It was unintentional so far as it was an act of *incest*, for he did not know and did not believe that Jocasta was his mother. If a person puts what is in fact ordinary sugar into another man's tea, mistakenly believing it to be arsenic and expecting it to cause death, part of his intention is to poison that man, although his belief is false and his action ineffective. Using "intentional" in the widest sense, we may say that an act is intentional in respect of *all* those properties and *only* those properties which the agent, correctly or incorrectly, ascribes to it when he considers it before doing it. These properties include the hedonic tone which he expects it to have, the non-causal relations which he thinks it will have, and the consequences which he expects it to have. I shall call them the *ostensible* properties of the act.

Now an act may have properties which the agent neither believed it to have nor disbelieved it to have when he considered it, because he never considered that aspect of it all. This is true of all the very remote consequences of any act. I shall call such properties of an act *extra-intentional*, i.e. *outside* the agent's intention. Again, it may have properties which the agent disbelieved that it would have, or it may lack properties which he believed it to have. An instance of the first is if a boy points what he believes to be an unloaded gun at a person and pulls the trigger and the gun really is loaded. An example of the

second is if a person administers a harmless substance, like sugar, in the mistaken belief that it is a poison, like arsenic. I shall call such properties *contra-intentional*. An act is unintentional in respect of all those properties and only those which are either *extra-intentional* or *contra-intentional*.

In ordinary life we often use "intention" in a narrower sense than this. Suppose, e.g., that an anarchist throws a bomb at a ruler in a public procession. Suppose that the anarchist expects that the bomb will kill the ruler, that it may injure and perhaps kill the chauffeur and some of the spectators, and that it will break the glass in neighbouring shop-windows. Then, on my definition, all these effects are part of the anarchist's intention in throwing the bomb. But it would be quite usual to say that his intention was to kill the ruler; and not to regard the killing or injuring of other persons or the breaking of shop-windows as part of his intention. On the other hand, it would be very strange to say that these other consequences, which the anarchist foresaw and was prepared to bring about, were *unintentional*. We can deal with this complication by distinguishing between a person's *primary* intention and his *secondary* or collateral intentions in doing an act. The primary intention is to do an act with those properties which attracted the agent towards doing it and constituted his reasons or motives *for* doing it. The secondary or collateral intentions are concerned with those ostensible properties of the act which *either* repelled the agent from doing it and constituted reasons or motives *against* doing it *or* left the agent indifferent between doing it and avoiding it. Thus we may suppose that the primary intention of the anarchist was to kill or injure the ruler; whilst killing or injuring other persons and breaking shop-windows were secondary or collateral intentions.

It is worth noticing that for legal or ethical purposes actions are sometimes classified by their primary intentions, without regard to their actual results, and sometimes by their intentions and their actual results jointly. E.g. a statement made with the intention of producing a false belief in a person is called a *lie*, whether it does in fact produce a false belief or not. And a statement which does in fact produce a false belief without being intended to do so is not called a lie. But in order that an action may count as a *murder* it is necessary both that it shall be done with the intention of killing another person and that it shall in fact do so. If the intention is absent but the result follows it, the act counts as *homicide* or *manslaughter*. If the intention is present but the result is not achieved, it counts as *attempted murder*.

4.342. Motive

This is a complicated business, and there are many distinctions to be drawn and confusions to be cleared up.

4.3421. Ambiguity in the word "motive". Suppose that *A* has murdered *B*

and that we ask: "What was *A*'s motive in murdering *B*?" One quite reasonable answer would be "*A*'s motive was revenge". Another equally reasonable answer would be: "*A*'s motive was his belief that *B* had done him an injury". Both these statements might be true. The former refers us to a certain *emotional or conative disposition* in *A*, viz. a standing desire to injure those who have injured him. The latter refers us to a certain *belief* in *A* about *B*, which excites this conative-emotional disposition and directs *A*'s emotion and desire towards injuring *B*. The two kinds of answer refer to two different aspects of a single fact. In accounting for anyone's doing a certain intentional action at a certain moment it is always necessary to refer to two different kinds of cause-factor. One of them is the agent's conative-emotional dispositions. Unless he had such dispositions nothing that he believed about an action and its alternatives would either attract him or repel him. And unless different persons had a different balance of conative-emotional dispositions we could not account for the fact that two people who agree in their beliefs about the same alternatives will choose differently. The other cause-factor is certain of the agent's actual cognitive states at the time, i.e. his perceptions, knowledge, and beliefs about the alternatives. If it were not for these we could not account for the fact that now this and now that conative emotional disposition is excited.

Anyone who is about to deliberate between alternatives, or to act intentionally without previous deliberation, does so with a complex system of pre-existing conative, emotional, and cognitive dispositions. This has, no doubt, been gradually built up, and it will no doubt gradually change in future. But, for the period during which a process of deliberation is likely to last, it may be taken as a permanent system of dispositional cause-factors. The cause-factors which vary during the process of deliberation are, in the main, one's acts of attending now to this and now to that ostensible characteristic of the various alternatives under consideration. These excite now this and now that one of our conative-emotional dispositions and thus make each alternative attractive in certain respects and repulsive in others.

In order to avoid ambiguities we will introduce the terms *motive-factors* and *total motive*. We shall say that the total motive in any intentional action always contains two kinds of motive-factor, viz. conative-emotional and cognitive.

The conative-emotional factors are mainly dispositional, e.g. pre-existing sentiments, etc. But they may not all be merely dispositional. One may enter on a process of deliberation with some conative-emotional disposition already excited, e.g. in a mood of anger or depression or fear. If so, one's deliberation will tend to be biassed, and this bias may take two forms.

In the first place, the deliberation will tend to be biassed on the cognitive side. One will tend to notice and dwell upon features in the alternatives which

seem to justify anger or depression or fear. One will tend to ignore or pass lightly over features which, if fairly attended to, would excite other conative-emotional dispositions. Secondly, it will tend to be biassed on the conative-emotional side. Since one set of conative-emotional dispositions is already excited, the emotions and conations connected with these will have a start. They may inhibit the excitement of other conative-emotional dispositions, even if one attends fairly to the circumstances which would normally excite the latter.

Completely unbiassed deliberation is an ideal limit which is perhaps never attained. Even if we do in fact sometimes attain it, we can never be sure that we have done so. But we can certainly approximate to it, especially if we are acutely aware of the tendencies towards bias and are on guard against them. And we can be quite certain that we have come nearer to it in some deliberations than in others, and in some stages of one deliberation than in other stages of it.

The cognitive motive-factors can always be divided into dispositional and occurrent factors. The dispositional ones are our acquired system of knowledge and belief, and especially that part of it which is closely connected with the alternatives about which we are deliberating. The occurrent cognitive motive-factors are our constantly varying acts of attention, thought, perception, inference, etc., directed now to one aspect and now to another aspect of the alternatives under consideration.

4.3422. Absolute and relative attraction and repulsion. Let us suppose that a person is considering several alternative courses of action *A*, *B*, *C*, etc. We will first confine our attention to a single one of them, let us say *A*.

In respect of some of its ostensible characteristics *A* will attract the agent towards doing it, in respect of others it will repel him from doing it, and in respect of others it will perhaps neither attract him nor repel him. So we can divide the ostensible characteristics of any alternative which an agent considers into *attracting, repelling,* and *neutral.* Corresponding to each attracting ostensible characteristic will be what I will call a *component of absolute attraction.* Corresponding to each repelling ostensible characteristic there will be a *component of absolute repulsion.* The resultant of all the components of absolute attraction and all the components of absolute repulsion which alternative *A* has for the agent will be called the *resultant absolute motive-force* which that alternative has for that agent. This may be positive, in which case it will be called the *resultant absolute attraction.* It may be negative, in which case it will be called the *resultant absolute repulsion.* Or it might happen to be neither. In that case I should say that the agent was in a state of *resultant absolute indifference* towards that alternative.

So far we have considered each alternative separately. But deliberation is a

matter of comparing several alternatives and choosing between them. So we must now deal with *relative* attraction and repulsion.

(1) The simplest possible case is this. Suppose that A has a resultant absolute attraction for the agent, and suppose that each of the other alternatives B, C, etc. is either resultantly repulsive or resultantly indifferent to him. In this case, and in this only, we can identify his *total motive for choosing A* with the resultant absolute attraction of A for him.

(2) This simple case rarely arises. (i) The agent may dislike all the alternatives and choose the one that he dislikes least on the whole. (ii) He may like all the alternatives and choose the one that he likes best on the whole. Or (iii) some of the alternatives may be resultantly attractive and all the rest either resultantly repulsive or indifferent. In that case his choice will eventually be between those that are resultantly attractive, and the others will pass out of the picture. So in the end there are only two cases to be considered, viz. (a) choice between alternatives which are all resultantly attractive, and (b) choice between alternatives which are all resultantly repulsive. The principles are the same in both cases, so we can confine ourselves to the former.

Let us suppose that I have to choose between three alternatives A, B, and C, each of which has a resultant absolute attraction for me. It is plain that my total motive for choosing A must be composed of my motive for preferring A to B and my motive for preferring A to C.

Let us consider the motive for preferring A to B. Any two alternatives which come under consideration in a single deliberation will have a good deal in common. The differences between A and B can be brought under the following three heads. (i) Factors ostensibly present in A and absent in B. (ii) Factors ostensibly present in B and absent in A. (iii) Generic characteristics, ostensibly present in both A and B, but in different specific forms in the two.

Now any of the following three sorts of factor will move me to prefer A to B. (i) Attracting features ostensibly present in A and absent in B. (ii) Repelling features ostensibly present in B and absent in A. (iii) Common generic features ostensibly present in a more attractive or a less repulsive form in A than in B. Corresponding to each of these three will be what I will call a *component of preference for A to B*. Similarly there will be three kinds of factor moving me to prefer B to A. Corresponding to each of these will be a *component of preference for B to A*.

The resultant of all the components of preference for A to B and all the components of preference for B to A will be the *resultant motive of choice between A and B*. This may favour A, or it may favour B, or it may happen to be exactly balanced. According to which of these three possibilities is realised we call it either the *resultant motive for choosing A in preference to B* or the *resultant motive for choosing B in preference to A* or the *resultant motive for indifference between A and B*.

Exactly similar remarks would apply to *A* and *C*. Suppose now that I choose *A* and reject *B* and *C*. Then my *total motive for choosing A* is composed of my resultant motive for choosing *A* in preference to *B* and my resultant motive for choosing *A* in preference to *C*.

There are two remarks worth making before we leave this part of the subject.

(1) We must distinguish between being indifferent *towards* a certain alternative *A*, and being indifferent *between* two alternatives *A* and *B*. Indifference between *A* and *B* is compatible with there being a strong resultant absolute attraction or repulsion towards both of them. We must also distinguish between *balanced* and *uninterested* indifference towards an alternative. The former means that it *both* attracts *and* repels us and that the two components are equal and opposite. The latter means that it *neither* attracts *nor* repels us. The two experiences are quite different.

(2) Suppose that one has to choose between alternatives each of which is resultantly repulsive to one, e.g. between being exposed and ruined or paying money to a blackmailer or committing suicide. A person in such a situation performs an act which is in one sense voluntary and in another sense contra-voluntary. His act is voluntary, in the sense that it is the result of a considered choice between alternatives. It is contra-voluntary, in the sense that he would prefer not to have to choose any of these alternatives. In such cases we can say that the action, though voluntary, is *enforced.*

4.3423. Motives in **acting and motives** for **acting.** Let us go back to our example of an anarchist who deliberately throws a bomb at a ruler in a procession expecting to kill or injure the ruler, to kill or injure some of the spectators, and to break a number of windows in the neighbourhood. Let us suppose that the ostensible property of causing the death of the ruler is an attractive characteristic, that the ostensible property of causing death or injury to harmless spectators is a repelling characteristic, and that the ostensible property of causing windows to break is neither attractive nor repellant to the anarchist.

Suppose we were to ask: What was the anarchist's motive in throwing the bomb? Many people would say that it was the expectation of killing the ruler, and they would not mention the expectation of injuring innocent people. But, if the anarchist was a humane man, the expectation of injuring innocent people *was* a motive-component. It was so in precisely the same sense in which the expectation of killing the ruler was, and in a sense in which the expectation of breaking windows was *not*. It made the choice of the action more difficult; and its presence or absence would make a considerable difference to one's moral judgment on the agent's character. To omit all reference to it is like omitting all reference to the force of gravitation in the case of a balloon which is being moved upwards by stronger forces which overcome the gravitation.

In order to deal with this point we can distinguish between a person's motive *in* an action and his motive *for* the action. The former is the resultant of all the components both of attraction and of repulsion. The latter is the resultant of the components of attraction only.

We can now state the relations between intention and motive. The characteristics in respect of which an action is *intentional* will in general fall into three classes, viz. attracting, repelling, and neutral characteristics. The action is *motived* in respect of the first two of these, and not of the third. We say that the agent does the action *because* of the attracting characteristics and *in spite of* the repelling ones. We also say that he decides to *put up with* the repelling ones. What I have called his *primary intention* is to do an act having those characteristics which attract him. So his *primary* intention coincides with his motive *for* doing the act. His *secondary* or *collateral* intentions include both doing an act with certain characteristics which repel him, and doing an act with certain characteristics which leave him unmoved. Thus, e.g., the anarchist's *primary* intention is to kill the ruler, and it is the expectation of killing the ruler which is his motive *for* throwing the bomb. A part of his collateral or secondary intention is to kill or injure innocent spectators. But this is something which repels him and is a motive-component *against* throwing the bomb.

4.3424. Purity and mixture of motives. (1) Let us begin with the simple case of just two alternatives A and B, and let us suppose that the agent chooses A in preferance to B. Let x, y, and z be the components of preference for A to B and let u, v, and w be the components of preference for B to A. (There is no special significance in the fact that I have supposed that there are three of each kind. There might be any number of either kind.) Since the agent in fact chose A we know that the combination x, y, z, u, v, w was sufficient to constitute a resultant motive for choosing A in preference to B. But we can now raise the following questions. Was this combination *more than* sufficient? Suppose that the anti-components u, v and w had all been present but one or other of the pro-components x, y and z had been absent. Would the resultant motive still have been for choosing A in preference to B? In particular would any *one* of the pro-components x, y, and z have sufficed, in presence of the anti-components u, v, and w, to ensure that the resultant motive would be for choosing A in preference to B?

There are the following possibilities. (i) There might be one and only one of the pro-components, e.g. x, which would suffice, in presence of the anti-components u, v, and w and in absence of the other pro-components y and z, to give a resultant motive for choosing A in preference to B. We could sum this up by saying that x was *sufficient* and was *not superfluous*, and that y and z are *collectively insufficient* and *severally superfluous* in presence of x. (ii)

There might be several of the pro-components, e.g. x and y, each of which would suffice, in presence of the anti-components and in absence of the other pro-components, to give a resultant motive for choosing A in preference to B. We could sum this up by saying that x is *sufficient but superfluous in presence of* y, that y is *sufficient but superfluous in presence of* x, and that z is *insufficient and superfluous in presence of either* x *or* y. (iii) It might be that none of the pro-components x, y, or z, would have sufficed by itself, in presence of the anti-components u, v and w, to give a resultant motive for choosing A in preference to B. We could sum this up by saying that x, y and z are *severally insufficient* but *collectively sufficient*. It might still be the case that one or more of them was superfluous. E.g. it might be that the combination x, y would be sufficient. In that case we could add the statement that z is *superfluous in presence of* x, y.

Let us now apply these results. Suppose we are told that a person chose the alternative A of subscribing a sum of money to a hospital in preference to alternative B of buying a new car with it. Then, if we take the attractions of the car alternative as fixed, we can raise the following questions. (i) Was there just one *single* component of preference for the hospital alternative, or were there several compounded with each other? E.g. was the *only* component of preference the desire to relieve suffering, or was this compounded with the desire that the neighbours should see one's name in the subscription-list? When and only when there is just one component of preference for the alternative that is actually chosen we can say that the motive was *homogeneous* and that the choice was *single minded*. (ii) Suppose that the motive was heterogeneous. Then we can raise the following questions. (a) Was there one component which was both *sufficient and not superfluous* to ensure the choice that was actually made? E.g. would he *still* have subscribed to the hospital if he had believed that it would relieve suffering, even though he had *not* believed that this name would appear in the papers as a subscriber? And would he *not* have subscribed to the hospital unless he had believed that it would relieve suffering even if he had believed that his name would appear in the papers? If the answer to this question is in the affirmative we can say that there was a *governing* motive-component *alloyed with* other secondary ones. In that case we can say that the motive for the choice of A in preference to B was *monarchic*. (b) Suppose that the motive was not monarchic. Then we can raise this question. Were there several components, each of which was sufficient by itself and therefore superfluous if any of the others of them were present? E.g. would he *equally* have subscribed to the hospital in preference to buying the car (α) from the belief that it would relieve suffering, in the absence of any hope of notoriety, and (β) from the hope of notoriety, in the absence of any belief that it would relieve suffering? If so, we might say that the motive is *polyarchic*; since there are a number of components of pre-

ference, each of which might be said to "govern" as much as any of the others. (c) If the answer is in the negative, then none of the various components is individually sufficient, whilst collectively they are sufficient. In that case we can say that the motive is *cooperative*. We could then raise the question whether any selection from them would be sufficient or whether nothing less than the whole collection of them would suffice. On the second alternative we might say that the motive is *minimal*; since it contains no factor which is superfluous to account for the actual choice.

We can now summarise this as follows. The motive for choosing A in preference to B may be either *homogeneous* or *heterogeneous*. If it is *heterogeneous*, it may be either *monarchic* or *polyarchic* or *cooperative*. And if it is cooperative, it may be either *minimal* or *non-minimal*.

Now I think that the phrase "pure" or "unmixed" motive is used sometimes to mean a homogeneous motive, and sometimes to mean a motive which is heterogeneous but monarchic. In its strictest sense it should be confined to homogeneous motives. A motive which is either cooperative or polyarchic would certainly be called "mixed", and it is important to see that the phrase "mixed motive" covers these two very different cases.

The importance of these distinctions for moral judgment is obvious. Suppose that a person's motive in making a choice is homogeneous. Then he can be unreservedly praised for it if it is good, and unreservedly blamed for it if it is bad. Suppose that it is heterogeneous but monarchic. Then he can be praised if the governing motive is good, in spite of its being alloyed with other components which are bad or indifferent; and he can be blamed if the governing motive is bad, in spite of its being alloyed with other components which are good or indifferent. Suppose that it is heterogeneous and polyarchic, and that some of the sufficient components are good and others are bad. Then it is impossible to make any determinate judgment of praise or blame on the agent in respect of his motive. The same is true if his motive is heterogeneous and cooperative and contains good and bad components.

(2) We must now consider the case where there are more than two alternatives. Suppose that A is chosen out of the three alternatives A, B, and C. At this stage there enters a new possibility of mixture. Even if my motive for preferring A to B is homogeneous and my motive for preferring A to C is homogeneous, they may be quite different in kind. If so, my motive for choosing A will be heterogeneous. Suppose, e.g., that I have to choose between inviting A or B or C to be my guest at a College feast. I do not enjoy either A's or B's company, but I am under obligation to both of them. I enjoy C's company very much, but I am under no special obligation to him. Suppose I decide to invite A in preference to C simply because I want to repay my obligation to A. And suppose that I decide to invite A in preference to B simply because his conversation is slightly less boring than B's. Then my

motive for inviting A in preference to both B and C is heterogeneous. When one chooses an alternative out of several one's motive for that choice will be homogeneous if and only if one's motives for preferring it to each of the others are each homogeneous and are all of the same kind. By saying that they are all of the same kind I mean that they all arise through the excitement of the same conative-emotional disposition. Suppose, e.g., that I invited A in preference to B simply because I was under a stronger obligation to A than to B, and that I invited A in preference to C simply because I was under an obligation to A and under no obligation to C. Then my motive would be homogeneous, for the only conative-emotional disposition concerned would be my sense of duty. Similarly, one's motive for choosing a certain alternative out of several cannot be monarchic unless there is (a) a governing motive for preferring it to each of the other alternatives, and (b) all these governing motives of preference are of the same kind.

4.3425. Motives of different orders. A being who is capable of reflexive cognition not merely has motives but is capable of *reflecting upon* his own motives and appraising them. When he does so he may find that he approves morally of some of them and disapproves morally of others. He may remember that acting from certain motives in the past has led to unpleasant consequences. He may notice that some of his conative-emotional dispositions are specially weak or specially difficult to excite, and that others are specially strong or easily excitable. Suppose that he has acquired such a system of knowledge and beliefs about his own conative dispositions by reflecting on his own past life. He will now enter on any deliberation in a different state from that of a being who is incapable of reflecting on his own motives, and in a different state from that in which he would himself have entered on a deliberation before he had acquired this system of reflexive dispositions. The reflexive dispositions will be both cognitive and emotional. We may divide our conative-emotional dispositions into those of the first-order and those of the second-order. They will be of the second-order if the conations and emotions to which they give rise are directed to one's own conative-emotional dispositions. They will be of the first-order if the conations and emotions to which they give rise are directed to something other than one's own conative dispositions. E.g. the dispositions to feel pity or sexual desire are of the first-order. But a person may have acquired a disposition to feel contempt for feeling pity or a disposition to feel disgust at feeling sexual desire. If so, these dispositions are of the second-order.

Let us suppose that such second-order dispositions have been formed. Then a certain alternative may evoke a first-order component of attraction through some feature in it which excites sexual desire. And the very same

alternative may evoke a second-order component of repulsion because the property of exciting sexual desire excites the second-order disposition to feel disgust at feeling sexual desire.

It is true that in any deliberation we have to use such conative-emotional dispositions as we have at the time, and that these may be taken as a fixed factor for that deliberation. But it is important to notice that they are not all of the first-order. For this implies that beliefs about one's own first-order motives may give rise to second-order motives which may have a profound influence on one's choice.

It is also important to notice that a person's present knowledge and beliefs about his own first-order dispositions may lead him to make certain resolutions about his own procedure in future deliberations about certain subjects. He may have learned from experience that a certain conative disposition will almost inevitably pass into action if in any future deliberation he pays more than a fleeting attention to certain aspects of certain alternatives. He may also have learned from experience that such actions have been unsatisfactory in themselves or unfortunate in their consequences. Or he may regard this particular conative disposition as contemptible or disgusting when he reflects on it in a cool hour. Therefore he may resolve that, in future deliberations in which this conative disposition will be involved, he will attend mainly to the characteristics which excite other conative dispositions. That is, he may decide now on the basis of his knowledge of his own nature and his own past actions, to try to distribute his attention in a certain way in future deliberations on certain subjects. The effects of this resolution may persist and may modify the outcome of such deliberations in future.

This is just one more instance of the enormous importance of the fact that men have reflexive cognition, conation, and emotion.

4.3426. Mistaken motives. We often talk of a person acting from "mistaken motives". We often say that his "ostensible motive was *X* but his real motive was *Y*". I shall now try to clear up these statements.

The first point is to distinguish between being mistaken *in* one's motives and being mistaken *about* one's motives. The first is a non-reflexive false belief, and the second is a reflexive false belief.

A person is mistaken *in* his motives if (a) he mistakenly believes an alternative to have certain characteristics which it does not in fact have, and (b) this false belief excites a conative-emotional disposition and makes the alternative attractive or repulsive to him. (A variant would be if the characteristic really were present, but in a different form or to a different degree from that in which the agent believed it to be present.) Suppose, e.g., that I mistakenly believe that a certain road, which would be the most convenient for my purpose, is under repair. And suppose that I am moved by this false belief

to take another road which I should not otherwise have taken. Then I am mistaken *in* my motives.

We can now deal with mistakes *about* one's own motives. If I believe myself to be attracted in *some way or other* and to be repelled in *some way or other* by a certain alternative which I am contemplating, I am hardly likely to be mistaken. But suppose I go further and attempt to say *what* are my motives for this mixed attitude of attraction and repulsion. Then I may be mistaken in at least four ways, two of which are about the cognitive factors and two about the conative-emotional factors. (1.1) I may fail to notice some of my own beliefs about the alternative. I may notice my belief that it has X and my belief that it has Y, but may fail to notice my belief that it has Z. Yet I may in fact believe that it has Z, and this belief may in fact be exciting a strong feeling of attraction or repulsion. (1.2) Suppose that I am aware of all my beliefs about the characteristics of the alternative. I may think that my belief that it has X is exciting attraction or exciting repulsion, and that my belief that it has Y is not doing so. But really the opposite may be the case. (2.1) There is no reason to suppose that each of us is aware of all the conative-emotional dispositions which he in fact has. Therefore my beliefs about the characteristics of an alternative may be exciting some conative-emotional disposition which I am not aware of possessing. (2.2) Suppose that my belief that the alternative has the characteristic X is in fact exciting a certain disposition D which I am aware of having. I may mistakenly believe that it is exciting, not D, but another disposition D' which I am also aware of having. It is possible to be mistaken about one's own motives in any of the 15 different ways in which these 4 fundamental ways of being mistaken may occur separately or in combination.

The most common cause of mistakes about one's motives is probably the following. As we saw in discussing second-order motives, there are some conative-emotional dispositions which a person likes to think of himself as possessing, and there are others which he dislikes to think of himself as possessing. An example of the former might be the desire to improve other people's characters, and an example of the latter might be taking pleasure in the sufferings of others. Now suppose that a certain alternative is under consideration which one knows would hurt or humiliate a person whom one dislikes and which one believes would be likely to improve his character. One will tend to ignore all knowledge and belief about the aspects of the alternative which appeal to the disposition of cruelty, and to concentrate attention on those parts of one's knowledge and belief which appeal to the disposition to improve other people's characters. One may then persuade oneself that the disposition towards cruelty is not influencing one's decision at all. Yet it may have been much the most important conative motive-factor. And one's ignored knowledge that the action will hurt and humiliate may have been

much the most important cognitive motive-factor. The desire to produce moral improvement may hardly have been excited at all, and the belief that the action will improve the other man's character may have been a mere idle accompaniment.

There is another point to be made before leaving this part of the subject. I have been assuming that the agent really has made some analysis of the various alternatives and really does believe that each of them has such characteristics. I have supposed only that he may be unconscious of some of these beliefs, and that he may be mistaken as to which belief is exciting which of his conative dispositions. But there is another possibility. Suppose that a person is deliberating about alternative courses of action, and that he has not much time for deliberation and not much power of analysis. Then he may just have a vague impression of *A*, *B*, and *C* as each partly attractive and partly repulsive, and a vague impression that on the whole *A* is more attractive or less repulsive than *B* and *C*. Perhaps an external observer, or the agent himself on subsequent reflexion, could point out the ostensible feature in *A*, *B*, and *C* which are in fact making them attractive or repulsive. But, if the agent himself did not distinguish them at the time when he made his choice, we can hardly say that he was moved by the belief that they were present. And, unless we can say this, we cannot, strictly speaking, say that he was acting from motives at all. What I have been saying in this paragraph should be compared with what I said in paragraph 4.13 about motived and unmotived *emotions*.

4.3427. First-hand and second-hand motives. In paragraph 4.15 I drew a distinction between first-hand and second-hand emotions. I think that a similar distinction must be drawn among motives. When we deliberate about alternatives we may really be inspecting each alternative and trying to distinguish and contemplate its intrinsic qualities, its non-causal relationships, and its probable consequences. If we do this, and if our decision is influenced by our beliefs about these characteristics I say that we are deliberating and choosing at *first-hand* and that our motives are *first-hand motives*.

But very often we are not doing this. We are merely being moved by the emotions called up by certain *names*. We say, e.g., "that alternative would be ungentlemanly, that would be unpatriotic, that would be undemocratic", and so on. These labels may be attached to the alternatives on the most trivial and external grounds. But, once they are attached, the associations of the names exert a strong influence over us. In such cases I say that we are deliberating and choosing at *second-hand* and that our motives are *second-hand motives*. When a person says that he is "acting on principle", or that so-and-so would be "against his principles", it is often merely a sign that they will not take the trouble to analyse the actual situation or to reflect on the relevant features of the various alternatives.

4.3428. Motive and intention. Suppose that a certain person has done a certain act, e.g. taken steps to bring about the prosecution of a thief. Then we can be quite sure that the act was intentional and that the agent had some motive or other for doing it. For such an act is obviously a considered act, and it is most unlikely that the agent had not *some* reason for doing it. Again, one can often be practically certain about at least part of the agent's intention in doing the act. It is practically certain in our example that part of the agent's intention was to bring about the punishment of the thief by the authorities. But one may be uncertain about many details of the intention. E.g. one may be uncertain whether the agent foresaw the effects that his action would have on the thief's wife and family, and so one may be uncertain whether it was part of his intention to bring suffering on them. And one may be quite uncertain about the agent's motives *in* acting and *for* acting as he did. E.g. he may have been attracted by the thought that justice would be done and society protected, and repelled by the thought that the thief and his wife and family would suffer; and he may have acted for the former motive and in spite of the latter. Or he may have been quite indifferent to justice and the protection of society, but attracted by the thought that the thief would suffer because the thief had done him an injury, and repelled by the thought that the thief's wife and family would suffer. If so, his motive for his action was desire for revenge on the thief; and he acted for this in spite of his humane desire to spare the thief's wife and family. When one is uncertain of the agent's motives for acting one is *ipso facto* uncertain as to what was his *primary* intention, even if one knows for certain every feature in respect of which the act was intentional.

These facts are important in connexion with the distinction between purely legal and specifically moral judgments about a person's actions and character. In the main the law is concerned with a person's overt actions and his intention as a whole, and not with his motives or with the associated question as to which part of his intention was primary, and which part secondary or collateral. On the other hand, morality is very much concerned with motives and primary intentions. Thus, an act which is legally criminal may be morally right and even a moral duty; and one which is legally innocent or even a legal duty may be done for such a motive and with such a primary intention that it is morally wrong.

4.343. Action; its antecedents and its consequences

So far I have taken the notion of "an action" as sufficiently familiar. I want now to analyse it a little, and in particular to consider the distinction between an action itself and its antecedents and its consequences.

Actions may be divided into mental and bodily; though of course all mental actions have bodily accompaniments and all bodily actions have mental ac-

companiments. An example of a mental action is performing a calculation or making an inference. An example of a bodily action is singing a song or lifting a weight. For the present we will confine our attention to bodily actions.

We describe actions by means of phrases which contain what grammarians call active verbs. Thus, if you ask a person: "What were you doing at such and such a time?" you expect some such answer as: "I was singing a song" or "I was lifting a weight". It will be noticed that these two answers describe actions of two fundamentally different kinds, and that there is a grammatical difference in the sentences which corresponds to the difference in the kinds of action. In the sentence "I was singing a song" the verb is intransitive and the word "song" is what grammarians call an "internal accusative". In the sentence "I was lifting a weight" the verb is transitive and the word "weight" is an external accusative. The corresponding difference is this. When a person sings a song his primary intention is simply to make a series of harmonious sounds and not to make alterations in other physical objects. When a person lifts a weight his primary intention is to cause a certain kind of change in a certain external physical object. An intermediate case would be if the answer was "I was cutting my nails". There the primary intention is to cause a certain change in a certain physical object, but that object is a part of the agent's own body. Cutting one's nails and lifting a weight may be described as *transitive actions* or *transactions*. They may be divided into *reflexive*, e.g. cutting one's nails, and *non-reflexive*, e.g. lifting a weight. Singing a song or dancing a hornpipe may be described as *non-transitive* actions. Of course non-transitive actions nearly always do produce certain changes in the external world; e.g. dancing a hornpipe may cause the floor to shake and may bring down the ceiling. But the agent's primary intention in doing a non-transitive action is not to produce changes in his own or other bodies, whilst that *is* his primary intention in doing a transitive action.

Now suppose that I perform a transitive action, such as writing a letter. This will go on for some time, say ten minutes. We can distinguish the mental and the bodily aspects of this transaction. On the mental side there is the process of deciding to write and setting oneself to do so and keeping oneself at doing so in spite of distractions and so on. Let us call this *initiation* and *perseverance*. There are also the processes of thought which accompany and control the writing. Finally, there are the sensations which come from the muscles, joints, skin, etc. as they are used in writing. On the bodily side are certain changes in the brain and nervous system. These determine certain changes in the muscles and joints. These in turn determine overt movements in the eyes and fingers. The fingers grasp the pen and keep it moving over the paper, producing a series of conventional signs intended to convey to the recipient of the letter certain ideas.

Now the causation in all this is very complex, and it runs backwards and

forwards. All that the mental processes of initiation and perseverance and thinking can *directly* accomplish is to make and keep up certain changes in the brain, of which the agent is quite unaware. If and only if the nerves are in proper order, the consequences of this will be that certain changes are transmitted through them to the muscles. The agent is also quite unaware of these processes of transmission. If and only if the muscles are in order, they will contract in certain ways and cause the fingers to move the pen in the way which the agent desires. So far I have described only the causal chain formed from the mind to the fingers and the pen. But there is a causal chain backwards from the fingers and the eyes through certain nerves to the brain and the mind which is equally essential. The sight of what one has just written in part determines the thought of what one will write next, and this thought determines the movements of muscles, fingers, and pen in the immediate future.

For these reasons it is not at all easy to draw a hard and fast line between the action itself, its antecedents, and its consequences. Everyone would admit that the process of deliberating whether to write the letter or not, and of deciding to write it, was an antecedent and not a part of the action. And everyone would admit that the receipt of the information which is contained in the letter is a consequence and not a part of the action. But I think that it would be ridiculous to identify the action with the overt movements of the fingers and pen; and say that the changes in the muscles, nerves and brain and the mental processes of initiation, perseverance, and directed thinking were mere *antecedents*. Similarly, I think it would be absurd to call the visual and muscular sensations, which arise as one writes and which in part determine one's thoughts and through them one's next movements, mere *consequences* of the action.

I think that the point is made clear by the following two examples. Suppose that a person has rheumatism and that the actual movements which he makes in dancing produce painful sensations. Then it would be natural to say that the act of dancing *is* painful for him, and it would be pedantic to say that the act itself is not painful but causes painful sensations. But suppose that a person has a duodenal ulcer, and that eating is following some hours later by severe pain in the stomach. Then it would be natural to say that eating *causes* painful sensations, and misleading to say that the act of eating is painful.

I shall therefore include as *parts* of an act, and not mere antecedents or consequences, all the factors that I have mentioned. I shall include the mental processes of initiation, perseverance, and relevant thinking, and the sensations which arise directly from the bodily movements which are part of the act. I shall include the bodily processes in the brain, the nerves, the muscles, and the joints as well as the overt movements of the fingers, eyes, etc. We could distinguish the bodily factors in an act into *overt movements* and *intra-*

somatic processes. And we must realise that each phase of the mental part of an act helps to determine indirectly the next phase of the overt bodily movement, and each phase of the overt bodily movement helps to determine indirectly the next phase of the mental part of the act. So one has a complex interweaving of cause and effect *within* any act; quite distinct from the external effects *of* that act, such as the production of intelligible marks on paper.

When an act is described by means of an intransitive verb and an internal accusative, e.g. as "dancing a hornpipe" or "singing a song", the description is simply by reference to the nature of the overt bodily movements themselves or their immediate consequences in the way of sounds. Suppose that an act is described by means of a transitive verb and an external accusative, e.g. as "writing a letter". Then the description is by reference (a) to the immediate consequences which are presumed to have been an essential part of the agent's primary intention, viz. intelligible marks being made on a previously clean surface for the purpose of recording or conveying ideas, and (b) to the nature of the overt bodily movements and the instrument used. E.g. "writing" differs from "scribbling" in its intended immediate effects. It differs from "printing" or "typewriting" in the nature of the overt bodily movements and the instrument.

4.344. Means and end

We often say that a person chooses to do a certain act as a *means* to a certain *end*. E.g. he writes as a means to making money. Then we distinguish between the *proximate* and the *remoter* ends for which an act is done. E.g. he may write as a means to make money, and he may want to make money as a means to getting married. In that case writing is chosen as a means to the proximate end of making money and as a means to the remoter end of getting married. In a series of terms which starts with an act chosen as a means to a proximate end, which is itself desired as a means to another proximate end, and so on, there will eventually be a term which is desired, not as a means to some further end, but for some other reason or for no reason. This term is called the *ultimate* end for which the act at the beginning of the series is done. That act may be called the *initial means*, and the intermediate term may be called the *intermediate means*, to the ultimate end for which the act was done. The term immediately before the ultimate end may be called the *proximate means* to the ultimate end.

Now suppose that a person desires something E, either for no reason or for some other reason than as a means to something else. And suppose that he sees that no act which is within his power would immediately produce E. Then he must cast about for a series of events having the following properties. (i) The last term of it must be such that it would immediately produce E. Let us

call this term D. (ii) The first term of it must be some act which is or will be in his power to do if he chooses. Let us call this A. (iii) Every term in the series except A and E must be a *causal descendent* of A and a *causal ancestor* of E. By calling Z a "causal descendant" of W I mean that either W is the immediate cause of Z or that there is a series of terms, e.g. X and Y such that W is the immediate cause of X and X is the immediate cause of Y and Y is the immediate cause of Z. Now in some cases one will see that there are several alternative series each of which would fulfil all these conditions. They might be symbolised, e.g., by

$$\begin{array}{llllll} A & \rightarrow B & \rightarrow C & \rightarrow D & \searrow & \\ A' & \rightarrow B' & \rightarrow C' & \rightarrow D' & \rightarrow E \\ A'' & \rightarrow B'' & \rightarrow C'' & \rightarrow D'' & \nearrow & \end{array}$$

(There is of course no need for there to be the same number of terms in each series.) Here, A, A', and A'' represent alternative acts, any one of which is in the agent's power to do if he chooses. The assumption is that if the agent does any of these acts it will initiate a causal chain leading up to the production of E; whilst, if he does none of them, there is no reason to believe that E will be produced. I shall call any such series a *way to E*.

The following remarks are worth making about such series.

(1) Two such series might begin to coincide before the last term E. E.g. you might have the alternatives

$$\begin{array}{llll} A & \rightarrow B & \rightarrow C & \searrow \\ & & & D \rightarrow E \\ A' & \rightarrow B' & \rightarrow C' & \nearrow \end{array}$$

(2) The intermediate term might or might not be, or contain, further acts on the part of the agent. Sometimes one initiates such a series by an act and then it goes on without further interference. This happens, e.g., if one posts a letter. Sometimes the act which initiates such a series merely prepares the conditions in which one will be able to perform another act which will be a term in the series. This happens, e.g., if one posts a letter in order to arrange for an interview with a person.

(3) In real life you can never be certain that a series which you initiate by an act A will lead to the ultimate end E, as a means to which you did A. Take any intermediate term C, e.g.; you cannot be sure that if you do A, C will arise as a causal descendant of it. And you cannot be sure that if C is a causal descendant of A it will be a causal ancestor of E. In all such cases we can only make more or less probable conjectures.

(4) Suppose a person desires A as a means to *B*, which he desires as a means to *C*, and so on. *A*, *B*, etc. will have plenty of other characteristics beside this. So he may also desire *A*, *B*, etc. for other reasons. Or, again, he may be at the same time attracted by *A*, in so far as he believes it to be a means to *B*, and repelled by it in so far as he believes it to have some other characteristic, e.g. to be painful or to be morally wrong.

(5) Sometimes the same thing, or different things of the same kind, are desired on many occasions as means to many different ends. The most obvious example is money. Another example is the acquisition and retention of power by the political party of which one is an active member. In such cases it very often happens that eventually that thing or things of that kind begin to be desired directly and not simply as means to something else. This is what happens in the case of a miser or a fanatical partizan. I shall call this process *finilisation*, since it consists in coming to desire as an *end* what was originally desired only as a means. It may be compared with *fetishism*, which consist in transferring to a symbol the emotions which are normally directed to the thing symbolised. The two processes are often associated.

(6) Suppose that a person is considering whether or not to try for a certain ultimate end *E*. He cannot confine his attention to the attractive and repulsive features of *E*, for he cannot get *E* without initiating a series of means. Suppose he believes that there are several alternative ways *S*, *S'*, etc. each of which would start with a different act within his power and would probably lead to the end *E*. Suppose he believes that these are the only alternatives open to him which would lead to *E*. Then he will have to consider all the following points. (i) With regard to each alternative initial action and each intermediate term in each series of means he will have to ask himself how far it is attractive or repulsive on other grounds than that of leading to *E*. (ii) With regard to each intermediate term in each alternative series he will have to ask himself (a) How likely is it that this state of affairs will be produced if I initiate this series? and (b) How likely is it that, if it should be produced, it will be a causal ancestor of *E*? These considerations may point in opposite directions. E.g. the series *S* may be on the whole more attractive or less repulsive than the series *S'*. But I may be more likely to produce *E* if I initiate *S'* than if I initiate *S*.

Now it may be that all the alternative series of means which I could initiate in order to produce *E* are either so repulsive on the whole or so uncertain in their outcome that I should decide not to try for *E*. There might be an alternative end *E'* which is less attractive to me than *E* But perhaps *E'* could be more likely to be effective than any which I could initiate in order to get *E*. In that case I might prefer to try for *E'* rather than to try for *E*. We express this by saying that my *antecedent preference* is for *E* but my *consequent preference* is for *E'*. E.g. most people would have an antecedent preference for a larger income over a smaller one. But suppose that all the alternative

series of means which one could initiate to increase one's income involved either dishonesty, or excessive drudgery, or living in an unhealthy climate. And suppose that there was at least one honest, easy, and healthy way of earning a smaller income. Then one's consequent preference might be for the smaller income earned in one of those ways.

Suppose that one has finally decided to try for a certain end E by one or other of the alternative ways which one can initiate. One will then have to consider which way to take. In doing this one will have to estimate the relative efficiency of the various ways for bringing about the end, and the relative attractiveness or repulsiveness of the various ways on other grounds than their efficiency. The resultant attractiveness of each way will depend jointly on these two considerations; attractiveness in other respects must be discounted for inefficiency, and efficiency must be discounted for unattractiveness in other respects.

(7) The vast majority of our considered acts are done primarily as means, and often this is the only motive for doing them. But we may have additional motives for doing an act, e.g. the experiences which are part of it may be pleasant. And some considered acts are not done as means to anything but simply because the agent likes doing such acts. Many acts of bodily exercise, e.g. swimming, dancing, etc. are of this nature. But nearly all transitive acts, e.g. writing, cutting, lifting, etc. are done primarily as means.

4.345. Desiring X *as part of a desired whole* W.

A person may desire, for one reason or another, to bring into existence a thing or a state of affairs which consists of a number of interrelated simultaneous parts or successive phases. Such a whole may *consist of* a number of interrelated acts, e.g. dancing a dance or playing a game of tennis. Or it may be the *product of* a number of interrelated acts, e.g. a picture or the series of sounds which constitute a tune played on a violin. We will consider these two cases in turn.

(1) Suppose that a person wants to produce a complex series of interrelated actions, e.g. to dance a certain kind of dance. Then his reason for desiring at any stage to make such and such a movement is the fact that he believes this movement to be the appropriate phase in the dance at that moment and that he desires to carry out the dance as a whole. Suppose again that a person is producing a ballet. Then his reason for wanting a certain dancer to make a certain movement at a certain moment and for wanting other dancers to make certain other movements at the same moment is that he desires a certain pattern of rhythmic bodily movements and that such and such movements by such and such dancers are the appropriate items in that pattern at that time. In both these examples the reason for desiring that a certain act shall be done at a certain moment is the belief that it is the appropriate phase or part of a certain

complex whole of action which is desired.

It is important to notice that the relevant relationship is *not* that of means to end. The successive movements of a single dancer in a *pas seul* or the simultaneous movements of the various dancers in a ballet are related, not as cause to effect, but as parts to whole. Nevertheless there are certain analogies between desiring X as a part of a desired whole Y and desiring X as a means to a desired end Y. The analogies are these. (i) In both cases one part of a person's reason for desiring X is that he desires something else Y and that he believes X to stand in a certain relation to Y. (ii) Though the relation is different in the two cases, there is the following important resemblance. In each case it is believed that Y will not come into being unless either X or some one of a limited number of alternatives to X is done now. A whole depends for its existence on the simultaneous or successive existence of all its parts or phases. And an end depends for its existence on the previous occurrence of those events which are its causal ancestors. The nature of the dependence is very different in the two cases, but there is dependence in both.

(2) Let us next consider a whole which does not consist of interrelated acts but is a product of a number of interrelated acts. Here we have a combination of the two kinds of reason. Each act is done as a means to a desired end, and each such end is desired as a part of a desired whole. The violin-player desires that a certain whole, which will consist not of acts but of sounds, shall be produced. He performs each act of bowing and fingering the strings as a means to producing the sound which he thinks will be the appropriate phase at that moment in the series of sounds which he desires to produce.

4.346. Subordinate and ultimate desires

When X is desired either as a means to Y or as a part of Y we can say that the desire for X is *subordinate to* the desire for Y. For a person would not be attracted towards X by the belief that it was a means to Y or that it was a part of Y unless he were attracted towards Y.

There are, however, other instances of subordinate desires. Suppose that a person is attracted towards a certain alternative because (i) he believes that it has a certain property P and that in consequence of having P it will have a certain other property Q, and (ii) Q is for him an attracting characteristic. Then his desire for that alternative as having P is subordinate to his desire for it as having Q. E.g. suppose that a person is asked a question and he considers the alternatives of giving a true answer or giving one or other of several false answers. He may be attracted towards giving a certain answer because he believes that it is *true* and that to give a true answer is as such *morally right*, and because he desires to do what is morally right. In that case his desire to do that act because it is a true answer is subordinate to his desire to do what is morally right. If he did not believe that to answer truly is as such morally

right, or if he did not desire to do what is morally right, his belief that a certain answer would be true might not attract him towards giving that answer.

If a desire is not subordinate to any other desire it may be called an *ultimate* desire. It is plain that for any given person at any given moment there must be desires which are ultimate. It does not follow that the same kinds of desire will be ultimate for a person throughout the whole of his life. Nor does it follow that there are certain kinds of desire which are ultimate for all men. Nevertheless it is probably true that there are certain kinds of desire which are ultimate for all men at all times. E.g. if a person thinks that a certain experience would be pleasant, that is an ultimate reason for desiring it; and if he thinks that it would be unpleasant, that is an ultimate reason for trying to avoid it. Of course a reason may be *ultimate* without being *sufficient* or *conclusive*. A person may for other reasons prefer to forego an experience which he thinks would be pleasant or to endure an experience which he thinks will be unpleasant. But, if so, that is because the alternatives have other properties which evoke other desires either of the same or of different kinds. E.g. the person may believe that if he allows himself the pleasure of eating a certain kind of food he will have to endure the *pains* of indigestion. Or he may believe that to have a certain experience which would be *pleasant* would be or would involve doing something morally wrong; and he may have an ultimate aversion to doing what is morally wrong.

4.347. Pluralism v. monism of ultimate desires

On the face of it there seem to be a number of different kinds of ultimate desires which all or most men have. E.g. the desire to get pleasant experiences and to avoid unpleasant ones, the desires to gain an exercise power over others, the desire to do what is right and to avoid doing what is wrong; and so on. Very naturally philosophers have tried to reduce this plurality.

They have tried to show that there is one and only one kind of ultimate desire, and that all other desires which seem at first sight to be ultimate are really subordinate to them. I shall call the view that there really are several different kinds of ultimate desire *pluralism of ultimate desires*; and I shall call the view that there is really only one kind of ultimate desire *monism of ultimate desires*. Even if a person were a pluralist about ultimate desires he might hold that there were certain important features common to all the different kinds of ultimate desire.

Now much the most important theory on this subject is that all kinds of ultimate desires are *egoistic*. This is not in itself a monistic theory. For there might still be several irreducibly different kinds of ultimate desire even if they were all egoistic. Moreover, there might be several irreducibly different senses of the word "egoistic"; and some desires might be egoistic in one sense and some in another, even if all were egoistic in some sense. But the theory often

takes the special form that the only kind of ultimate desire is to prolong and to get pleasant experiences and to cut short and avoid unpleasant experiences. This *is* a monistic theory. I shall call the wider theory *psychological egoism* and this special form of it *psychological hedonism*. I shall now discuss these two theories in turn.

4.3471. Psychological egoism. I shall begin by enumerating all the kinds of desire that I can think of which are undoubtedly "egoistic" in one sense or another. (1) Everyone has a special desire for the continued existence of himself and a special dread of his own cessation. This may be called the *desire for self-preservation*. (2) Everyone desires to acquire and prolong experiences of certain kinds and to avoid and cut short experiences of certain other kinds, because the former are pleasant and the latter unpleasant. This may be called the *desire for one's own happiness*. (3) Everyone desires to acquire and to keep certain mental and bodily powers and dispositions and to avoid or get rid of certain others. In general he wants to be a person of a certain kind and wants not to be a person of certain other kinds. This may be called the *desire for self-culture*. (4) Everyone desires to feel certain kinds of emotion towards himself and his own powers and dispositions and not to feel certain other kinds of reflexive emotion. This may be called the *desire for self-respect*. (5) Everyone desires to acquire and to keep for himself the exclusive use of certain material objects or the means of buying and keeping such objects. This may be called the *desire to acquire and to keep property*. (6) Everyone desires to acquire and to exercise power over others, so as to make them do what he wishes regardless of whether they wish it or not. This may be called the *desire for self-assertion*. (7) Everyone desires that other persons shall believe certain things about him and feel certain kinds of emotion towards him. He wants to be noticed, to be respected by some, to be loved by some, and so on. Under this head come the *desire for self-display*, for *affection*, and so on. (8) Some desires, which are primarily concerned with other things or persons, either would not exist at all or would be very much weaker or would take a different form if it were not for the fact that those things or persons already stand in certain relations to oneself. I shall call such relationships *egoistic motive-stimulants*. The following are among the most important of these. (i) The relationship of ownership. If a person owns a house or a wife he feels a much stronger desire to improve the house or to make the woman happy than if the house belongs to another or the woman is married to someone else. (ii) Family-relationships. A person desires the well-being of his own children much more strongly than that of other children. (iii) Relations of love and friendship. A person desires strongly to be loved and respected by those whom he loves. He may desire only to be feared by those whom he hates. And he may desire only mildly to be loved and respected by those to

whom he is indifferent. (iv) The relationship of being fellow-members of an institution to which one feels loyalty and affection. An Englishman will be inclined to do services to another Englishman which he would not do for a foreigner; and an Old Etonian will be inclined to do services to another Old Etonian which he would not do for an Old Harrovian.

I think that the above is a reasonably adequate list of motives which could fairly be called egoistic in some sense or other. The next business is to try to classify them and consider their mutual relationships.

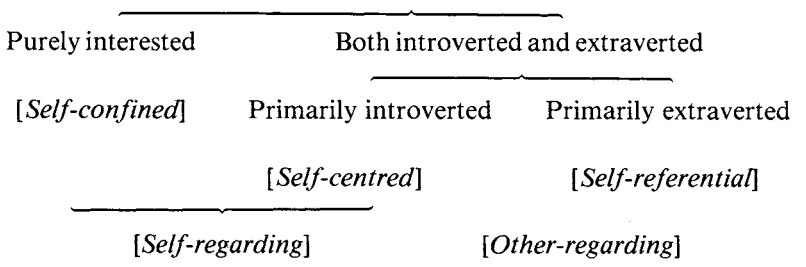
(1) In the first place we may ask ourselves: Which of these motives could act on a person if he had been the only person or thing that had ever existed? The answer is that he could still have had desires for *self-preservation*, for *his own happiness*, for *self-culture*, and for *self-respect*. But he could not, unless he was under the delusion that there were other persons and things, have desires for *property*, for *self-assertion*, for *self-display*, or any of the desires which presuppose family or other relationships. I shall call those desires, and only those, which could be felt by a person who knew or believed himself to be the only existent in the universe *self-confined*.

(2) Any motive which is not self-confined may be described as *extraverted*, for the person who has such a desire is necessarily considering, not only himself and his own qualities, dispositions, and states, but also some other thing or person and its relations to himself. It will also be *introverted*, if it is egoistic; since the person who has such a desire will also be considering *himself* and his relations to this other person or thing. Thus a self-confined motive is purely introverted, whilst a motive which is egoistic but not self-confined is both introverted and extraverted. Now we may divide motives which are both introverted and extraverted into two classes, according as the primary emphasis is on the former or the latter aspect. Suppose the person is concerned primarily with himself and his own acts and experiences, and is concerned with the other person or thing only or mainly as an object of such acts and experiences or as the other term in a relationship to himself. Then I shall call the motive *self-centred*. I shall use the term *self-regarding* to include both motives which are self-confined and motives which are self-centred. Under the latter head come the desires for *property*, for *self-assertion*, for *self-display*, for *affection*, and so on.

(3) We come finally to desires which are both introverted and extraverted, but when the primary emphasis is on the other thing or person and its states. Here the relationship of the other person or thing to oneself acts as a strong egoistic motive-stimulant, but one's primary desire is that the other person or thing shall be in a certain state. I will call such desires *other-regarding*. A desire which is other-regarding but involves a self-referential motive-stimulus may be described as *self-referential*. The desire of a mother to render services to her own children which she would not render to other children is an example

of a self-referential but other-regarding desire. So too is the desire of a man to inflict suffering on an enemy.

The above classification may be summarised as follows:



I shall now say something about the interrelations of these various kinds of egoistic desire.

(1) It is obvious that self-preservation may be desired as a necessary condition of one's own happiness, since one cannot prolong or acquire pleasant experiences unless one continues to exist. So the desire for self-preservation *may* be subordinate to the desire for one's own happiness. But it seems pretty clear that it is also an independent desire. It seems that a person often desires to go on living even when there is no prospect that the remainder of his life will contain a balance of pleasant over unpleasant experiences.

(2) It is also obvious that property and power over others may be desired as means to self-preservation or happiness. So the desire to get and keep property, and to get an exercise power over others, *may* be subordinated to the desire for one's own happiness. But it seems fairly certain that the former desires are sometimes independent of the latter. Even if a person begins by desiring property or power only as a means (and it is very doubtful whether we do always begin in this way) it seems that he often comes to desire them for themselves and to sacrifice happiness, security, and life itself for them. Any miser and almost any dictator provides an instance of this.

It is no answer to this to say that a person who desires power or property enjoys the experiences of exercising power and amassing property, and to argue that therefore his ultimate desire is to give himself these pleasant experiences. The premiss is true, but the argument is self-contradictory. The experiences of exercising power and amassing property are pleasant to a person only in so far as he desires power or property. This kind of pleasant experience presupposes desires for something other than pleasant experiences, and therefore the latter desires cannot be derived from desire for that kind of pleasant experience.

Similar remarks apply to the desire for self-respect and the desire for self-

display. If one already desires to feel certain emotions towards oneself or to be the object of certain emotions in other people, the experience of having those emotions or believing that others have them will be pleasant because it will be the fulfilment of a pre-existing desire. But this kind of pleasure presupposes the existence of these desires, and therefore these desires cannot be reduced to the desire for this kind of pleasure.

(3) Although the various kinds of egoistic desire cannot be reduced to a single ultimate desire, e.g. the desire for one's own happiness, they are very much mixed up with each other in many cases. Take, e.g., the special desire which a mother feels for the health, happiness, and prosperity of her own children. This is predominantly an other-regarding but self-referential desire. The mother is quite directly attracted by the thought of her child as surviving, and having good dispositions and pleasant experiences, and being the object of love and respect to other people. She is quite direclty repelled by the thought of him dying, or having bad dispositions and unpleasant experiences, or being the object of hatred and contempt to other people. The desire is therefore other-regarding. It is self-referential because the fact that it is *her* child and not another's acts as a powerful motive-stimulant. She would not be prepared to make the same sacrifices for the welfare of a child which was not her own. But this self-referential other-regarding motive is almost always mixed with other motives which are self-regarding. One motive which a woman has for wanting her son to be happy and respected is the desire that other women should envy her as the mother of a happy, healthy and respected son. This motive is subordinate to the self-centred desire for self-display. Another motive which she might have is the dislike of being burdened with the trouble and expense of an ailing, unhappy and unsuccessful son. This motive is subordinate to the self-confined desire for one's own happiness. But, although the self-referential other-regarding motive is almost always mixed with motives which are self-centred or self-confined, we cannot plausibly explain the behaviour of certain mothers towards their children without bringing in the other-regarding motive.

Precisely similar remarks apply to the case of a man who desires to inflict suffering on another person because he hates the latter. His motive is primarily self-referential but other-regarding. This is shown by the fact that he may be prepared to sacrifice his happiness and even his life to securing revenge on his enemy. But this motive is often mixed with self-regarding motives. He may believe that the death or imprisonment of his enemy is essential to his own safety. In that case the self-regarding motives of desire for self-preservation and for one's own happiness come in. He may know that the experience of seeing his enemy suffer will be pleasant, and part of his motive may be the desire to give himself the pleasant experience of gloating.

We can now consider the various forms that psychological egoism might take.

(1) The most rigid form is that all human motives are ultimately egoistic, and that all egoistic motives are ultimately of one kind. If all the egoistic motives are ultimately of one kind, the only kind which could be suggested with any plausibility is the desire for one's own happiness. Thus this theory amounts to saying that the only ultimate motives are self-confined, and that the only ultimate self-confined motive is the desire for one's own happiness.

I have already tried to show by examples that this is false. E.g. among self-confined motives the desire for self-preservation cannot be reduced to the desire for one's own happiness. Then again there are self-regarding motives which are self-centred but not self-confined, such as the desire for power over others. And, finally, there are motives which are self-referential but are not self-regarding, such as a mother's desire for her children's welfare or a man's desire to injure his enemy.

(2) It follows that the only form of psychological egoism that is worth discussing is the following. It might be said that all ultimate motives are *either* self-confined *or* self-centred *or* self-referential, some being of one kind and some of another. This is a much more modest theory. I think that it covers an enormously wide field, but I am not quite certain that it is true without exception. I shall now discuss it in the light of some examples.

(A) Take the case of a man who does not expect to survive the death of his body and who makes a will whose contents will be known to no one but himself during his lifetime.

(i) The motive of such a testator cannot possibly be the expectation of any experiences which he will enjoy through the provisions of his will being carried out, for he believes that he will have no experiences after the death of his body. The only way in which this motive could be ascribed to such a man is by supposing that, although he is intellectually convinced of his future extinction, yet in practice he cannot help imagining himself as surviving and witnessing events which will happen after his death. I think that this kind of mental confusion is possible and not uncommon; but I do not think that it is a plausible account of such a man's motives to say that they all involve this confusion.

(ii) Can we say that his motive is the desire to enjoy during his life the pleasant experience of imagining the gratitude which the beneficiaries will feel towards him after his death? The answer is that this may be one of his motives, but it cannot be primary. Unless he desired to be thought about in one way rather than another after his death the experience of imagining himself as the object of certain thoughts and emotions in future would be neither attractive nor repulsive to him now.

(iii) I think it is plain, then, that the ultimate motive of such a man cannot be desire for his own happiness. But it might be desire for power over others. For he may be said to be exercising this power when he makes his will even though

the effect will not begin until after his death.

(iv) Can we say that his motive in making the will is simply to ensure that certain people will think about him and feel about him in certain ways after his death, i.e. that his motive is self-display? The answer is that this may be a motive and a very strong one, but it can hardly be his sole motive. A testator generally considers the relative needs of various possible beneficiaries, the question whether a certain beneficiary will appreciate and take care of a certain picture or house, the question whether a certain institution is doing work that he thinks important, and so on. In so far as he is influenced by these considerations his motives are other-regarding. But they may all be self-referential. In making his will he may desire to benefit persons only in so far as they are *his* relations and friends; and institutions only in so far as *he* is a member of them; and so on. I think that it would be quite plausible to hold that the motives of such a testator are all either self-regarding or self-referential; but that it would not be in the least plausible to say that they are all self-confined or that none of them are other-regarding.

(B) Let us next take the case of a man who subscribes anonymously to a certain charity. His motive cannot possibly be that of self-display. Can we say that this motive is to enjoy the pleasant experience of self-approval and of seeing an institution in which he is interested flourish? The answer is again that these motives may exist and may be strong but that they cannot be ultimate. Unless he already wants the institution to flourish, there will be nothing to attract him in the experience of seeing it flourish. And unless he subscribes from some other motive than the desire to enjoy a feeling of self-approval, he will not obtain a feeling of self-approval. So here again it seems to me that some at least of his motives must be other-regarding. But it is quite possible that his other-regarding motives may all be self-referential. An essential factor in making him desire to benefit the institution may be that it is *his* old school or that a great friend of *his* is at the head of it.

(C) The question that remains is this. Are there any cases in which it is reasonable to think that a person's motive is not egoistic in any of the senses mentioned? In practice this comes down to the question whether there are any cases in which an other-regarding motive is not stimulated by an egoistic motive-stimulus, i.e. whether there is any other-regarding motive which is not also self-referential.

Let us consider the case of a person who deliberately chooses to give up his life to working among lepers with the full knowledge that he will almost certainly contract leprosy and die in a particularly loathsome way. To give the psychological egoist as strong a case as possible we will suppose that the person is a Roman Catholic priest who believes that his action may secure for him a place in heaven in the next world and a reputation for sanctity and heroism in this, that it may be rewarded posthumously by canonisation, and

that it will rebound to the credit of the Church of which he is a priest.

It is difficult to see what self-regarding or self-referential motives for the action there could have been beside desire for happiness in heaven, desire to gain a reputation for sanctity and heroism and perhaps to be canonised after death, and desire to glorify the Church of which one is an officer. Obviously there are extremely strong self-regarding motives *against* choosing such an action. And in many cases there must have been very strong self-referential motives *against* it. For often the person who made such a decision has been a young man of good family and brilliant prospects whose parents were heart-broken at his decision and whose friends thought him an obstinate fool for making it.

Now there is no doubt at all that there was an other-regarding motive, viz. a direct desire to alleviate the sufferings of the lepers. No one would deny this unless he were dying in the last ditch for an over-simple theory of human nature. The only questions that are worth considering are these. (i) Is this other-regarding motive stimulated by an egoistic motive-stimulus and thus rendered self-referential? (ii) Suppose that this motive had not been supported by the various self-regarding and self-referential motives *for* deciding to go to work among the lepers. Would it have sufficed, in presence of the motives *against* doing so, to ensure the choice which was actually made?

(i) As regards the first question I cannot see that there was any special pre-existing relationship between a young priest in Europe and a number of unknown lepers in Asia, which might serve as an egoistic motive-stimulus. The lepers are neither his relatives, nor his friends, nor his benefactors, nor members of any community or institution to which he belongs.

Perhaps the psychological egoist might say that the intending medical missionary found the experience of imagining the sufferings of the lepers intensely unpleasant, and that his primary motive for deciding to spend his life working among them was to get rid of this unpleasant experience. About this suggestion there are two remarks to be made. (a) This motive cannot have been primary. Unless this person desired that lepers should not suffer, there is no reason why the thought of their sufferings should be an unpleasant experience to him. A malicious man finds the thought of the sufferings of an enemy a very pleasant experience. This kind of pleasure or unpleasure presupposes a desire for the well-being or the ill-being of others. (b) If his primary motive were to get rid of the unpleasant experience of imagining the sufferings of the lepers, he could hardly choose a less effective means than to go and work among them. For the imagination would then be replaced by actual perception; whilst, if he stayed at home and devoted himself to other activities, he would have a reasonably good chance of diverting his attention from the sufferings of the lepers.

In this connexion it is important to notice the following facts. For most people the best way to realise the sufferings of strangers is to imagine oneself or one's parents or children in the position in which they are placed. This, as we say, "brings home to one" the sufferings of strangers. A large proportion of the cruelty which decent people applaud or tolerate is applauded or tolerated by them only because they are either too stupid to put themselves in imagination into the position of the victims or because they deliberately refrain from doing so. And one important cause of their deliberately refraining is the abstract notion of retributive justice, i.e. the belief that these persons have *deserved* to suffer and the desire that they shall get their deserts. But this does not make the desire to relieve the suffering of strangers self-referential. Imagining oneself in their place is merely a condition for becoming vividly aware of their sufferings. Whether one will then desire to relieve them or to prolong them or will remain indifferent to them depends on motives which are not primarily self-regarding or self-referential.

(ii) As regards the sufficiency of the other-regarding motive in the absence of an egoistic stimulus and of self-regarding motives tending in the same direction, no conclusive answer can be given. I cannot prove that a single person in the whole course of history *would* have decided to work among lepers if all the motives against doing so had been present, and if the hope of heaven, the desire to gain a reputation for sanctity and heroism, and the desire to glorify and extend one's own Church had been absent. Nor can the psychological egoist prove that no single person would have so decided under these hypothetical conditions. Factors which cannot be eliminated cannot be proved to be necessary and cannot be proved to be superfluous; and there we must leave the matter.

I will now sum up about psychological egoism. (1) If it asserts that all ultimate motives are self-confined; or that they are all either self-centred or self-confined, some being of one kind and some of the other; or that all self-confined motives can be reduced to the desire for one's own happiness; it is certainly false. It is not even an approximation to the truth. (2) If it asserts that all ultimate motives are either self-regarding or self-referential, and that all other-regarding motives require a self-regarding or self-referential stimulus, it is a close approximation to the truth. It is true, I think, that in most people and at most times other-regarding motives are very weak unless stimulated by a self-regarding or self-referential stimulus. (3) But it is very doubtful if taken as a universal proposition. Some people at some times are strongly influenced by other-regarding motives which cannot plausibly be held to be stimulated by any self-regarding or self-referential stimulus. It seems plausible to hold that the presence of these other-regarding components is *necessary* to account for their choice of the actions which they do choose, though this cannot be positively proved. Whether it is also *sufficient* cannot be decided with certainty,

for self-regarding and self-referential components are always present also in one's total motive for choosing the action.

4.3472. Psychological hedonism. In refuting the narrower forms of psychological egoism I have incidentally refuted psychological hedonism, i.e. the theory that the only ultimate motive is the desire to prolong and to get pleasant experiences and to cut short and avoid unpleasant experiences.

But psychological egoism in general and psychological hedonism in particular have seemed very plausible and have continually recurred as theories of human motives. I believe that their plausibility depends, not so much on empirical facts, as on certain verbal ambiguities and misunderstandings. I shall therefore try to complete the refutation by exposing the fallacies which make these theories plausible.

I shall begin by mentioning the following obvious facts.

(1) Every choice, whatever its motive may be, is *made by* a self. Whether I decide to make myself as comfortable as possible and to ignore the claims of others, or to give all my property to charity and to spend my life working in the slums, the choice and the subsequent action will be *my* choice and *my* action. This is a mere tautology from which nothing substantial can be deduced.

(2) It may be said that a person always chooses that alternative which, on the whole and when all aspects are taken into account, attracts him most or repels him least of the various alternatives under consideration. The difference, it may be said, between a selfish and an unselfish man is not that the former always does what he likes best or dislikes least, whilst the latter often chooses an alternative which he likes less or dislikes more than some other which is open to him. The difference, it may be said, lies wholly in *what* attracts or repels the two men or in the *relative strenght* of the attraction or repulsion exercised by the same alternatives on the two men. The unselfish man is one who is strongly attracted by the thought of others being happy and strongly repelled by the thought of their being miserable and is not so strongly moved by the thought of his own happiness or unhappiness. The selfish man is one who is comparatively indifferent to the happiness or unhappiness of others, and is strongly moved by the thought of his own happiness and unhappiness. It seems to me that these statements are certain so far and only so far as they are tautological. They are certain if and only if your sole test of the relative attractiveness of various alternatives for a given person at a given time is to notice which of them he does in fact choose. Possibly this is the only test available. But in that case the proposition becomes a tautology and nothing substantial can be deduced from it.

(3) Whatever a person may aim at, and whatever his motives may be in aiming at it, the pleasure of fulfilled desire will be *his* if he succeeds, and the

unpleasure of frustrated desire will be *his* if he fails. If he decides to spend his life in the slums trying to improve the lot of the poor and succeeds in doing so, the pleasure of success will be his and not that of the poor. If he fails the unpleasure of failure will be his and not that of the poor. The same remarks apply to the pleasure of successful activity and the unpleasure of continually obstructed activity. These will equally be states of the agent, and of no one else, whether his activity be what we call "selfish" or what we call "unselfish".

Now I think that the primary fallacy which has led people to accept the narrower forms of psychological egoism is this. They start from the premiss that all choice is choice *by* a self of that alternative which on the whole attracts *him* most or repels *him* least. They jump from this to the conclusion that all motives of choice must be self-regarding. Now the first part of the premiss is a tautology, and the second part of it is certain only in so far as it is so interpreted as to be tautological. But the conclusion is a synthetic proposition. For it amounts to saying that no state of affairs can be attractive or repulsive *to* a self unless it is either (i) the continuation or cessation *of* that self; or (ii) the occurrence of experiences of certain kinds *in* that self, or (iii) the acquisition, retention, or loss of something *by* that self; or (iv) that self being the object of certain kinds of belief and emotion. Now this proposition, whether true or false, cannot possibly follow from the truism that every choice is made by a self of that alternative which on the whole attracts him most or repels him least. No conclusion as to *what* will or will not move a person's desires in one direction or another can be inferred from the mere fact that it is *his* desires which will be moved, that he will choose whatever on the whole attracts *him* most or repels *him* least, and that it is *he* who will experience the pleasure of fulfilled desire if he succeeds and the unpleasure of frustrated desire if he fails. It is as if one should jump from the truism that everything which a person sees must affect his eyes to the conclusion that the only things which a person can see are spots in his own eyes.

When this initial fallacy has been committed, and a person has to defend psychological egoism against objectors who produce apparent counter-instances, a second fallacy nearly always appears. The objector produces an instance of an apparently other-regarding desire, e.g. the desire of a mother for the well-being of her children or that of a malicious man for the discomfiture of his enemy. The psychological egoist answers by pointing out that it is a highly pleasant experience for the mother to imagine or to see her children flourishing and that it is a highly pleasant experience for the malicious man to imagine or to see his enemy suffering. He then asks us to believe that the motive of the mother in sacrificing her happiness for the children, or of the malicious man in sacrificing his happiness to secure the downfall of his enemy, was to gain these pleasant experiences. He fails to

notice that this motive cannot be primary, since the experiences in question are made pleasant only by the pre-existing desire for the children's welfare or the enemy's downfall.

I think that there is also a purely verbal ambiguity which has tended to make psychological hedonism seem plausible. Suppose I am choosing between alternative possible experiences. It sounds quite reasonable to say that the two statements "I like the experience *X*" and "I find the experience *X* pleasant" are equivalent; and that the two statements "I dislike the experience *X*" and "I find the experience *X* unpleasant" are equivalent. Now suppose we add to this the tautology that I shall always choose out of the experiences under consideration that one which I expect on the whole to like most or dislike least. Then we can draw the conclusion that in choosing between possible experiences I shall always choose that one which I expect on the whole to find most pleasant or least unpleasant. And this is psychological hedonism as applied to our motives in seeking to get or to avoid experiences.

The fallacy consists in identifying "I like the experience *X*" with "I find the experience *X* pleasant" and "I dislike the experience *X*" with "I find the experience *X* unpleasant". The statement "I find the experience *X* pleasant" is equivalent only to the restricted statement "I like *X* for its *intrinsic* properties as an experience", e.g. for its sweetness, for its ticklishness, etc. But *X* will also have extrinsic characteristics, viz. relational properties, some causal and some non-causal. And, although I like *X* for its intrinsic properties, I may dislike it very much for some of its extrinsic properties. Similarly, although I dislike an alternative experience *Y* for its instrinsic properties, I may like it very much for some of its extrinsic properties. And so, on the whole, I may prefer *Y*, which I expect to find unpleasant, to *X*, which I expect to find pleasant.

Now up to a point any reasonable psychological hedonist would admit this. But he would say that the *only* extrinsic property which can induce a person to seek or to shun an experience is its ostensible tendency to produce in himself further experiences wich will be pleasant or unpleasant. In our general discussion of psychological egoism we have seen that this is certainly false, and that the causes which have made it seem plausible are not good reasons.

4.3473. Summary on pluralism v. monism of ultimate desires. I have tried to show that psychological egoism, in the only form in which it could possibly fit the facts of human life, is not a monistic theory of motives. On this form of the theory the only feature common to all motives is that every motive which can *act on* a person has one or another of a large number of different kinds of special *reference to* that person. I have tried to show that it is by no means certain that there is even this amount of unity among human motives. I think that psychological egoism is much the most plausible attempt to reduce the

plurality of ultimate desires to a unity, and that, if it fails, it is most unlikely that any alternative attempt on a different basis will succeed. So I accept a radically pluralistic view of human motives.

This does not of course entail that the present irreducible plurality of ultimate motives may not have evolved, in some sense, out of fear in the history of each individual or in that of the human race. About this I express no opinion here and now.

Now, if psychological hedonism had been true, all conflict of motives would have been between motives of the same *kind*. It would always be of the form: "Shall I go to the dentist and certainly be hurt now but probably avoid frequent and prolonged tootache thereby in future? or shall I take the risk in order to avoid the certainty of being hurt by the dentist now?" According to me there is also conflict between motives of different *kinds*. e.g. between aversion to painful experiences and desire to be thought manly, or between desire to be thought witty and aversion to hurting a sensitive person's feelings by a witty but wounding remark. In our moral judgments about ourselves and about others we always assume that there can be and often is conflict between motives of different kinds. If psychological hedonism or any other purely monistic theory of motives had been true, we should have to begin our study of ethics by recognising that most moral judgments are made under a profound misapprehension of the psychological facts and are largely vitiated thereby. As it is, there is no reason to believe this.

4.348. Conflict and cooperation of desires

I shall begin by analysing and defining the notion of "desire" rather more carefully. We will begin with a concrete example and then generalise from it.

Supose that *A* desires that *B* shall be appointed to a certain office. Then in the first place we must distinguish what I will call the *content* and the *intent* of *A*'s desire. *A* contemplates a certain possible future state of affairs, viz. that *B* should be appointed to this office. This possible future state of affairs is the *content* of his desire. He desires that this possible future state of affairs shall be realised. I call this the *intent* of his desire. He might have had a desire with the same content but with the opposite intent. This would have been the case if he had desired that *B* should *not* be appointed to that office. *A* would have been contemplating the same possible future state of affairs, but desiring that it should *not* be realised.

Now the content of *A*'s desire contains several constituents, viz. the person *B*, the office in question, and the relation of being appointed. I shall call these the *referents* of *A*'s desire. So the referents of a desire are the several terms which are the constituents of the content of that desire.

The *object* or *desideratum* of *A*'s desire is that the possibility that *B* will be appointed to the office shall be realised. So the desideratum of a desire is the

unity of its content and its intent, where the content functions as subject and the intent as predicate.

Now the content of a desire can take two different forms. It may be either a *present actual* state of affairs or a possible *future* state of affairs. Suppose that *A* desires that *B*, who already holds a certain office, shall continue to hold it. Then the content of his desire is a present actual state of affairs. The intent of his desire is that this shall continue.

The intent of a desire can take two opposed forms if the content is an actual state of affairs, and two other opposed forms if the content is a possible future state of affairs. In the first case the intent may be either the *continuance* or the *cessation* of the actual state of affairs which is the content. In the second case the intent may be either the *realisation* or the *non-realisation* of the possible future state of affairs which is the content.

We can now deal with the opposition of *desire* and *aversion*. This can be done conveniently in the following way. I shall say that a person may have either a *pro-desire* towards a certain content *C* or an *anti-desire* towards it. These terms may be defined as follows. A pro-desire towards *C* is either (i) a desire for the continuance of *C* if *C* is an actual state of affairs or (ii) a desire for the realisation of *C* if *C* is a possible future state of affairs. An anti-desire towards *C* is either (i) a desire for the cessation of *C* if *C* is an actual state of affairs, or (ii) a desire for the non-realisation of *C* if *C* is a possible future state of affairs. So an anti-desire towards *C* is the same as an *aversion to C*, and we need not complicate our statements in future by introducing aversion as well as desire.

I think that the analysis which I have just given covers every case of desire. But I must point out that there are certain phrases in common use which seem at first sight to conflict with it. It would be quite usual to say that a certain person on a certain occasion desired a glass of port or a sum of money. Now port and money are *things*, not actual or possible states of affairs. But really this is no exception, for these expressions are elliptical. What is meant is that this person desired to *drink* a glass of port or to *possess* a sum of money. Now the drinking of port by a person or the possession of a sum of money by him are actual or possible states of affairs. In fact the content of this person's desire was the possibility of his drinking a glass of port in the future. And the intent was the realisation of that possibility. The port was just one of the referents of the desire. The other referents were the person himself and the act or relation of drinking; but these are tacitly assumed and not explicitly mentioned.

Next we can define the terms "fulfilment" and "frustration" of desire. A pro-desire towards a content *C* is *fulfilled* if *C* continues to exist or becomes realised, as the case may be. It is *frustrated* if *C* ceases to exist or fails to become realised, as the case may be. An anti-desire towards a content *C* is *ful-*

filled if *C* ceases to exist or fails to become realised, as the case may be. It is *frustrated* if *C* continues to exist or becomes realised, as the case may be. To put it generally, a desire is fulfilled if the relevant present or future events accord with its intent; it is frustrated if the relevant present or future events conflict with its intent.

Lastly we can define what is meant by "indulging" or "foregoing" a desire. We must notice, in the first place, that the word "desire" is used more narrowly than the word "wish" or "hope". One can wish for what one knows to be impossible, e.g. that something which has already happened should be altered. One can hope for a future state of affairs which one can do nothing towards bringing about. But we use "desire" in such a way that a person would be said to desire only such objects as he knows or believes to be possible and to be in part at least dependent on his own actions. A person *indulges* a desire if he acts with the intention of fulfilling it. He *foregoes* a desire if he deliberately omits to act with the intention of fulfilling it or acts with the intention of frustrating it. E.g. suppose that a person has an anti-desire towards working on a certain evening. He fulfils it if he stops working after a short time or does not begin to work. He foregoes it if he begins and continues to work in spite of his anti-desire towards working.

Two desires in a person *conflict* if to indulge one of them would involve, either directly or indirectly, foregoing the other. This may happen in various ways, and I shall now consider some of them. (1) A person may have two desires with the same content and opposite intents. E.g. a soldier in battle may desire to run away through fear of death or injury, and he may at the same time desire not to run away through sense of duty or fear of being court-mashalled and shot or dislike of incurring the contempt of his comrades. (2) A person may have two desires with different contents, and they may be both pro-desires or both anti-desires. But the contents may be so related that the desires cannot both be fulfilled. E.g. a person may have two invitations to dinner at different houses on the same night and he may desire to accept each of them. Since he cannot be in two places at once the contents of the two desires are so related that he cannot indulge either without *ipso facto* foregoing the other. (3) It may be that the only means which a person could take to fulfil one desire would have consequences which would frustrate the other. E.g. a traveller who has to spend a night in a forest may desire to avoid attacks by wild animals and to escape the notice of savages. But anything that he can do, such as lighting a fire or beating a gong, in order to frighten off wild animals will automatically betray his presence to the savages. Under this head come all cases of conflict that arise through limitation of means or time or energy. In order to fulfil one desire you may have to expend so much money or time or energy that you will not have enough left to fulfil the other.

Let us now go back to the first case, viz. two desires with the same content

and opposite intents. We can talk of desires springing from certain *conative tendencies*. E.g. when a person believes himself to be in danger he desires to escape. This may be said to spring from the conative tendency of *fear*. When he is hurt or thwarted by another he has a desire to return the injury. This may be said to spring from the conative tendency of *resentment*. And so on. Now desires with the same content and opposite intent may spring from the same or from different conative tendencies. The desire of the soldier to run away is based on the fear-tendency. Suppose that his simultaneous desire not to run away is based on his fear that if he does so he will be court-marshalled and shot. Then we have two desires with the same content and opposite intent based on the same conative tendency. But suppose that his simultaneous desire not to run away is based on his sense of duty or his dislike of appearing cowardly to his comrades. Then we have two desires with the same content and opposite intent based on different conative tendencies.

In both cases the different desires are connected with different aspects of the total situation. The desire to run away is felt in respect of the immediate present danger of remaining. The desire not to run away, even if it is also based on the fear-tendency, is felt in respect of the subsequent dangers which will arise as a consequence of running away. Or suppose that the desire not to run away is based on sense of duty or dislike of appearing cowardly. In the former case the content is considered by the soldier in reference to a whole set of relations in which he stands to his country, his officers, his comrades, etc. In the latter case it is considered by him in the light of the effect which the knowledge of it by others would have on their attitude towards himself.

Just as two desires may conflict so they may cooperate. This happens if they can both be indulged and the indulgence of one either directly or indirectly involves or facilitates the indulgence of the other. E.g. the soldier may have a desire not to run away based on fear of the consequences, a desire with the same content and intent based on dislike of appearing cowardly, and a desire with the same content and intent based on his sense of duty. If so, these three desires cooperate; and he is much less likely to run away than if there is only one of them present to oppose his desire to do so based on fear of present danger.

4.3481. Conative tendencies of different orders. There is a certain sense in which we can say that sense of duty is a conative tendency of a "higher order" than dislike of appearing cowardly, and that the dislike of appearing cowardly is a conative tendency of a "higher order" than fear of death or wounds or punishment. I will now try to explain this notion more fully. What I shall say is closely connected with Plato's account of the soul in the *Republic* and with Butler's account of human nature as a system in his *Sermons*.

(1) Primary propensities. We have a limited number of conative tendencies which I will call *primary propensities*. Each of them is aroused only in situations of a characteristic kind which recur fairly often in everyone's life. And, if the desire is indulged, it leads to a characteristic kind of action. Examples of primary propensities are the propensity to eat when hungry, to dring when thirsty, to retaliate when hurt or thwarted, the sexual impulse, and so on.

There are certain things which can be said of all of them. (i) They exist in animals as well as or in human beings though they may be modified and complicated in all kinds of directions in men. (ii) When they are aroused the person nearly always feels a strong emotion of a characteristic kind. And often there are also characteristic bodily sensations. Examples are the emotion of anger which is felt when the propensity to retaliation is aroused; the characteristic sensations of hunger and thirst, and so on.

There are other things which are true of some primary propensities but not of all. Some of them are connected with recurrent states of bodily depletion or repletion, and the desire and action to which they give rise are directed towards restoring the depletion or evacuating the repletion. We can take hunger as a typical example. Suppose that a person is hungry and indulges his desire to eat. Before he begins, and during the earlier stages, he has certain characteristic sensations in his stomach. These are not unpleasant when slight, but they increase and become intensely painful if the desire to eat cannot be indulged. As he eats these sensations gradually disappear. But the process of eating is necessarily accompanied by sensations of taste and smell and contact, which may be pleasant or unpleasant. Finally, if he eats as much as or more than he needs, he will begin to get a different set of bodily sensations, arising from the presence of the food in his stomach and the process of digesting it. These may be mildly pleasant, but they may be slightly unpleasant or acutely painful. We must therefore distinguish (i) the *premonitory*, (ii) the *concominant*, and (iii) the *consequent* sensations connected with any primary propensity, such as hunger, which depends upon bodily depletion or repletion. Again, the sensations which are concomitant to the process of satisfying the hunger fall into two classes, viz. the bodily sensations arising from the actual processes of biting, chewing, swallowing etc., and the sensations of taste, smell, etc. which depend on the kind of food which is eaten. I shall call the former *intrinsic* and the latter *extrinsic*. In the case of eating the intrinsic sensations are comparatively unimportant. They are seldom positively pleasant, and are unpleasant only if one has some bodily defect such as a tender tooth or a sore throat. The extrinsic part of the concomitant sensations is what matters here. But in the indulgence of certain other primary propensities, such as the sexual impulse, the opposite is true. There the intrinsic part of the concomitant sensations is highly pleasant, and the extrinsic part is relatively unimportant.

Let us consider a primary propensity, such as retaliation, which has nothing to do with restoring a depletion or evacuating a repletion. The propensity is aroused in an individual by his being hurt or threatened or thwarted by another individual or even by an inanimate thing. An emotion of anger is then felt towards that individual or thing, and a desire is aroused to hurt, injure, or destroy it. The sensations concomitant to the process of indulging the desire may be of the most various kinds; they will often be extremely unpleasant, since one may have to struggle desperately and receive further injuries in doing so. There are no characteristic sensations, such as arise when one has satisfied one's hunger, consequent upon the process of retaliating an injury.

Let us now consider the possible further developments of primary propensities. I will take the propensity to eat when hungry as an example. Take first the case of a hungry animal. It would be absurd to say that he desires to eat as a means to getting certain pleasant sensations. That would assume that he can remember the past and imagine the future, and that he has the idea of certain causes leading to certain effects. Can we say that he desires to eat as a means to getting rid of certain unpleasant sensations? This is again far too intellectual. It is true that the unpleasant sensations of hunger are a necessary condition of arousing the desire to eat. If he were anaesthetised, he would not desire to eat, although he might need food just as much. What we must say is this. He is so constituted that, when he has a certain kind of sensation, which would be extremely unpleasant if it were prolonged and intensified, he desires to act in a certain way towards certain kinds of external objects, viz. to seize and to chew and swallow them. This kind of act, if it is successful, will *in fact* lead to the renewal of the unpleasant premonitory sensations of hunger, and it may *in fact* give pleasant concomitant sensations of taste. But it is not desired by the animal as a means to those ends. It is desired directly on the occasion of and at the instigation of the characteristic premonitory sensations.

The only further development that can take place in an animal is this. Certain ways of satisfying hunger do in fact have pleasant concomitant sensations and are not followed by painful consequences. Others involve unpleasant concomitant sensations, or are followed by painful consequences such as a beating. If this happens often and regularly the animal may become conditioned. When presented with two alternative ways of satisfying its hunger it will now tend to adopt the sort that has had pleasant concomitant sensations and no painful consequences. But the animal will never reach the stage of *deliberating* and *choosing* that alternative *for that reason*. Now consider the case of a man.

(a) A starving shipwrecked sailor comes nearest to an animal. But even here there are important differences. (i) He is moved, not only by his present

painful sensations, but by the knowledge that they will get worse and worse and the imagination of himself dying in agony if he does not soon eat some food. (ii) He may have a very limited amount of food available, and he may overcome the temptation to eat it all at once, because he knows that he will soon become hungry again and then will have nothing to eat. (iii) He may feel repulsion, based on moral, religious, or aesthetic grounds to the only food available. E.g. if he is a Jew he may think it wrong to eat pork; and whatever his religion may be he may have a strong repulsion to eating a human body.
(b) A mildly hungry healthy man in ordinary circumstances eats primarily to satisfy his hunger; but, if he has a choice of food, he considers which will give him the pleasantest concomitant sensations, and which, if any, is likely to disagree with him and give him unpleasant consequent sensations. He may have to weigh opposite considerations against each other. Then, again, if he is a person of limited means, he will have to consider that, if he spends so much on a meal, he will have so much less to spend on something else which he would like more, or would be unable to pay a debt which he owes, and so on.
(c) An epicure will be primarily concerned with getting the maximum of pleasant concomitant sensations and the minimum of unpleasant consequent sensations. He may even take steps to *make* himself hungry, e.g. by taking vigorous exercise, in order to increase the pleasure that he will get from eating.
(d) Lastly an invalid recovering from an illness may, as we say, have to *force* himself to eat. He does not feel hungry and he does not get pleasant concomitant sensations. Perhaps he gets mainly unpleasant sensations of nausea. He forces himself to eat because he knows that this is a necessary condition of keeping alive and getting better; and he wants to live and regain his health.

(2) Conceptual extension of primary propensities. This is the next state in the hierarchy, and it can occur only in creatures like men who are capable of conceptual cognition. A man is not only capable of being hungry or resentful or afraid, he is capable of reflecting on his propensity and considering how he can best satisfy it on the whole. An example is the hungry sailor who deliberately refrains from devouring all his available food at once in order to have some left to satisfy future hunger. Another example is the timid soldier who checks his desire to run away from fear of the enemy by his greater fear of being court-martialled and shot if he does so. A third example is that of the man who has suffered an injury or insult and checks the tendency to react with an immediate blow because he believes that he can inflict a greater injury if he waits and plans the ruin of his enemy.

The essential point to notice is this. The indulgence of a certain primitive propensity may be checked by a conceptual extension of that very same

propensity, which looks beyond the present occasion to the effects of a present indulgence on the possibility of future indulgences. Again the indulgence of one primitive propensity may be checked by the conceptual extension of another. E.g. the desire of a hungry man to take and eat some food which does not belong to him may be checked by the thought of being detected and punished in future, i.e. by a conceptual extension of the primary fear-propensity.

(3) Organising desires. These come next in the hierarchy. Examples are the desire to acquire and to keep property, the desire to aggrandise one's family, and various forms of the patriotic desire. Each of these is a persistent and far-reaching desire under which several simpler and more primitive propensities are organised. There is no one characteristic kind of situation, such as being hungry, which calls forth one of these organising desires. And there is no one characteristic kind of action, such as eating, to which any one such desire leads when it is indulged. The object of an organising desire can be attained only by a whole sequence of actions of the most varied kinds. Some of these actions will involve the frustrating of a desire based on a primary propensity; others will involve the indulging of such a desire.

In accordance with what we said when we discussed psychological egoism organising desires may be classified as follows. We divide them first into *self-regarding* and *other-regarding*. Then we sub-divide those which are self-regarding into *self-confined* and *self-centred but not self-confined*. And we sub-divide those which are other-regarding into *self-referential* and *not self-referential*.

Before I discuss desires which come under these headings, I must draw some distinctions concerning desires about persons. Let us begin by considering a desire which a person *A* has about another person *B*. (1) In the first place *A*'s desire may be favourable to *B* or unfavourable to him. So it may be either a *pro-desire* or an *anti-desire* towards *B*. (2) Whether it be pro or anti it may take two forms. (i) It may be a desire that *B* shall remain or become or cease to be a *person of such and such a kind*, e.g. a hero or a drug-fiend. This means a desire that *B* shall have such and such mental and bodily dispositions, organised or disorganised in a certain way. We may call such desires desires about *B*'s *personality*. (ii) *A*'s desire might be that *B* should lead a *life of such and such a kind*, i.e. that he should have such and such experiences and do such and such actions in such and such an order. Now this kind of desire can take two different forms. (a) *A* may desire primarily that *B*'s life shall be predominantly *happy* or *unhappy*. He may be indifferent to whether it is good or bad in other respects except in so far as this has a bearing on its being happy or unhappy. (b) *A* may desire primarily that *B*'s life shall be *good* or *bad* in the widest sense, i.e. he may desire *B*'s *welfare* or *illfare*. He may be

concerned with *B*'s happiness or unhappiness only in so far as this has a bearing on the goodness or badness of *B*'s life in the widest sense. Suppose, e.g., that *B* were by nature a cruel or a lustful person. Then, if *A*'s desires were primarily for *B*'s *happiness*, he might wish to give *B* every opportunity to indulge his cruelty or his lust. But suppose *A*'s desires were primarily for *B*'s *welfare*, i.e. that *A* desired that *B*'s life should be as good a one, in the widest sense, as such a man as *B* could lead. Then *A* might wish to remove from *B* opportunities to indulge his cruelty or his lust, even though this would make *B*'s life less happy.

We can sum this up as follows. If *A* has a desire about another person *B*, it may be either about *B*'s personality or about *B*'s life. If it is about *B*'s life it may be either confined to *B*'s happiness or unhappiness; or it may be concerned with *B*'s welfare or illfare, in which his happiness or unhappiness is only one factor among others. And in all three cases *A*'s desire about *B* may be either pro or anti.

Now *A* may have desires about his *own* personality and his *own* life, as well as about another person's personality and life. I doubt if it is psychologically possible for a person to have anti-desires towards himself. But he can certainly have pro-desires about his own personality, and about his own happiness or unhappiness, and about his own welfare or illfare.

We can now enumerate the main kinds or organising desire as follows:

(1.1) Self-confined. (i) Desire for *self-preservation*, regardless of whether one's life will be happy or good or useful or not. (ii) Desire to have as much pleasant experience and as little unpleasant experience as possible throughout one's life as a whole, regardless of whether one's life will be good or bad in other respects. This may be called *desire for one's own greatest enjoyment.* (iii) Desire to have as good a life, in the widest sense, as is possible for one to have. This will include as much and only as much pleasure and as little unpleasure as is compatible with or conducive to a life which is good in the widest sense. This may be called *desire for one's own greatest welfare.* (iv) *Desire to multiply and intensify one's bodily and mental powers as much as possible and to organise one's personality, as an end in itself and not merely as a means to leading a pleasant or a good life or being useful to others. This may be called desire for self-development.* (v) Desire to respect and approve oneself and one's actions, experiences, and dispositions.

(1.2) Self-centred but not self-confined. (i) Desire to get and keep possession of things or persons as an end in itself and not merely as a means to pleasure, power, security, etc. This may be called *desire for possession.* (ii) Desire to get and to exercise power over others as an end in itself and not merely as a means to pleasure, wealth, security, etc. This may be called *desire for power.* (iii)

Desire to be the object of certain kinds of emotion in other persons, e.g. to be an object of affection or respect or fear. This may be called *desire for emotional reaction*.

(2.1) Other-regarding but self-referential. (1) Under this head are included all desires which are directed towards the personality or the life of another person, provided that the desire depends on some special relationship between that other person and one's self. Such a desire may be either pro or anti; it may be directed towards the other person's personality or his life; and if it be directed towards his life it may be concerned only with his happiness or unhappiness or with his welfare or illfare. Under the first head comes the desire to improve and develop or to corrupt and stunt another man's personality. Under the second head comes (i) the desire to make another person's life as happy as possible without regard to whether it will be in other respects good. (ii) The desire to make another person's life as miserable as possible without regard to whether it will be in other respects bad. Under the third head comes (i) the desire to make another person's life as good as possible in all respects, and to give him as much happiness and as little unhappiness as is compatible with or conducive to this. (ii) The desire to make another person's life as bad as possible in all respects, and to give him as little happiness and as much unhappiness as is compatible with or conducive to this.

We are confining our attention to cases where these desires are felt by a person only towards those persons who stand in certain special relations to him. So I will now mention some of the most important relationships on which such desires are based. The pro-desires are based on such relationships as parent-to-child; child-to-parent; common parentage; love or friendship independent of blood-relationship; membership of the same group of inter-related persons other than blood-relations, e.g. being a fellow-countryman, a schoolfellow, and so on. The anti-desires are based on such relationships as envy, jealousy, having suffered an injury at the hands of the other person, and so on.

(2) Under this head we must also include pro or anti desires directed, not to other individuals, but to groups of inter-related individuals. A man may desire to aggrandise his family or his nation quite regardless of the welfare or illfare of the individuals which compose it. In the same way he may desire to depress or ruin another family or a foreign nation, without having any hostile desires towards the individuals who compose it. He may realise that they will be involved in the ruin of their family or their nation; but this is not what he primarily desires, and he may even regret it. The relationship at the basis of such anti-desires is generally the real or imagined hostility of the foreign group to a group of which one is a member.

It should be noticed that all the other-regarding but self-referential desires are closely bound up with certain *sentiments* which one has formed about certain individuals or groups, e.g. the parental sentiment, the sentiment of loyalty or patriotism, the sentiment of love for a person who is not a blood-relation, and so on. In fact these organising desires are the conative factor in such sentiments.

(2.2) Other-regarding and not self-referential. The pro-desires under this head are the following. (i) The desire that the life of every sentient being as such should be as *happy* as possible, regardless of any special relationship in which he may stand to oneself. This may be called *desire for the general happiness.* (ii) The desire that the life of every sentient being should be as *good* in all respects as possible, and that it should contain as much happiness and as little unhappiness as is compatible with or conducive to its maximum goodness. This may be called *desire for the general welfare.*

I do not think that the corresponding anti-desires, viz. for the general unhappiness and for the general illfare exist in sane human beings. I think that anit-desires about persons always depend upon some special relationship of the other person to the person who feels the anti-desire, i.e. that they are always self-referential.

I think that desire for general happiness or for general welfare are extremely weak in most people, and are very easily and frequently overcome by self-referential anti-desires arising from special relationships such as jealousy, envy, etc. They are also very easily and frequently overcome by self-referential pro-desires arising from special relationships, such as the parental relation. These restrict one's pro-desires to the members of certain limited groups, and thus inhibit the desire for *general* happiness or *general* welfare. It has been said that there is no crime which a good father of a family is not capable of committing.

The following remarks may be made about this list of organising desires. (i) They may be compared with each other, in respect of the extent of their object, in two different ways. In one sense, the self-confined desires and the pro and anti-desires connected with love or hatred of a particular individual have very narrow objects. For they are concerned with only a single person. In another sense their objects may be very extended. For such desires may be concerned with the whole future history of that person. (ii) They may be compared, in respect of their duration, in two different ways. We may consider the date at which the conative disposition was first formed; and we may consider the frequency with which it is excited after it has been formed. Now the pro-desires which are felt by a person towards himself score in both respects. Unlike his loves or hates of other persons, they are present as a disposition from the cradle to the grave. And they are nearly always in action.

For their characteristic object, viz. one's self and one's own experiences, is perpetually and intimately present to one in a way in which no other object can be. (iii) There is yet another respect in which the desire for one's own maximum happiness stands in a peculiar position. As we have seen, the indulgence of *any* desire is as such pleasant, and the foregoing of *any* desire is as such unpleasant. Therefore, in so far as a person was activated by the desire for his own maximum happiness, the thwarting of *any* of his desires would be at best a regrettable necessity. Such a person would be willing to check a desire only when its indulgence would be incompatible with the indulgence of some stronger desire or would lead to some unpleasant consequences. Now that is not true of any other organising desire. Take the case, e.g., of a person who is dominated by the desire to amass property as an end in itself. He will have to check many of his other desires. But in so far as his ruling desire is simply to amass property, this suppression will not be in itself a regrettable necessity. For what he is after is possessions, not happiness; and the thwarting of a desire does not diminish his property though it does diminish his happiness.

4.3482. Temperamental hindrances and temperamental energizers. There are certain dispositions in every human being which hinder the carrying out of *any* far-reaching desire. There are others which are almost necessary conditions for carrying out any such desire. I shall call them respectively *temperamental hindrances* and *temperamental energizers*. The most important temperamental hindrances are *laziness*, bodily and mental, and *timidity*, physical and moral. Everyone finds it easier to be physically and mentally passive than to initiate active bodily work or hard thinking and to carry them on against growing discomfort, fatigue and boredom. But no far-reaching desire can be realised without persistent work which will often go against the grain. Again, everyone shrinks from bodily pain, from the risk of being injured or killed, and from incurring blame, unpopularity, ridicule or hostility. Yet hardly any plan can be carried out without running some of these risks. It should be noticed that this is true even if the organising desire is simply for one's own greatest happiness. Through laziness or timidity a person will omit to have experiences and to acquire bodily and mental powers which he knows quite well would give him more pleasure in the long run than he can hope to get by a passive unenterprising safe mode of life.

The temperamental energizers are certain dispositions which tend to counteract laziness and timidity. As Plato pointed out, a very important temperamental energizer is connected with the primary impulse to retaliate and feel anger. This may be called *combativeness*. In its most primitive form it shows itself as a tendency to resent actively attempts by others to thwart one's desires, and in particular any attempt to take away something which

one has. Some animals are habitually combative; but even timid creatures, like hens and deer, may become so in special circumstances, e.g. when their young are threatened or in the mating season. When combativeness is aroused an animal or a man will make exertions and endure pain and danger which he would otherwise shirk.

In men, with their powers of conceptual and reflex cognition, this primitive combativeness develops in very elaborate ways which are quite unknown in animals. One development is *emulation*, i.e. the desire to surpass others and the dislike of being surpassed by them. Another development is that courage and endurance are admired and praised, whilst cowardice, irresolution, and laziness are blamed in nearly all communities. Now each person wants to be admired and respected by his fellows, and so he may force himself to behave resolutely from fear of incurring contempt or blame. A further development is connected with reflex cognition and emotion. Each person tends to judge himself by the standards of the society in which he lives. If he knows that he is being lazy or cowardly and thus failing to carry out his own long-range desires, he will feel angry with himself on that account. This reflex anger will be a stimulant against laziness and cowardice. Again, even when no other person is concerned, a man may challenge *himself* to undertake and carry through a difficult feat. He may come to treat the obstacles which inanimate nature or his own weaknesses put in his way almost as if they were the deliberate opposition of a human rival or enemy. In these ways the obstacles to carrying out a desire may very greatly strengthen it. Often one does not very strongly desire to do a certain thing; but, if that weak desire is opposed, one may make a point of doing it simply to assert one's independence and get one's own way.

We must distinguish an *initiating* and a *persevering* aspect of the general tendency which I have called "combativeness". Some people will seek difficulty and labour and danger with enthusiasm on the slightest provocation, but will not show much persistence in carrying out their desires in face of them. This is the initiating aspect. Others will rather shun them and prefer a quiet lazy life. But, if they are provoked or challenged beyond a certain point or involved against their will, they will show great persistence. This is the perseverving aspect. Of course the two may be combined. Then again, some people who are courageous are lazy, and others who show great resolution in face of fatigue and boredom and obstruction are timid. Lastly, a person who will readily face bodily pain and danger may be very timid and irresolute in face of hostile public opinion.

We can sum this up as follows. The primary impulse to retaliate when thwarted may be very detrimental to carrying out any far-reaching desire, if it remains in its crude form and leads to uncontrolled outbursts of anger and violence. But, if it is sublimated in certain ways which I have described, it

becomes a very powerful help to the carrying out of such desires. The temperamental hindrances are strong in everyone at first, and they remain strong in most people throughout life; and any attempt to carry out any far-reaching desire is certain to come up against them. Therefore no one who does not possess a fair initial dose of combativeness is likely to be able to organise his life successfully on *any* plan. An important part of a person's moral education is to have such combativeness as he possesses trained and sublimated so that it becomes a source of useful energy and not a destructive explosive.

4.3483. Conflict and cooperation of organising desires. Suppose that a person had only one organising desire, e.g. desire for his own maximum happiness. Then there could be only two kinds of inner conflict in his life, and they would all be capable in principle of a single kind of solution. (1) There might be occasions on which two of his primary propensities would conflict, or when the conceptual extension of one of them conflicted with that one or with another one. E.g. hunger might spur him to seize certain food, and fear might deter him from doing so. (2) Any of his primary propensities, or the conceptual extension of it, might conflict with his one organising desire, i.e. in the case supposed with his desire for his own greatest enjoyment. E.g. primitive resentment might move him to strike out at someone who had insulted him, but a calm consideration of his chances of future happiness might move him to dissemble his annoyance.

In principle the solution of such conflicts would always be the same. If the person could check his tendency to immediate action and give himself time to think, the one question which he would ask himself is: "Would this action be conducive to my happiness on the whole or would it not?" And, if his one organising principle were strong enough, he would indulge the propensity if and only if his answer were "Yes".

But in point of fact everyone has many more than one organising desire. And so there is a possibility of conflict and cooperation at a higher level, viz. between two or more organising desires. E.g. a man may desire to maximise his own happiness and he may also desire power for its own sake. Suppose now that he has the offer of some post, such as that of Prime Minister, which would give him very great power, but involve much drudgery and make him the object of much ill-feeling. His desire for power will move him to accept the offer. His desire to maximise his own happiness will in part move him in the same direction and partly in the opposite. It will move him in the same direction because he knows that, if he declines the office, his desire for power will be thwarted; and the thwarting of any strong desire is itself unpleasant and likely to be a constant source of discontent. It will move him in the opposite direction because he knows that, if he accepts the office, he will be

exposed to all kinds of unpleasant experiences and will have to forego many pleasures which require leisure and absence of responsibility.

So the next question that arises is this. Is there any supreme organising principle in human life, which stands to the various organising desires in somewhat the same relation in which each of them stands to the various primary propensities? Butler said that there is, and that the supreme principle is *conscience*. But he also drew a distinction between the *moral authority* of conscience and its actual *psychological power*. He said that if conscience had the psychological power which corresponds to its moral authority it would be the supreme organising principle in every man's life. I shall now discuss these questions in my own way.

4.3484. Types of unification A person's character would be completely unified if he had a single all embracing plan or ideal for his life by reference to which he decided whether and how far each other organising desire and each primary propensity should be indulged or checked. We need not suppose that this plan or ideal would always be explicitly before his mind, or that every trivial action or decision would be deliberately considered in relation to it. In many cases it would be a matter of indifference which alternative he chose. In many other cases he would automatically act in accordance with the scheme and avoid actions which would conflict with it. But, whenever a serious conflict did arise and persist, he would refer to this plan or ideal. He would ask himself whether the indulgence on a certain occasion of a certain primary propensity, e.g. anger, or the pursuit of a certain organising desire, e.g. desire for maximum enjoyment, would fit in with or conflict with his over-ruling plan of life.

We need not suppose that the supreme organising desire would always succeed in checking other desires when they conflict with it. It might sometimes be overcome by a primary propensity, e.g. resentment, or by some other organising desire, e.g. desire for power or property for its own sake. On such occasions the person will regret the victory of the primary propensity or the other organising desire. He will feel it as a kind of personal defeat and as a frustration of a desire with which he is more fully identified. And he will feel angry with himself for failing to act in accordance with his supreme organising desire.

Suppose that such a person indulges a desire which he sees to be helpful to the realization of his supreme desire. Then he will have two kinds of satisfaction, viz. (i) that which arises from indulging that particular desire, and (ii) that which arises from knowing that in so doing he is forwarding his over-ruling plan of life. Suppose that on another occasion he checks a desire because he believes that to indulge it would be detrimental to his over-ruling plan of life. Then he will have a certain pleasure to set against the unpleasure

of that particular desire being thwarted. For he would have the satisfaction of knowing that, in thwarting this desire, he was contributing to the fulfilment of his supreme organising desire. And he would have a pleasant feeling of self-approval at having had the strength to check a desire which he had decided that he ought not to indulge on that occasion.

Obviously many people never attain to any high degree of unification. Some never even manage to gain an habitual control of their primary propensities, because all their organising desires are very weak or some of these propensities are abnormally strong. Such a person will eat greedily when hungry, be violently angry when thwarted, and so on, without consideration of the remoter consequences. If the circumstances should evoke two primary propensities leading to opposite kinds of action, e.g. resentment and fear, the conflict will not be solved on any principle. There will just be a kind of mechanical clash of forces, leading to vacillating action or the victory of the stronger impulse. I shall say that such a person is *disorganised at the primary level.*

Another kind of disorganisation is possible. A man's organising desires may be strong enough, as compared with his primary propensities, to control the latter on most occasions. But he may have several strong organising desires and none which is supreme over all the rest. I shall say that he is *disorganised at the secondary level.* E.g. his desire to maximise his own happiness, to have power over others, and to aggrandise his family may all be quite strong. Each may be quite strong enough to organise his primary propensities and to resolve conflicts between them on its own principle. And in many cases there will be cooperation and not conflict between these organising desires. Often the very same act which gives him power over others will help to aggrandise his family and to secure opportunities for pleasant experiences. But there will certainly be many cases when they would conflict, i.e. where each would move him to a different course of action. In such cases he has no principle by which to decide how far one is to be indulged and another is to be checked.

If a personality is to be completely unified there must be some one organising desire which is habitually predominant over all the rest. But this kind of predominance may take two different forms which I will call the *despotic* and the *constitutional.* A personality is unified *despotically* if one organising desire is from the first, or gradually becomes, very much stronger than all the rest without the person having considered the relative value of his various organising desires and of the various kinds of life to which they will lead. In this case the other organising desires tend to atrophy, and thereafter never have a chance to assert themselves in case of conflict with the dominant desire. A typical example is the miser. In such a man the desire to acquire and to keep property as an end in itself has become an obsession. The man does

not deliberately compare it and the kind of life to which it leads with other organising desires and the kinds of life to which they lead. He does not set it over the rest because he judges it or the life to which it leads to be the best. It just usurps control; and other organising desires, such as the desire for happiness, for affection, and so on, never get a fair hearing and at length atrophy almost completely.

A personality which is unified on despotic lines is necessarily cramped and lopsided to some extent. The extent to which it is so will depend on the nature of the desire which becomes dominant. If this has a very restricted kind of object, e.g. acquirement and retention of property for its own sake, the personality will be proportionately cramped. But the despotic desire might, e.g., be the desire to develop one's powers and capacities to the utmost as an end in itself. In that case the personality would not be specially cramped. But there would still be a certain lopsidedness about the life. For enjoyment and personal affection, e.g., would be unhesitatingly sacrificed when they conflicted with the acquirement of new capacities or the development of old ones. And this would not be done in consequence of a deliberate judgment that enjoyment and personal affection are less valuable than the possession of great and numerous powers of mind and body.

We come now to the notion of a *constitutionally* unified personality. It seems to me that there is a certain ideal of human life which would be accepted by most sane grown-up human beings if you put it to them in their calmer and more reflective moments. It is very vague in its positive details and it is unlikely that there would be general agreement about these. But it is fairly definite in certain negative respects. (1) A life in which the various organising desires simply take turns with each other to dominate and be dominated, on no plan or principle, is felt to be unsatisfactory. It is condemned as incoherent. (2) A life, like that of the miser, in which one organising desire just happens to become predominant and then simply represses the rest, is also felt to be unsatisfactory. If any desire is to be repressed, one wants to see some good reason for its being suppressed. And, if one desire is habitually to dominate over all the others, one wants to see some good reason why that one rather than another should occupy this predominant position. (3) We have the ideal of a kind of life in which the various primary propensities and the various organising desires would all play their parts like the various instruments in a properly conducted orchestra. None would be completely suppressed and none would be completely dominant. The extent to which any one would be suppressed and the circumstances under which it would be repressed or indulged would vary from person to person. But it would be determined by general principles of fittingness and unfittingness which all rational beings recognise. The extent to which each would be exercised, and the order in which they would alternate with each other, would be deter-

mined, in accordance with general principles, by the value which each contributes to human life as a whole. (4) There exists in all sane human beings the desire to do what is morally right, as such, and to avoid doing what is morally wrong, as such. This may be called the *conscientious desire*. The following things may be said about it. (i) It is one organising desire among others; and it may conflict with others, e.g. with desire for our own happiness or the patriotic desire, just as those two desires may conflict with each other. (ii) In such conflicts it sometimes wins and sometimes loses, just as desire for one's own happiness sometimes inhibits patriotic action, and patriotic desire sometimes inhibits action for one's own greatest happiness. (iii) So far the conscientious desire is just like any other organising desire. But there are the following three differences. (a) When there is a conflict between the conscientious desire and any other desire we have a peculiar experience which we call "sense of duty" or "feeling of obligation". We feel as if we were being *ordered* to act in one way and *forbidden* to act in another. If the opposing desire overcomes the conscientious desire we have a peculiar emotion which we call *remorse of conscience*. Now we do not have these experience in other cases where one organising desire overcomes another in a conflict between the two, unless conscientious desire is itself supporting the one and opposing the other. Suppose, e.g., that the patriotic desire and the desire for the welfare of humanity in general were in conflict in a particular situation. And suppose, e.g., that the patriotic desire won. The person would feel a sense of obligation during the conflict only if he had decided that it was *right* in this instance to indulge one of these desires and *wrong* to indulge the other. And he would feel remorse afterwards only if the desire which he had indulged (e.g. the patriotic desire) was the one which he believed it to be wrong to indulge in that situation. (b) When we consider the question in a cool hour we think that it is always *desirable* that the conscientious desire should win, and always *regrettable* that it should be overcome in any case of conflict. Now one's attitude towards any other desire is quite different. We should say of *any* desire that it was *unfortunate* that it had to be checked, because the thwarting of any desire is an unpleasant experience. But, apart from that, we do not say of any one desire, e.g. the desire for one's own happiness, that its nature is such that it is always desirable that it should win and always regrettable that it should lose on any occasion when it conflicts with another desire. On the contrary, we say that it depends on circumstances; in some circumstances it is desirable that the desire for one's own happiness should take precedence over the patriotic desire, and in others it is desirable that the opposite should happen. (c) Closely connected with this is the fact that we should all be inclined to say that, in some important sense, it *is* always right for a person to do what he *believes* to be right, and it *is* always wrong for him to do what he *believes* to be wrong. No doubt all kinds of qualifications are needed, e.g. that the person shall have

seriously considered all the aspects of the case, that he shall have sought proper advice, that he shall have taken all possible precautions against his own ignorance or eccentricity or bias. But suppose that all these conditions are fulfilled. Then, even if *A* thinks that *B*'s opinion about what is right or wrong in a given situation is mistaken, *A* will hold that it is right for *B* to do what he believes to be right, and wrong for him to do what he believes to be wrong.

I think that Butler had in mind these peculiarities of the conscientious desire when he said that conscience has supreme *authority* though not by any means always predominant *psychological power* in human nature.

I will sum up this question of the unification and organisation of human personality.

Suppose one were to put to oneself, or to any reasonable person, the following question: "In what way, if any, would you wish your various impulses and desires to be organised? What would be your ideal in this respect for yourself or for anyone whom you wished well?" Suppose that the person addressed considers the question carefully and honestly, after any ambiguities and obscurities in it have been cleared up, and at a time when he is not subject to any violent emotion or desire which is likely to warp his judgment. Then I think that he would be inclined to give the following answer, or at any rate to accept it if one put it to him. "I should wish my life and personality to have some kind of systematic unity, and not to be just a mob of uncontrolled primary propensities or of organising desires which simply take turns with each other without rhyme or reason. I should not wish this unification to arise merely through one organising desire happening to be or to become so much stronger than the rest that they never got a chance to assert themselves and so gradually atrophy. That would be a lopsided personality and a hag-ridden life. I should wish that no desire of mine should be checked except so far as it would be morally wrong to indulge it or the indulgence of it would frustrate some other desire which would contribute more of value to my life as a whole. I should wish that my desire to do what is right and to avoid what is wrong should in every case be strong enough to induce me to do what I believe to be right and to avoid doing what I believe to be wrong, no matter how attractive the other alternatives may be and no matter how repellant in all other respects this alternative may be. Within the sphere of what is morally right or morally indifferent I should wish always to choose what I believe to be more valuable on the whole in preference to what I believe to be less valuable on the whole. I should wish my preference for what is more over what is less valuable to be strong enough to induce me to pursue the former and neglect the latter, even though the latter appeals strongly to my laziness and my desire for immediate passive pleasure, and the former involves exacting mental and bodily work and perhaps danger or unpopularity. Lastly, I should wish that, so far as is humanly possible, my beliefs about what is right and what is wrong, and

about what is more and what is less valuable on the whole, should be *correct*. Even if they are mistaken I should wish to have the strength of character to live up to them; but I should prefer not, through ignorance or wrong-headedness, to be tenaciously engaged in a wild-goose chase."

I have said that I think that any reasonable person would give some such answer as this if he understood the question, considered it carefully and honestly, and were not at the time under the influence of any violent desire or passion. But I do not want to exaggerate the amount of agreement, and I think it is important to remind ourselves of the following facts. People like ourselves, who ask and try to answer such questions, are a very small, and in some respects a rather old, minority of the human race. Most men are, and most men have always been, unintellectual and inarticulate. (By calling them "unintellectual" I do not mean to imply that they are unintelligent, though many of them are that also. It is quite possible to be intelligent without being intellectual, and many unintellectual persons are far more intelligent than many intellectuals.) Now most men have neither the desire nor the capacity to formulate such abstract questions as: "What kind of person should I really wish to be, and what kind of life would really satisfy me?" It is doubtful if they could be made to understand the question, with all the needful explanations and qualifications, if we were to formulate it for them and put it to them. And it is certain that they would be too inarticulate to make a coherent answer to it.

You might say that this does not matter, because we can formulate the question and answer it on their behalf. But I always doubt how far the articulate and intellectual minority are entitled to speak for the inarticulate and unintellectual majority on such a matter. It might be said that we have in ourselves all the necessary data for imagining their mode of life, whilst they have neither the data nor the capacity to imagine the other alternatives which we contemplate and prefer. Therefore our judgment of what is and what is not the most desirable kind of personality and life is founded on a wider basis than their unformulated preferences, and so is more likely to be correct. Again, it might be said that our fundamental desires and emotions are the same as theirs, and that the only differences are on the cognitive side. We can introspect better; we can draw more subtle distinctions and contemplate more alternative possibilities, and deduce more remote consequences, and express ourselves more clearly. So we can formulate explicitly the ideal which is common to them and to us, but which they cannot disentangle and express coherently.

There is obviously a considerable measure of truth in these contentions. There are these non-cognitive resemblances and these cognitive differences between the articulate intellectual minority and the inarticulate unintellectual majority. But I think that two qualifications must be made.

(1) Is it not very likely that there are also fairly deep differences in our desires and emotions and standards of value? It seems reasonable to judge the unformulated ideals of the inarticulate majority by their actions, their conversation, the papers that they read and the films that they enjoy, and the qualities which they ascribe to their heroes and their gods. If we do so, I think we may be inclined to come to the following tentative conclusion. Their main interests seem to be in making as much money with as little effort as possible; in eating, drinking, and exchanging gossip; and in the excitement provided by alcohol or other stimulants, by betting, by taking part in or watching or reading about sport, and by sexual activities and fantasies.

(2) How far is it true that we can imagine for ourselves these other kinds of life which we reject as inferior? The kind of person who has the interests and the capacity to become a moral philosopher is probably rather peculiar in his tastes and desires from the start. In any case he has spent many years in undergoing the special kind of training needed to fit him for his job, and has lived a very sheltered life with a fixed secure income. It seems likely that he can form only a very inadequate and superficial imagination of the satisfactions and disappointments of the ordinary workman or clerk, the tough businessman, the film-star, or the sexual athlete.

I am not saying this in order to suggest that the unformulated popular ideals of life and personality are as likely to be correct as those which are explicitly formulated by moral philosophers and accepted by the intellectual minority. That would be quite absurd. It is obvious that intelligent and interested and technically trained expert onlookers are likely to see more of the game than players who are wholly immersed in it. Often, e.g., we can observe that certain kinds of life which are popularly thought to be highly desirable do not in fact bring lasting satisfaction, and we can see why they cannot do so. And we may be able to notice and to formulate certain by no means obvious conditions for any satisfactory kind of life. All that I want to suggest is the following two warnings.

(1) It is not fair for us to claim that we have the tacit agreement of practically the whole human race to the ideals of personality and life which we explicitly formulate. The practice of the inarticulate majority makes it very doubtful whether their silence can be interpreted as assent.

(2) We must always suspect that there may be a trace of intellectualistic bias in the ideals formulated by intellectuals, even when they have done their best to allow it. But this at least can be said. We *are* aware of the danger of personal, professional, or class bias and we *do* strive to avoid it. We are not sublimely oblivious to it like the plain man; and we do not glory in it and cherish it like the communists and the fascists.

4.349. Conscience

I want at present to give a purely descriptive account of conscience which could be accepted by anyone quite independently of what ethical theories, if any, he might hold.

(1) All civilised languages contain adjectives like "right" and "wrong", "morally good" and "morally evil", or their equivalents. This shows that human beings have from very early times had certain experiences which they took to be cognitions of actions, intentions, motives, etc. as having certain peculiar characteristics, viz. *moral* ones, which can take opposite forms. Again, retrospection assures most of us that we too have had such experiences when we have contemplated certain actions, dispositions, or characters, whether our own, or those of other real persons, or those of fictitious characters in novels or plays. I am not assuming at present that there really are moral characteristics or that we really do cognise them. I am concerned here only with the plain psychological and historical fact that most of us and most of our ancestors back into prehistoric times have had experiences which they *took to be* cognitions of moral characteristics in acts, dispositions, characters, etc. I shall call such experiences "ostensibly moral cognitions".

(2) It is an equally plain psychological fact that, when a person contemplates an action or disposition or character in which moral characteristics seem to him to be present, he is liable to feel certain kinds of emotion which he would not otherwise feel. All languages have words like "remorse", "feeling of guilt", "feeling of moral approval", and so on; and most of us know from our own experiences what such words denote. I shall call these experiences "morally directed emotions".

No doubt morally-directed emotions are nearly always mixed with others which are felt in respect of the non-moral characteristics of the same act or disposition or character. Suppose, e.g., that a friend grants me a favour unfairly at the expense of another person, because he likes me and does not like him. I shall tend to view this act with a non-morally directed emotion of complaisance in respect of its non-moral characteristic of being an act of love and favour towards myself. In such a case it is easy to distinguish the two emotions, because one is *pro* and the other is *anti* and both are directed at the same object. But often we feel a mixture of morally and non-morally directed emotion, both towards the same object and both *pro* or both *anti*. E.g. remorse may be blended with fear of being blamed or punished, and moral indignation may be blended with malice. It is therefore quite possible to think that one is feeling an *unmixed* morally directed emotion when one is really feeling a mixed emotion which contains a non-morally directed constituent. It is even possible to mistake a *purely* non-morally directed emotion for a morally directed emotion, e.g. one of non-moral repugnance for one of moral disapproval. I suspect that people often make this mistake about the

emotions which they feel towards those kinds of sexual desires and actions which make no appeal to them. But the possibility and even the frequency of such mistakes has no tendency to show that these are not morally directed emotions.

(3) It is also a plain psychological fact that the belief that a certain course of action would be right does exert a special attraction or compulsion on most people, and thus provides them with a motive for doing it. It is still more obvious that the belief that a certain course of action would be wrong exercises a certain repulsion or inhibition on most people, and thus provides them with a motive against doing it. All civilised languages have words like "ought", "duty", "obligation", etc. These words all refer to the fact that the supposed rightness of an action gives rise to a motive for doing it, that its supposed wrongness gives rise to a motive against doing it, and that these specifically *moral* motives may conflict with others which arise from one's beliefs about the non-moral characteristics of the action or its consequences. I shall refer to these psychological facts as "moral motivation".

Moral motives are generally combined with and supported by non-moral motives based on the attractiveness or repulsiveness which an alternative derives from the non-moral characteristics which the agent believes it to have. Therefore a person may often think that he is being moved by *purely* moral motives, when really his complete motive includes non-moral as well as moral components. He may even mistake a *purely non-moral* motive such as desire for safety or for the good opinion of his neighbours, for the moral motive of desire to do what is right as such. But the possibility and even the frequency of such mistakes has no tendency to show that there is not moral motivation.

We may sum up these facts by saying that the vast majority of sane adult human beings have and often exercise the capacities of ostensibly moral cognition, of morally directed emotion, and of moral motivation. Now every such person is also capable of *reflexive* cognition, i.e. of contemplating himself, his dispositions, experiences, intentions, motives, actions, etc., from various points of view. To say that a person "*has a conscience*", when that phrase is used in the widest sense, means the following.

(1) That he has and exercises the *cognitive* disposition to reflect on his own past and future actions and to consider whether they are right or wrong; to reflect on his own motives, intentions, experiences, dispositions, and character, and to consider whether they are morally good or bad; and to reflect on the relative moral value of various alternative ideals of character and conduct.

(2) That he has and exercises the *emotional* disposition to feel towards himself and his own actions, dispositions, etc. certain peculiar emotions in respect of the moral characteristics which he believes them to have. Examples are remorse, feeling of guilt, moral self-approval, and so on.

(3) That he has and exercises the *conative* disposition to do what he believes

to be right and to seek what he believes to be good, as such, and to avoid what he believes to be wrong and to shun what he believes to be bad, as such. I shall describe this as "*having a conscience in the widest sense*".

The next point to notice is that a person can have a conscience, in this sense, no matter what ethical theories, if any, he may hold. Ethical theories are concerned in the main with three different, though interconnected, questions, viz. (i) What is the correct analysis of the facts which are expressed by ethical sentences in the indicative? (ii) What kind of non-moral characteristics make an act right or wrong, and what kind of non-moral characteristics make an experience or a state of affairs or a person morally good or bad? (iii) What is the nature of ostensibly moral cognition, and what kind of evidence is there for our beliefs about what is right and wrong, good and evil? Now a plain man with no theories on any of these subjects can have a conscience and act conscientiously. So too can persons who hold the most various theories on these subjects. A man can be a conscientious utilitarian, a conscientious intuitionist, a conscientious holder of the Moral Sense theory, and so on. All that is necessary is that he should believe that, in some way or another, he can form a reasonable opinion about the rightness or wrongness, goodness or badness, of his own acts, motives, intentions, dispositions, etc. and that his opinions on such matters shall be capable of evoking his emotions and influencing his decisions.

Perhaps the only doubtful case is that of the ethical sceptic. Suppose a person has come to the conclusion that words like "right", "morally good", etc. do not really stand for characteristics, as words like "square", "red", etc. do. He holds that moral sentences in the indicative are fundamentally misleading in their grammatical form. On his view, such sentences are really of the nature of interjections or commands; but they masquerade as statements which ascribe certain peculiar characteristics to actions, dispositions, persons, etc. Could a person who held this particular kind of ethical theory be said to have a conscience even in the widest sense?

(1) There would be no reason why he should deny that people who do not hold this theory have consciences. For it is certain that most people *believe* that they are aware of the presence of moral characteristics, and that they *believe* that they make moral judgments. And granted that a person believes this, there is no reason why such beliefs (however mistaken they may be) should not evoke specific emotions in him and influence his conduct. The ethical sceptic will have to regard such emotion rather as a disbeliever in ghosts might regard the fear which a superstitious person would feel in a room which he believed to be haunted. And he would have to regard action which was influenced by such motives as like the action of such a person in putting his head under the bedclothes in order to avoid seeing the ghost.

(2) I think that this analogy shows that even the ethical sceptic himself might

have a conscience in the widest sense. A convinced disbeliever in ghosts might nevertheless feel fear and take precautions if he had to sleep in a room which was said to be haunted; though he would regard such fear and such precautions as unreasonable. In the same way a convinced ethical sceptic might continue to have ostensibly moral cognitions and they might continue to evoke in him certain emotions and to influence his actions. He would if he were consistent regard this as unreasonable; but he would have a conscience in my sense of that phrase.

The next point to notice is that the fact that nearly everyone has a conscience, in this wide sense, does not tend to *support* or to *refute* any particular ethical theory. This is quite a different point from the one that I have just been discussing. It is one thing to say, e.g., that a person could equally well have a conscience whether he *accepted* or *rejected* utilitarianism. It is quite another thing to say, e.g., that a person could equally well have a conscience whether utilitarianism is in *fact true* or in *fact false*. I assert that, on my definition of "having a conscience", *both* these statements are true; and that they would be equally true if any other ethical theory were substituted for utilitarianism.

4.3491. Narrower sense of "conscience". There is no doubt that the phrase "to have a conscience" is often used in certain narrower senses than this. I will now say something about these.

On the face of it one seems to be under obligations of two different kinds. (i) The first is the obligation to maintain and increase the amount of good and to diminish the amount of evil of every kind in the lives of other persons whom we can affect by our actions. I will call this a *teleological* obligation. (ii) *Prima facie* we seem to have other obligations, not derivable from this, which limit it and may conflict with it. E.g. the mere fact that a person has made a promise to another seems to impose an obligation on him to keep it unless the promisee should release him. The obligation seems to be independent of any good that may be produced or evil that may be averted by keeping the promise. I call this an example of an *ostensibly non-teleological* obligation. There seem also to be some ostensibly non-teleological obligations which bear upon the direction and the range of our teleological obligations. Granted that one has a duty to do good to and avert evil from others, it seems obvious to most people that one has a more urgent duty of this teleological kind towards one's parents or one's benefactors than towards complete strangers. There seem to be a number of ostensibly non-teleological obligations, e.g. to keep one's promise, to tell the truth, and so on. And they may conflict with each other and with the teleological obligation to produce as much good and as little evil as possible.

Now the word "conscience" is often used in such a way that conscience is

concerned only with ostensibly *non-teleological* obligations. Suppose, e.g., that a person is in such a situation that he must either tell a lie to *A* or break a promise to *B*, and that he is trying to decide what he ought to do. Suppose he tries to settle this question by direct inspection and without considering the goodness or badness of the consequences of the alternative actions. Suppose that he considers merely the natures of the two alternative actions and their relation to the immediate situation; and tries to judge, simply on these data, whether he has a more urgent obligation to keep his promise or to speak the truth. Then he *will* be said to be using his conscience. But suppose he considers the probable remote consequences of the lie, and compares them with those of the breach of promise, and tries to estimate which would be more good or less bad on the whole. And suppose he decides what he ought to do by reference to those considerations. When he will *not* be said to be using his conscience. I call this narrower sense of the word the "intuitional sense of conscience".

A convinced utilitarian would not have a conscience in the intuitional sense. He would not need to deny that people who are not utilitarians have a conscience in this sense. But he would have to say that all such people are under a delusion when they use their consciences. For, according to him, the only ultimate obligation is teleological; and the ostensibly non-teleological obligations to tell the truth, to keep promises, etc. are really binding so far and only so far as they can be derived from it.

The word "conscience" is sometimes used in a still narrower sense. It is held by some people, not only that there are non-teleological obligations, but that some of them are so urgent that a person ought not under any conceivable circumstances to do an action which would infringe one of these. This claim has been made, e.g., for the obligation to answer a question truthfully if at all. Now I think that the words "conscience" and "conscientious" are sometimes used in such a way as to imply that a person could not have a conscience unless he holds this opinion, and that his conscience is in operation only when his action or refusal to act is based on his belief that one of these unconditional non-teleological obligations is involved.

I think that there is at least one further narrowing of the word "conscience". Sometimes it is used in such a sense that a person would be said to be following his conscience only in so far as he bases his decisions about what he ought to do on some alleged divine revelation. In many cases, I think, this comes to little more than the previous usage decorated with theological frillings. The person regards the pronouncements of his conscience that certain kinds of act would be unconditionally right or wrong as in some sense the voice of God speaking in and to himself. So he can take them to be infallible without arrogating too much to himself. In other cases, however, the situation is quite different. Here the agent regards certain kinds of

act as unconditionally right or wrong, not because he thinks he sees this for himself by direct inspection, but because he believes that God has given a ruling on the matter either in inspired writings or in the traditions of a divinely founded and guided church. I shall call this sense of "conscious', in either of its two forms, the "theological sense of conscience".

I think that it is unfortunate to confine the word "conscience" to the intuitional or the theological sense. In any ordinary use of language one would say that John Stuart Mill and Henry Sidgwick were extremely conscientious men. But both of them were utilitarians, and neither of them had any assured belief in the existence of God. It seems to me most awkward to use "conscience" in such a restricted sense that one must deny that Mill and Sidgwick had consciences and acted conscientiously. Again, the theory that there are unconditional non-teleological obligations, and that one can discover by mere inspection with complete certainty what one ought or ought not to do in a given situation, is almost certainly false. It would be unfortunate to use the word "conscience" in such a way that no one would be said to have a conscience unless he were mistaken on an important point of ethical theory, and that no one could be said to be following his conscience except when he was under the influence of that delusion. For these reasons I shall always use the word "conscience" in the wider sense which I have defined.

Chapter 3

ETHICAL PROBLEMS: RIGHT AND WRONG

Introduction

I will begin by reminding you of some things which I said at the beginning of the course before we embarked on the description of moral psychology.

I said that the raw material of Ethics is *moral phenomena* and that moral phenomena are what we refer to when we use *deontic and evaluatory* sentences in a *specifically moral sense.* As examples of such sentences we may take the following: "You *ought* to keep your promises" and "You *ought not* intentionally to cause needles pain". These are deontic sentences. They should be compared with "You ought to change your clothes if you get wet" and "You ought not to eat peas with a knife". These are also deontic sentences, but they are not used in a specifically moral sense. They should also be compared with "You will often be tempted to break your promises" and "Most people disapprove of the intentional infliction of needless pain". These are not deontic at all; they are what I call *purely expository* sentences. As examples of evaluatory sentences used in a specifically moral sense we may take "Lying is wrong" and "Nero was a wicked man". These may be compared with "It is wrong to wear a white tie with a dinner-jacket" and "Nero was a bad actor". These are evaluatory, but not specifically moral. Lastly, they should be compared with "Lying is a habit which we acquire at school" and "Nero had his mother drowned". These are not evaluatory at all; they are purely expository.

I will also remind you that many sentences and adjectives which seem at first sight to be purely expository, really involve a mixture of pure exposition with an evaluation based on it. As examples I gave such words as "democratic", "reactionary", "propaganda", etc. Cf., e.g., the following sentences, either of which might be used to describe a well-attended meeting of members of the same political party. "An enthusiastic and numerous audience of Mr. X's political supporters listened eagerly to his persuasive address." "A howling mob of Mr. X's partizans greedily swallowed the dope which he handed out."

1. Right and wrong

1.1. Right-inclining and wrong-inclining characteristics

I shall begin with the notions of "right" and "wrong", in the moral sense. The first remark to be made is that we must distinguish between rightness and wrongness themselves and what I call "right-inclining" and "wrong-inclining" characteristics. If an act is said to be right or to be wrong, it is always sensible to ask "Why? What *makes* it right or *makes* it wrong, as the case may be?" The answer that you expect is "It is right because it is the keeping of a promise or the doing of a kindness or so on. It is wrong because it is a breach of promise or an intentionally deceptive answer to a question or so on". I call such expository characteristics as being the keeping of a promise, the doing of a kindness, and so on, "right-inclining". I call such expository characteristics as being a breach of promise, being an intentionally deceptive answer to a question, etc., "wrong-inclining". I call them right-*inclining* or wrong-*inclining*, and not right-*making* or wrong-*making* for the following reason. A lie, as such, always *tends* to be wrong and an act of promise-keeping as such always *tends* to be right. But there may be special circumstances in which it is right to tell a lie and wrong to keep a promise. We shall go into this point more fully later on.

1.2. Rightness and moral justification

The next point is that there are two fundamentally different senses in which the words "morally right" and "morally wrong" are used. This can be shown as follows. Consider the following paradox "It is always wrong to do an act which one honestly believes to be wrong. But some acts which one honestly believes to be wrong are in fact right. Therefore it is sometimes wrong to do a right act". Or again "It is always wrong deliberately to omit doing what one honestly believes to be right. But some acts which one honestly believes to be right are in fact wrong. Therefore it is sometimes wrong deliberately to omit doing a wrong act". How are we to deal with these paradoxes?

(i) There is no doubt that the reasoning in them is formally correct. Therefore, if there is anything amiss, it must be in the premisses. (ii) I think it is plain that there is a sense of "right" and "wrong" in which the second premiss of each argument must be admitted. Anyone who denies it is assuming that each man is infallible in his judgments about the rightness and wrongness of his own acts; and this is surely absurd. (iii) I think that most people would admit that there is a sense of "wrong" in which the first premiss of each argument is obviously true. (iv) The conclusion that one will sometimes be acting wrongly in doing a right act or in deliberately omitting to do a

wrong act is plainly very paradoxical. Since there is nothing formally amiss with the argument, and since each premiss is in some sense true, only one solution is possible. The words "right" and "wrong" must be ambiguous, and they must be used in different senses in different parts of the argument.

The paradox can be cleared up by introducing the notion of "being morally justified". The argument would then run: "One is never morally justified in doing what one believes to be wrong or in deliberately omitting to do what one believes to be right. Some acts which one believes to be wrong are in fact right, and some acts which one believes to be right are in fact wrong. Therefore one is sometimes not morally justified in doing an act which is in fact right or in deliberately omitting to do an act which is in fact wrong". It seems to me that the conclusion is no longer paradoxical.

We can find an analogy to this in logic. Consider the following argument. "A man is never logically justified in strongly believing a proposition if all the evidence available to him is against it, or in strongly disbelieving it if all the evidence available to him is for it. But a proposition is sometimes true when all the evidence available to a person is against it, and it is sometimes false when all the evidence available to him is for it. Therefore a man is sometimes not logically justified in strongly believing a proposition which is in fact true or in strongly disbelieving one that is in fact false". There is nothing paradoxical here.

Now suppose we compare the *rightness* or *wrongness* of a contemplated act with the *truth* or *falsity* of a contemplated proposition. Suppose we compare the *doing* or *the avoiding* of such an act with *strongly believing* or *strongly disbelieving* the proposition. And suppose we compare being *logically justified* in believing or disbelieving a proposition with being *morally justified* in doing or avoiding a contemplated act. Then the analogy becomes clear, and I think it is rather illuminating.

So it is plain that there are at least two common usages of the word "right" and "wrong". In the first sense to say that an act is right means that, if a certain agent were to do it, he would be morally justified in doing it. To say that it is wrong means that, if a certain agent were to do it, he would be morally unjustified in doing it. The second sense of "right" and "wrong" is certainly different. The question whether an act *is* right or wrong, in the *first* sense, depends, in some way and to some extent, on whether an agent who does it *believes* it to be right or to be wrong in the *second* sense.

1.3. Right in the objective sense

It is clear that "right", in the first sense, is more subjective than it is in the second sense. For it involves a reference to an agent's state of knowledge or belief. We might call the second sense the *purely objective* sense of "right".

We will begin with the objective sense of "right". Even so, the words "right", "ought", "duty" etc. are extremely ambiguous, and it is futile to pretend that there is just one *right* sense of "right" and one sense in which we *ought* to use "ought".

I think that the best plan in these circumstances is to proceed as follows. I shall begin by taking an artificially simple example and pointing out exactly what the simplifying assumptions are. Then I shall discuss this example pretty fully and introduce special technical terms to describe the various features in it which are typical of a very wide range of ethical facts. Then I shall gradually introduce the additional complications which generally exist in real life and which I have deliberately excluded to begin with. If necessary I shall modify and supplement the terminology in order to deal with these added complications. Finally it will be easy to distinguish the various senses in which familiar words like "right" and "ought" are used, and to define them in our own technical terms.

1.31. Discussion of an artificially simplified case

We can begin by taking a very simple case which Ross discusses in *The Right and the Good.*[1] One person X has borrowed something, e.g. a book, from another person Y. Almost everyone would agree that, in some sense, this situation gives to Y a *moral claim* on X, viz. to have his book returned in due course by X in almost as good condition as when he lent it. We should also agree that, in some sense, the situation imposes a correlative *moral obligation* in X towards Y, viz. to return the book to Y in due course intact. We all recognise that there is a difference between merely legal and moral claims or obligations. Some moral claims are enforceable by law and others are not. And some claims which are enforceable by law are not in themselves moral claims.

1.311. The simplifying assumptions

Now I am going to begin by making the following simplifying suppositions. (1) I shall suppose that X is aware of the fact that he has borrowed the book from Y and that he recognises that this gives Y a claim to have the book returned intact by his agency. It is evident that a person may be in a situation which undoubtedly would demand a certain kind of action from him if he were aware of the facts, but he may be quite unaware that he is in that situation. E.g. a certain boy Y may in fact be the child of a certain woman X. But X may not know that Y is her child. It would generally be held that the fact that Y is X's child would give to Y a special claim on X for help and kindness, if X were aware of the fact. But has X any special obligation to Y so long as she is ignorant of this special relationship in which she stands to Y? Again, a person may know himself to be in a situation of a certain kind, and we might

1. W.D. Ross, *The Right and the Good* (Oxford, 1930).

agree that, in some sense, it gives rise to a moral claim on him to act in a certain way. But he may fail to see that it does, or may mistakenly believe that it gives rise to a moral claim on him to act in a certain other way. E.g. it is commonly held that being a citizen of a country which is at war gives rise to a moral claim on a person of military age to give his services if they are legally demanded of him. This opinion must be either correct or incorrect. Let us suppose for the sake of argument that it is correct. There are people who know that they are in the situation described, and who fail to see that it imposes this moral claim on them, although, on our hypothesis, it does in fact do so. Is a person under an obligation to act in the way which the situation, by hypothesis, does in fact morally demand of him? You might be inclined to say "*Yes*. For his obligation is to do what the situation *in fact* morally demands of him, whether he recognises that it does so or not." And you might be equally inclined to say "*No*. For a person cannot be under an obligation to do what he does not see to be morally demanded of him. And he is under a positive obligation *not* to do what he believes to be morally forbidden to him."

It is obvious that factual ignorance and error about the situations in which one stands introduce considerable difficulties. And it is obvious that ethical ignorance and error about the claims imposed upon one by situations in which one knows oneself to stand introduce still greater difficulties. Therefore I want to start with a simple case, where the person under consideration knows about the situation which gives rise to a certain demand on him and also recognise that it gives rise to this demand on him.

(2) In actual life a person is nearly always subject to a number of different contemporary and successive claims, and it is literally impossible for him to fulfil them all completely. In some sense he has to make the best compromise that he can between them. This will have to be discussed in detail later. At present I shall make the simplifying supposition that the person under discussion is subject to no other claims beside the one that we happen to be discussing at the time.

(3) A person who intentionally fulfils or intentionally evades a claim on him may be moved to do so by the most various motives. He may *keep* a promise because he believes this to be his duty and he wants to do his duty, or because he wants to obtain other favours in future and believes that if he breaks his promise he will not be trusted again; and so on. He may *break* a promise because he wants to evade an unpleasant duty and thinks he can do so safely; or because he thinks that to keep it under present conditions would seriously injure the person to whom he has made it and he wants to preserve this person from harm; and so on. Obviously the motives which moved a person for or against doing an act are very important from an ethical point of view, and they must be dealt with in due course. But we can consider an intentional

action without considering the agent's motives in doing it. If a person packs, addresses, and posts a book which he has borrowed to the person from whom he borrowed it, we can be practically certain that he *intended inter alia* to fulfil the claim to return borrowed property to the lender. But we may be quite uncertain about what were his *motives* in doing this action. Now the intention, apart from the motive, is ethically important. So I shall begin by considering intentional actions without reference to their motives.

1.312. Discussion of the borrowed book under these assumptions

We will now discuss an example of the borrower, the lender, and the book, under these specially simplifying assumptions. When a person borrows a book there may be no definite time fixed for its return to the lender. And, in any case, it is understood that the borrower will keep it for some time, handle it, read it, and so on. Now so far as concerns the relationship of borrower and lender all that the situation directly demands is the following two interconnected things. (a) That the lender shall get back his book intact within reasonable time or when he asks for it. (b) That this result shall be brought about through the borrower doing an act intended to secure it. The former is demanded *for* the lender. It may be called the *demanded result*. The latter is demanded *of* the borrower. It may be called a *demanded action*. It will be seen at once that there are two different factors in a demanded action, which we may call its *intentional* and its *causal* aspects. The intentional aspect is that it is to be done with the intention of bringing about the demanded result. Its causal aspect is that in the actual situation it is to be such as will in fact lead to the demanded result. You will also notice that the intentional aspect of a demanded action is completely fixed by the nature of the demanded result. If the demanded result is that Y shall receive his book back intact from X, then the intentional aspect of a demanded act is that it shall be done by X with the intention of securing the return of the book intact to Y. And so on in every case. But the causal aspect of the demanded action is only partly determined by the nature of the demanded result. What sort of action by X will in fact secure the return of the book intact to Y depends largely on such factors as the post-office, the railway-companies, Y's present address, and so on. And several different alternative acts might be equally effective, e.g. posting the book to Y, going round to Y's house with it, inviting Y to one's own house and giving it to him there; and so on.

Now during the period while the book is in the borrower's possession many of his acts will have nothing whatever to do with it. We can leave these out of account. The acts which are concerned with the book may be subdivided into two classes, viz. those which are and those which are not directly intended to bring about the return of the book or to evade or prevent its return. Merely taking the book from a shelf, opening it, reading it, and so on, are acts of the

second kind. Packing it, addressing it to the lender and posting it is an act of the first kind. So is selling it to a second-hand bookseller and pretending that it has never been lent to one. Some acts of the second kind may be *relevant* to the demand which the situation makes on the borrower, although they are not intended directly to bring about or to evade the demanded result. If the borrower reads the book with buttery fingers at tea-time, he puts it out of his power to return it in as good condition as when he was lent it. If he takes it out in a canoe he runs the risk of dropping it into the river and thus being unable to return it at all. We may call such acts *indirectly relevant* to the demand of the lender-borrower relationship. Acts which are intended to bring about or to evade the return of the book may be called *directly relevant* to this demand. I shall call an act of the borrower which is intended to secure the return of the book to the lender a *formally claim-fulfilling* act with respect to the lender-borrower situation. Suppose that the borrower continually and wittingly omits to return the book in the hope that the lender may forget about it. This may be called a *formally claim-evading* act. Lastly suppose that he sells it to a bookseller and denies that it was ever lent to him. This may be called a *formally claim-frustrating* act.

Now an act of any of these kinds might in fact lead or not lead to the return of the book to the lender. An act of the first kind may be called *materially claim-fulfilling*. Acts which are materially not claim-fulfilling may be divided into those which positively prevent the return of the book to the lender, and those which fail to bring about either its return or its non-return. These may be called respectively *materially claim-frustrating* and *materially ineffective* with respect to the lender-borrower relationship. Since acts done by the borrower with the book can be divided into four classes in respect of their intention and into three classes in respect of their effect on the lender, and since any of the first possibilities can be combined with any of the second, it is clear that there are 12 possible cases. It would be tedious to illustrate them all. But I will illustrate a few of the less usual ones, partly from this and partly from other examples.

(A) *Indirectly relevant and materially claim-fulfilling.* X has borrowed a stick, and not a book, of Y's. One day X is playing with his dog and throwing the stick for the dog to fetch. The dog has often been taken to Y's house and has been given bones there. He takes the stick in his mouth, desposits it on Y's doorstep, and barks until Y opens the door and thus gets back his stick.

(B) *Formally claim-fulfilling and materially claim-frustrating.* X packs the book and addresses it to Y and stamps it. In order that there may be no risk of a mistake he takes it to the post-office himself intending to register it. On his way he is run over by a lorry, and the book is smashed to bits and covered with X's mangled remains.

(C) *Formally claim-frustrating and materially claim-fulfilling.* X sells the

book to a second-hand bookseller and maintains that Y never lent it to him. The bookseller happens to know that Y is interested in books of that kind and communicates with Y. Y calls at the shop and recognises his own book, and the bookseller hands it over to him.

It is evident that the act which the situation *ideally* demands from X is an act which would be both formally and materially claim-fulfilling. This follows at once from the definitions which I have given of these technical terms. But it is also evident that cases could arise in which no act of X's could fulfil both these conditions, because of limitations in X's knowledge about the causes operating in the external world at the time. This can be most easily illustrated by a different example. Suppose that Y asks X a question about something. What the situation *ideally* demands from X here is that he shall do an act intended to produce in Y's mind a true belief about the subject, and that this act shall lead to such a belief arising. Suppose that Y is a foreigner and that X speaks Y's language rather badly, though he understand the question quite well. X may believe that if he utters a certain sentence S_1 he will convey a true belief, if he utters a certain other sentence S_2 he will convey a false belief to Y. And these may be the only two sentences that occur to him. Now suppose it is the case that S_1 would really convey a false belief or leave Y still puzzled, while S_2 would produce a true belief. If X utters S_1 his act will be formally claim-fulfilling but materially claim-frustrating or claim-evading. If he utters S_2 his act will be materially claim-fulfilling but formally claim-frustrating. If he deliberately holds his tongue his act will be formally claim-evading and materially ineffective.

It is evident from this that this notion of the action demanded from a person by a situation is a kind of ideal notion, like that of a perfect gas or a frictionless fluid in physics. It is, nevertheless, important; because it is the notion of something quite objective and independant of a person's state of factual or ethical opinion. Other notions, which are more concrete and practical, have to be defined in terms of it.

1.3121. Obligations. *(A) Formal.* The next point to notice is this. Suppose the borrower believes that packing, addressing, stamping and posting the book to the lender will secure its return to the latter, and that he does all this with the intention of securing that result. Then there is an important sense of "obligation" in which he has completely discharged his obligation to the lender in respect of this loan. And this is quite independent of whether in fact the lender gets the book back or not as a consequence of the action. The criterion of whether the borrower has fully discharged his obligation, in this sense of the word, is to ask whether the lender is left with any legitimate moral grievance against him in respect of this particular claim. If he is not, the

borrower has discharged his obligation in this sense of the word. If the lender is left with a legitimate moral grievance against the borrower, then the latter has either evaded or broken his obligation, in this sense of the word. Now, provided that the borrower has done a formally claim-fulfilling act, such as packing, addressing, stamping, and posting the book, the lender has no legitimate moral grievance against the borrower in respect of this particular claim if he fails to get the book. He may of course have a legitimate moral grievance against the Post Office or the Railway Company, or Providence. Suppose, on the other hand, that the borrower does a formally claim-evading or claim-frustrating act, such as hiding the book or selling it to a bookseller and denying that it was ever lent to him. Then, even if the book should by that means return to the lender, the latter has a legitimate moral grievance against the borrower in respect of this act. In this sense of "obligation" the borrower had evaded or broken his obligation to the lender. This may be called *formal obligation*. A person is not under a formal obligation to do anything which is not wholly within the control of his will at the time. The reason why the lender has no legitimate moral grievance against the borrower when he fails to get his book back, provided that the borrower has done a formally claim-fulfilling act, is obviously the following. The borrower did all that was in control of his will to secure the return of the book; and the fact that the book failed to return was due to his ignorance or misinformation about certain relevant factors in the external world. Similarly, the lender has a legitimate moral grievance against the borrower when he gets the book back in consequence of the borrower doing a formally claim-frustrating act, for the following reason. The factor which was under control of the borrower's will was intended to prevent the return of the book, and it was only through factors about which he was ignorant or misinformed that the claimed result was secured.

(B) Material. Now there is no doubt that there is at least one other equally common sense of "obligation". A person who did a formally obligatory act which failed to bring about the return of the book would be quite likely to say "I thought, at the time when I acted, that I was doing what I ought. But I see now that I ought not to have done what I did. I ought instead to have done such and such another act, which *would* have secured the return of the book." When a person speaks in this way it is plain that the act which he says that he ought to have done is an act which would have been *materially* claim-fulfilling. He would have to admit that, in his then state of partial ignorance and error about the causes at work, no such act would have been formally claim-fulfilling. For, by hypothesis, he would not have *believed* this act to be materially claim-fulfilling, since he believed this about the different act which he in fact did. Therefore, if he had done what he ought in this sense, he would have done what he ought not in the formal sense. We will call obligation in this sense, *material obligation*.

1.3122. Limits of formal obligations. The next point to consider is exactly what are the limits of the borrower's formal obligation to the lender in respect of the loan. We have already seen that it does not include actually getting the book back to the lender intact. We might be inclined to say that it includes at least handing the book to the lender if he is present, or packing it and posting it or telling someone else to do so if the lender is away. But this will not do. We can at once eliminate putting it in the post; for the borrower might be run over on his way to the post through no fault of his own, and the lender would have no legitimate moral grievance against him in respect of the claim. Suppose then we confine the formal obligation to handing the book over or packing it, addressing it, stamping it, and making arrangements for it to be posted by oneself or another. This amounts to confining it to making certain movements of the hands and the speech-organs which it is reasonable to think will lead to the return of the book. But even this goes too far. It is true that the manual and vocal movements which a person wills generally do follow immediately after his volition to make them. We are therefore inclined to think that they depend wholly and directly on our volitions. But anatomical facts about nerves and muscles would make this very doubtful; and the fact that a man may at any moment be stricken with paralysis or aphasia shows that it is not true. In the end I think we must agree with Prichard that nothing is completely under the control of a person's will at any moment except a certain kind of change in his own mental state.[1] This is the kind of change which is followed almost immediately, unless a man is paralysed or drugged or hypnotised, by willed movements of his hands and his vocal organs. Its immediate consequence is presumably to set up certain changes, about which he knows nothing and over which he has no further control, in certain nerves. If these are in order, which again is out of his control at the time, some kind of physical change travels down them to certain muscles. If these are in order, they contract in certain ways. And, if his hands or vocal organs are in proper order, they move in the way that he has willed or make the sounds that he has willed. It is very important to notice that the integrity of all this elaborate intra-somatic mechanism is as much out of a person's control at any given moment as the telephone-system and the railway-system, and that he knows far less about what goes on in it when it works properly or goes wrong than he does about these external systems of communication. Since nothing is formally obligatory on the borrower which is not completely under the control of his will, his formal obligation extends only to making that kind of change in his own mental state which Prichard would call "setting himself to" pack, address, stamp and post the book to the lender. His formal obligation is to make that kind of change in his own mental state which experience tells him

2. H.A. Prichard, "Duty and Ignorance of Fact", *Proc. British Academy*, vol. 18 (1932). Reprinted in H.A. Prichard, *Moral Obligation* (Oxford, 1949).

has been followed in the vast majority of cases by such movements of his body. Let us call this the borrower's *formally obligatory effort* in respect of the loan-situation.

So far I have been engaged in whittling down the extent of the formally obligatory effort by cutting off everything that is not completely under the control of the agent's will. But there is another respect in which we must be careful not to whittle it down. What I have said so far might suggest that the agent would completely fulfil his formal obligation if he made a certain momentary change in his own mental state. Of course this is not so. In the first place, the bodily process which is needed in order to bring about the demanded result may be prolonged and difficult and painful. Cf., e.g., the case of a doctor who is called to the help of a mountaineer who has met with an accident. The effort that is formally obligatory is still purely mental, but it is not merely that the agent shall produce a momentary change in his own mental state. What he is under a formal obligation to do is to *keep up*, in face of constant temptations to let it stop, a continuous mental process, varying in detail with the varying external obstacles and increasing in intensity with the growing tiredness and weakness of his body. The point is that you have not discharged your formal obligation by merely setting yourself to do what you believe will discharge your material obligation, and then expecting to keep set without further effort. In many cases the hardest part of the formally obligatory effort is to *keep* oneself set to carry out the later stages of an increasingly tiring, unpleasant, and dangerous process. The second point is this. One effort to return the book may be completely unsuccessful, but its failure may leave the book intact on the borrower's hands. E.g. he may have packed it and posted it to the lender's last known address, but it may be returned to him because the lender has gone away and left no address. Obviously the borrower has not thereby completely discharged his formal obligation to the lender, and cannot henceforth just remain comfortably in possession of the book. The failure of one effort leaves him with a formal obligation to try other means, so long as there is a reasonable prospect of their being successful.

The next point to notice is that there is an ambiguity in the notion of the lender's rights corresponding to the ambiguity in the notion of the borrower's obligations. In one sense of "right" the lender has a right to get his book back from the borrower by some means or another; and, if and only if he does, so he has "got his rights" in the matter. This may be called a *material right*. In another sense of "right" the lender has a right to expect the borrower to make, keep up, and if necessary repeat the effort to return the book. If and only if the borrower does this, the lender has "got his rights" from him in the matter. This may be called a *formal right*. The lender gets his formal right in the matter if and only if the borrower sets himself to secure to him his

material right, i.e., if and only if the borrower discharges his formal obligation.

We are now in a position to discuss two questions which have been looming for some time. (1) What happens to an obligation when it becomes known to the agent that it is impossible to discharge it? (2) What happens to it when it becomes known to him that it is not impossible, but extremely difficult, to discharge it?

1.3123. Effects of change of conditions on an obligation. (1) Suppose that the borrower knows that the book has been destroyed in a fire or dropped into the river and fished out again. Then he knows that it is literally impossible to do anything that will give the lender his material right in the matter, viz., the return of the book intact to his possession. Now it is literally impossible for a person to set himself to bring about a result which he knows it to be impossible for any effort of his to bring about. But it is only such an effort which would discharge his formal obligation to the lender and give the lender his formal right. Therefore it is now impossible for the borrower to discharge his formal obligation and impossible for the lender to get his formal right.

In some cases this is almost all that can be said. Suppose that the book was the only copy of a certain ancient manuscript. Then the material right is for ever unobtainable. The formal obligation cannot be discharged and the formal right cannot be obtained. These facts give the lender a legitimate moral grievance against the borrower, if and only if it was the latter's carelessness which led to the impossibility of his fulfilling his material obligation. It is worth noticing that it would be legitimate for him to have exactly the same grievance, neither more nor less, if the borrower had behaved with the same carelessness but it had not led to these disastrous results.

In many cases, however, the changed situation makes new demands; and, if the borrower sets himself to discharge his new material obligation, the lender has no legitimate moral grievance. Suppose that the book has no special sentimental value to the lender, that it has no valuable notes by him scribbled in it, and that a new copy is obtainable by the borrower. Then he is under an obligation to offer to provide a new copy and to set himself to get it to the lender. If he does this, the lender will have no legitimate moral grievance in respect of the failure to discharge the original obligation. If the new copy arrives safely, he will have got something slightly better than his original material rights.

(2) Now take the second case. Suppose that a person has borrowed a cheap and easily replaceable book and has returned it by post to the lender's last known address. The lender has gone away and left no address and the book comes to the borrower through the post. Now the borrower knows that, if he

spent a good deal of time and money, there would be a fair change of tracing the lender and getting the book back to him. We should all think it fantastic to suggest that the borrower is under a formal obligation to spend hundreds of pounds and to employ private detectives in order to return a copy of a *Penguin* book to a lender who had vanished from his last known address. I think that there are two ways of dealing with such a case. (a) Suppose that this were the only claim on the borrower. Still it might be suggested that there is some reasonable proportion between the importance of the claimed result and the intensity and duration of the claimed effort. If failure to bring about the claimed result will inflict only a very small material wrong on the lender, it will be positively unfitting and inordinate for the borrower to make a collossal effort to give him his material right after a normal amount of effort has been made and has failed. (b) In fact there always will be many other claims on the borrower. I am deferring for the present any detailed treatment of a plurality of claims. But we can see at once that, if a person devotes an enormous amount of time and money to the satisfaction of one small claim on him, he will certainly be unable to meet a great many other claims, some of which will be far more urgent.

1.32. Supplementary remarks on claims

I think that the example of the borrower, the lender, and the book brings out most of the important points in the notions of claims and obligations and rights. It is an example of intermediate complexity and it will be as well just to indicate some other examples before attempting to generalise from it.

(a) A simple example is that of a person who asks another a question, e.g. a stranger who stops a person in the street and asks him the way to the railway station. It would generally be agreed that, if the person who is questioned hears and understands the question, the situation demands *for the questioner* that his state of uncertainty shall be replaced by true belief. It demands *from the questioned* an act which will bring about this change in the questioner's state of mind. The former is the material right of the questioner, and the latter is the material obligation of the person questioned. The formal obligation of the person questioned is to set himself to fulfil his material obligation and give the questioner his material right. He will completely discharge this formal obligation if and only if he sets himself to make such sounds and gestures as he thinks will give rise to a true belief about the question in the questioner's mind. If and only if he does this, the questioner will have got his formal right in the matter from the person whom he has questioned.

This case is simpler than the example of the lent book for the following reason. The situation which gives rise to the demand arises at a certain definite moment, and it must be dealt with intentionally in one way or another almost immediately. There is therefore no complication here about acts

which are relevant to the demand but are not directly intended to fulfil it or to evade it or to frustrate it, such as taking out a borrowed book in a canoe to read it.

(b) The folowing example of a claim is in some ways more complex than the case of the borrowed book. It would generally be held that a child, as such, has special claims on his parents. Now the peculiarity here is that what the situation demands of the parents is not one definite act leading to one definite result, such as returning the borrowed book or giving a true answer to the question. What is claimed of the parents is a general line of conduct towards their child which will involve a long series of acts which will differ very much from each other according to varying circumstances. It may include, e.g., occasional beatings and occasional administrations of caster-oil. No particular act discharges the obligation. It would be difficult to say that it is completely discharged at any particular point of the series, though it obviously becomes less and less urgent as the child becomes more able to look after himself. And the obligation might be well discharged by any one of a number of alternative series of acts on the part of the parents.

(c) Some claims arise through the deliberate actions of one or both of the parties concerned whilst others are independent of it. The claim to be given a true answer to a question arises through the deliberate act of the questioner and is independent of previous deliberate action on the part of the person questioned. The obligation to perform a promised action or to return a borrowed article depends upon the deliberate act of the person who made the promise or borrowed the thing. If it be admitted that a child is as such under a special obligation to his own parents, or that a citizen is as such under a special obligation to his own country, it follows at once that a person may be subject to obligations which arise from situations in which he did not deliberately place himself. No one can decide whether he will be born or not; or whether, if he is born, it will be in England or in Germany; or whether, if he is born in Germany, his parents shall be a particular Herr und Frau Schmitt or some other couple. Nevertheless most people would hold that, if a boy is born in Germany to this Herr und Frau Schmitt, this fact imposes on him special obligations to that particular couple and that particular country. I have sometimes heard it denied that there are special obligations of a child to its parents or of a citizen to his country, on the ground that these relationships are not entered into voluntarily. The tacit assumption here is that no relationship can impose a moral claim on a person unless he deliberately entered into it. I do not think that this general principle is self-evident, and it plainly leads to consequences which are completely at variance with commonsense. For, if there are any moral obligations at all, there are few about which commonsense feels more certain than the special obligations of a child to his parents and of a citizen to his country.

(d) The examples of the borrowed book and the question concern only two persons and their relations. We pass beyond this restriction in the case of parents and children and citizen and country. Here a person is under a special obligation to all the members of a certain limited group of other persons, in virtue of some special relationship between himself and them. Such groups may be wider or narrower. At the extreme of width we come to the group composed of all human beings and even all sentient beings. It is generally held that each person is under an obligation to help, or at any rate not to harm except for the sake of some more urgent claim on him, any other person or animal whom his acts may affect. This may be called the *claim of humanity* or the *obligation of general beneficence.*

It should be noticed that one's obligations to a group seem to be of two kinds, which might be called *distributive* and *collective.* A distributive obligation is one towards any individual member of the group equally. E.g. a person has a distributive obligation to his children. A collective obligation is towards the group considered as a collective unity composed of individuals interrelated in certain definite ways. This is the kind of obligation which an officer or a private soldier has to his regiment. No doubt, when one has a collective obligation to a group as a collective unity, this nearly always involves distributive obligations to any individual member of the group. But, as a rule, one will have different obligations to different members of such a group, depending on the position which each occupies in the group. Thus an officer's duty to his regiment involves different obligations towards those individuals who are his superior officers, towards those who are officers under him, and to the private soldiers.

(e) I have talked about what a situation "demands for" a person, what he has a "claim to", what he has a "right to", and so on. This may suggest that the claimed result is always something which the person for whom the situation demands it would be *glad* to get. No doubt this is true in many cases. A person who has lent a thing does generally want to have it back. And the obligation to return it does vanish if the lender tells the borrower that he no longer wants it back. In fact what is claimed for the lender, strictly speaking, is to have his loan returned unless he makes plain that he no longer wants it.

But it is not safe to assume that in every case the result which the situation demands for a person is something which he would welcome. Suppose, e.g., that a person who has authority over another, e.g., a parent or a schoolmaster, has forbidden a certain action to his children or his pupils and has stated that he will inflict a certain penalty for any breach of the rule by them. Then if they break the rule the situation demands that they shall receive the penalty which they have wittingly incurred. The authority has a moral obligation to inflict it, and the culprit has a moral claim on the authority to have the penalty inflicted upon him. But naturally in many cases he would prefer not

to get what he has a right to. It is important to be clear that I am using the words "right" and "claim" to cover what the situation demands for a person, even when he would much rather not be given his objective rights and would much rather that the other party concerned should evade the formal obligation to set himself to give them to him.

(f) A person is often said to have duties towards *himself*, e.g. a duty to develop his mental and physical powers and not let them lie fallow or run to seed. Some people would say that such duties are really obligations which we are under to other men or to God. If so, there is no need to say anything special about them. But I am not persuaded that this is a correct account of them. It is not clear to me that, if there is no God, a solitary Robinson Crusoe, with no hope of being rescued, would be justified in letting his mind and body go to seed. Now what I have said about claims and duties has so far assumed that there at least two persons concerned, viz. an agent and a claimant. Perhaps the best way of dealing with duties towards oneself is to say that the agent here is the person as he now is and the claimant is the person as he will be in future. No one would feel any difficulty in holding that intending parents have duties in respect of an as yet unborn child. They may be under an obligation not to have a child at all if it would be very likely to inherit some disease or defect. And, if they are going to have a child, they have a strong obligation to act in such ways as will give it the best chance to be born healthy and sane. I suggest that one might think of a man's present obligations to his future self as somewhat analogous to an intending parent's obligations to his unborn child.

1.33. Removal of simplifying assumptions

1.331. Ignorance and error

We will now begin to remove the simplifying assumptions with which we started. We will begin by considering the bearing of a person's possible ignorance or error on his obligations. We must divide this into two parts, viz. *purely factual* and *purely ethical* ignorance and error. When Oedipus married Jocasta, not knowing her to be his mother, he acted in ignorance of a certain relevant fact about blood-relationship. He was not ignorant or in error about the obligations which such a situation imposes, but was ignorant of the fact that he and Jocasta stood in that relationship. Suppose that a person knowingly hurts or teases an animal for amusement, and thinks that the situation demands no other treatment from him. Then he is ethically ignorant or mistaken. He makes no mistake about the fact that the animal can feel pain or annoyance and that his action is causing it to do so. But he fails to see that, in spite of the fact that animals differ profoundly from men, the mere fact that they are sentient and conative beings gives them a moral claim to be treated with consideration.

1.3311. Factual. For the present purpose we may divide factual ignorance and error into two kinds. (a) About the *effects* which this, that, or another of one's possible exertions will have in a given situation. (b) About the *nature of the situation* in which one is placed. We have already seen that the first kind of ignorance and error may have a great influence on whether efforts which are formally claim-fulfilling and therefore discharge one's formal obligation will also be materially claim-fulfilling and therefore discharge one's material obligation. But it has no bearing on whether one's efforts are formally claim-fulfilling. So we need not take further account of it here.

1.33111. Factual ignorance of situations. Factual ignorance about situations which in fact make demands on one is of the following kind. X does in fact stand in a certain relation to a certain other person Y who has in fact done or suffered certain things in the past and is in a certain state now. E.g. Y may in fact be X's mother, or may have helped X in his career, or may have a disease in the treatment of which X is the only available expert. If X were aware of these facts about Y there is no doubt that the situation would demand from him certain characteristic kinds of action. But, we will now suppose, X is not aware of Y's existence; or, if he is, he is not aware of these facts about Y. The question is: What is the ethical relevance of this ignorance of X's about Y? There is no doubt that we should often be inclined to say that X may have an obligation to Y in respect of a certain situation, although he is unaware of the situation which imposes the obligation. X might say "I didn't know that I had any special obligation to Y; but now that I have discovered that she is really my mother and has been seriously ill, I see that I have been under an obligation to help her for years past." On the other hand, we should also be inclined to say that no act done or omitted by X during his period of ignorance was a breach of his obligation.

I think that the best way of putting the case is as follows. The situation does demand a certain kind of action from X and a certain kind of result for Y. So X is under a material obligation to do an act which will bring about this result for Y, and Y has a material right to have this result brought about in him. But X has not a formal obligation to act with the intention of bringing about this result for Y. For he is unaware of the situation which gives rise to Y's material right, and therefore he cannot act either with the intention of securing it to Y or withholding it from Y. And he cannot be under a formal obligation to do what is impossible to him. Now let us look at the matter from Y's point of view. Plainly Y has no legitimate moral grievance against X for failing to attempt to secure him his material rights in the matter, so long as X is wholly unaware of the situation which demands these rights for Y. Therefore Y has no formal right to expect X to make an effort intended to give him these material rights.

If we like, we may say that in such a case X has a *potential* formal obligation to act with the intention of securing to Y his material right and that Y has a *potential* formal right to expect such action from X. This means only that everything necessary to impose this formal obligation on X and to give this formal right to Y already exists except X's knowledge of the situation. If ever and whenever X becomes aware of the situation which imposes this material obligation on him and gives this material right to Y, the corresponding potential formal obligation and potential formal right will become actual.

1.33112. Factual error about situations. We can now pass from factual ignorance to factual error. The latter is more positive than the former. We must now suppose that X believes himself to stand in a certain relationship to Y and believes that Y has done or suffered certain things in the past or is in a certain state now. If the situation were as X believes it to be, it would demand that X should bring about a certain change in Y's condition. And X knows that it would. But X is mistaken about the facts. E.g. X receives a letter purporting to come from his old nurse and saying that she is ill and in want. He believes that the letter comes from her and that it contains a true account of her present state. But really the old nurse died some time ago, and a dishonest relative, finding that X has helped her in the past, has written a lying letter in her name. Plainly X has no material obligation to help the nurse, and the nurse has no material right to be helped by him. On the other hand, he has a material obligation to refuse to send money and to bring the writer to justice. And the writer has a material right to be refused help and to be punished. But suppose that X is fully satisfied that the letter is genuine. Then there is certainly a sense in which he evades an obligation if he makes no attempt to send help. I think we must say that he is under a formal obligation to set himself to discharge what he knows *would* be his material obligation *if* the situation were as he mistakenly believes it to be. On the other hand no one has a formal right corresponding to this formal obligation. Plainly the deceased nurse has no formal or material right in the matter. As regards the dishonest letter-writer he has a *material* right to be refused money and to be punished. But he has no *formal* right to be refused and brought to justice by X, since X is ignorant of his existence and of his relationship to himself. So X has a formal obligation to set himself to bring about a certain result, viz. the relief of the old nurse, which is (a) impossible, since she is dead, and (b) corresponds to no material or formal right. On the other hand Y, the writer of the letter, has a material right to be refused money and punished, but has no corresponding formal claim on X, since X is ignorant of the relevant facts and cannot act intentionally in regard to them.

It will be worth while to compare and contrast the case when X is aware of the situation and is in error only about the effect that his action will have with

the case where he is in error about the situation. In each case there arises a *dislocation* between his formal obligation and the other party's material right. In the first case the act which discharges his formal obligation fails to give the other party his material right owing to the agent's mistakes about causation; in the second case it fails owing to his mistake about the nature of the claim-imposing situation. In the first case, however, Y has a *formal* right or claim on X corresponding exactly to X's formal obligation. But in the second case there is the further dislocation that Y has no formal right or claim on X corresponding to X's formal obligation.

1.3312. Ethical ignorance. We can now consider the relevance of purely ethical ignorance and error about claims and obligations. We will begin with ignorance. The situation here is as follows. X knows or correctly believes that he stands in a certain relationship to Y, and knows that Y has done or suffered certain things in the past or is in a certain state at present. These facts are such that they demand a certain kind of action from X leading to a certain change in Y's condition. But X fails to recognise that they demand this change for Y. Suppose, e.g., that X is a rather stupid and ignorant peasant and that Y is an animal whom he finds very badly injured in a trap. X correctly believes that the animal is in great pain and that it will continue to be so until it dies after many hours if it is left in the trap. He knows that he could put it out of its misery at once by knocking it on the head. I assume that this intention demands from X that he shall knock Y on the head, and demands for Y that it shall be put out of its misery. X however does not recognise any such material obligation on himself or any such material right in animals. He watches the animal for a while and then deliberately avoids taking the trouble to kill it and walks away. Undoubtedly he breaks or evades his material obligation. What about his formal obligation?

I said that a person's formal obligation is to do an act which is intended to fulfil the material obligation which the situation would impose on him if it were as he believes it to be. If a person fails to see that a situation imposes any obligation on him, he cannot act with the intention of fulfilling it or of evading it. Now he is not under a formal obligation to do what is impossible. Therefore we must say that no act which such a person may do either fulfils or breaks his formal obligation in the matter. He just has no formal obligation. And the animal in the trap has no formal right or claim against him.

But we also said that the criterion of whether X had discharged his formal obligation to Y was whether Y would be left with any legitimate moral grievance against X in respect of the situation. We cannot very well apply this test in the case of an animal. But suppose that Y were a negro who had met with an accident, and that X, who was present and could have helped, does not recognise any obligation to help suffering negroes. Should we not be in-

clined to say that an injured negro, whom X left unhelped, would have a legitimate moral grievance against X in spite of X's moral ignorance? I do not doubt that he would *in fact* feel such a grievance, but I do not think that it would be *legitimate*. The truth seems to me to be as follows. The negro would be extremely *unfortunate* in that the only person available to help him had a certain defect which is not very common and which prevents him from giving help. But he would also be very unfortunate if the only person who had passed near to the scene of the accident had been deaf and had thus failed to hear him shouting. The difference between the two cases is that in the former his misfortune depends on what is in one sense a *moral* defect in the other man, whilst in the latter it depends on a purely physical defect. The defect is *moral* in the sense that it is a defect in this person's powers of *moral cognition*, though not necessarily in his powers of *moral conation*. He may be extremely ready to make every effort to discharge those moral obligations which he recognises. Therefore all that we can say is that the negro might legitimately complain of his misfortune in respect of the moral cognitive defects of the only person who happened to come near him. This is entirely different from the legitimate moral grievance which he would have against a man who recognised that suffering negroes have a right to be helped but made no attempt to carry out his material obligation in the matter. It is very likely that the negro would in fact feel this kind of grievance, because he would not be sure that X's failure to help him sprang simply from moral ignorance and not from a deliberate neglect of his formal obligation. But, if the facts were as I am supposing, he would not be justified in doing so.

1.3313. Ethical error. We can now consider ethical error, which is something more positive than ethical ignorance. We start, as before, by supposing that X knows that he stands in a certain relationship to Y and knows that Y has done or suffered certain things in the past, or is in a certain state at present. These facts are such that they impose on X a material obligation to bring about a certain change α in Y's condition. But X mistakenly believes that they impose on him a material obligation to bring about a certain different change β in Y's condition.

It is not very easy to give examples of purely ethical cases of this kind which will satisfy most people. The following is the best I can think of. Many people have held that if X is publicly and deliberately insulted by Y and Y refuses to apologise, X is under a material obligation to challenge Y to a duel in which he will try to wound or kill Y, taking a corresponding risk himself. Many other people have held that X is under a material obligation to forgive the injury. One of these ethical opinions must be mistaken. Let us suppose, for the sake of argument, that the gentlemanly opinion is false and the Christian one true. Let X be an officer in the French or German army who sincerely be-

lieves that he is under a material obligation to fight a dual with *Y* who has insulted him and refuses to apologise. We are assuming that really *X*'s material obligation is to forgive *Y* and not challenge him to a dual. What further can be said about this situation?

The first thing to notice is that there is a certain ambiguity in our definition of "formal obligation" which now becomes troublesome. I said that *X* discharges his formal obligation to *Y* provided that he does an act which he believes will give *Y* his material right. But this might mean two things. (i) That *X* does an act which he *believes* (rightly or wrongly) will produce a certain change in *Y*'s condition, and that this change (whether *X* believes it or not) is *in fact* the one to which *Y* has a material right. (ii) That *X* does an act which he *believes* (rightly or wrongly) will produce a certain change in *Y*'s condition, and he *believes* (rightly or wrongly) that this change is the one to which *Y* has a material right. The ambiguity has not troubled us before because we have been supposing hitherto that *X*'s *ethical* belief was correct, even though his non-ethical beliefs might be mistaken. But now it becomes highly important. Consider an example. If *X* forgives *Y* he intentionally gives to *Y* what is in fact *Y*'s material right. But in doing this he is intentionally giving to *Y* what he believes that *Y* has no right to have, and is intentionally withholding from *Y* what he believes that *Y* has a right to have, viz. a challenge. Suppose that *X* challenges *Y*. Then he intentionally gives to *Y* something which he believes to be *Y*'s material right. But, in doing so, he is in fact (through ethical error) doing a material wrong to *Y*. I think it is plain that there is a sense of "obligation" in which we should say that *X*, who believes that he has a material obligation to challenge *Y*, *is* under an obligation to challenge *Y*; and that he breaks his obligation if he forgives *Y*, although in fact *Y* has a material right to be forgiven.

I propose to introduce the term *subjective obligation* to cover this case. *X* fulfils his subjective obligation if and only if he does an act which he *thinks* will produce a result to which he *thinks Y* has a right. I shall henceforth use the phrase *formal obligation* in the following sense. *X* fulfils his formal obligation if and only if he does an act which he thinks will produce a certain result, that result being the one which *Y really would* have a right to *if* the situation were as *X* believes it to be. As we have seen, if *X* makes ethical mistakes, it is quite possible that an act of his which would discharge his subjective obligation would break his formal obligation, and vice versa.

The following points are worth asking.

(1) The notion of material obligation may be called purely objective, for it takes no account of the agent's possible ignorance or error, either about matters of fact or about ethical demands. The notion of formal obligation may be called factually subjective and ethically objective. For it allows for the possibility of the agent's error or ignorance about matters of fact. But it is

concerned with what the situation *really would* demand if the facts were as the agent believes them to be; and not with what he thinks it would demand, if the two should differ. The notion of subjective obligation is both factually and ethically subjective.

(2) There is a close parallel in logic or epistemology to these three notions. Take the following three notions. (i) What is logically entailed by a set of premisses which are in fact true. (ii) What is logically entailed by a set of premisses which X believes to be true. (iii) What X believes to be logically entailed by a set of premises which X believes to be true. The first is a purely impersonal and objective notion, and it may be compared with the notion of what is *materially obligatory* in a certain situation. The second is what X would be logically bound to accept if he committed no *formal* fallacy. It might happen to be false, because some of the premisses which he believes are false. Or it might happen to be true, because the errors in the premisses which he believes cancel out. But in any case he will be *logically consistent*. This corresponds to the notion of what is *formally obligatory* in a certain situation. The third is what X would have to believe in order to be *psychologically consistent with himself*. If he believes p and believes that p entails q, then he would be psychologically inconsistent if he refused to believe q or rejected it. And yet q may not follow from p; so that his psychological consistency may involve logical inconsistency. This corresponds to the notion of what is *subjectively obligatory* in a certain situation.

(3) How would the test of whether Y has a legitimate moral grievance after X has acted work in this case? We must now go back to our example about the insult and the challenge and look at it from X's point of view. Let us assume, as before, that what the situation really demands for Y is forgiveness and not a challenge; and that what X thinks it demands for Y is a challenge and not forgiveness. (i) Suppose that X fulfils his subjective obligation and challenges Y to a dual, thus breaking his formal and his material obligations. Would Y have any legitimate moral grievance against him for this action? I do not think that he would. Y could legitimately complain that he had been unfortunate in being done out of his right to forgiveness and forced to risk his life or incur disgrace, and he would say that this misfortune was due to a *cognitive* moral defect in X. By hypothesis X did not exhibit a *conative* moral defect. He tried to give to Y what he believed to be Y's right in the matter. (ii) Suppose that X breaks his subjective obligation, and keeps his formal and material obligation, by forgiving Y although he believes that he ought to challenge him. Would Y have a legitimate moral grievance against X in the matter? Y could certainly congratulate himself on the fact that a combination of ethical error and moral weakness on X's part had happened to give him his material right in the matter. But he might nevertheless feel a legitimate moral

grievance against *X*. For the essential point is that *X* deliberately tried to do *Y* out of what he took to be *Y*'s right, and failed only because he was mistaken as to what these rights were. So it seems to me that, where formal and subjective obligations differ, the question whether the other party has or has not a legitimate moral grievance against the agent is a test for whether the agent has discharged his *subjective* obligation.

1.3314. Formal classification of all possibilities. We can now classify acts done by an agent in a situation into eight possible kinds in the following way:

(.111) An act which will *in fact* bring about a certain change, which the situation *as it really* is does *in fact* demand.

(.110) An act which will *in fact* bring about a certain change, which the situation *as it really* is *would appear* to the agent to demand, if he were aware of it.

(.101) An act which will *in fact* bring about a certain change, which the situation *as it appears* to the agent would *in fact* demand if it were as it appears to him to be.

(.100) An act which will *in fact* bring about a certain change, which the situation *as it appears* to the agent *appears* to him to demand.

(.011) An act which the agent *thinks* will bring about a certain change, that being the change which the situation *as it really* is does *in fact* demand.

(.010) An act which the agent *thinks* will bring about a certain change, that being the change which the situation *as it really* is would *appear* to the agent to demand if he were aware of it.

(.001) An act which the agent *thinks* will bring about a certain change, that being the change which the situation *as it appears* to the agent would *in fact* demand if it were as it appears to him to be.

(.000) An act which the agent *thinks* will bring about a certain change, that being the change which the situation *as it appears* to the agent appears to him to demand.

It is evident that these eight alternatives cover all the theorethical possibilities. Suppose we start by ruling out error and ignorance about causation. Then (.011) reduces to (.111); (.010) reduces to (.110); (.001) reduces to (.101); and (.000) reduces to (.100). For in that case the act which the agent thinks will bring about a certain change will in fact do so. Suppose we then proceed to rule out error and ignorance about the nature of the situation. Then (.101) reduces to (.111) and (.100) reduces to (.110). For in that case the situation as it appears to the agent will be identical with the situation as it really is. Suppose finally that we proceed to rule out ehtical ignorance and error. Then (.110) reduces to (.111). For in that case the change which the situation appears to the agent to demand is identical with the change which it really does demand.

Now it will be noticed that (.111) is the act which is *materially obligatory* on the agent and the one which will secure to the other party his *material right*. On the other hand (.000) is the act which is *subjectively obligatory* on the agent. It is thus evident that the subjectively obligatory act for a given agent in respect of a given situation would necessarily coincide with the materially obligatory act if and only if he were fully and correctly informed about (a) the relevant causes operating, (b) the nature of the situation, and (c) the kind of change which such a situation demands. If any of these conditions breaks down, the subjectively and the materially obligatory act would coincide only be luck, e.g. by the effects of different errors or ignorances happening to cancel each other out.

Of the intermediate cases between (.111) and (.000) I think that only (.101) and (.001) are worth further consideration from an ethical point of view. (.101) is an act which will *in fact* bring about a certain change which the situation as it appears to the agent *would in fact* demand if it were as it appears to him to be. (.001) is an act which the agent *thinks* will bring about a certain change, that being the change which the situation *as it appears* to the agent *would in fact* demand if it were as it appears to him to be. The peculiarity of these two cases is this. They both exclude *purely ethical* error. They both allow for the possibility of the agent being mistaken about the factual character of the situation. The second allows also for the possibility of his being mistaken about the effects which his action will have. I think that we can confine our attention to the second. If we are going to allow for the possibility of factual error at all, it is reasonable to allow for the possibility of being mistaken about the effects of one's action as well as for the possibility of being mistaken about the factual nature of the initial situation. Now I think that the notion involved in (.001) is important. This is what I call a *formally obligatory* act. It is the act which an agent, who might be mistaken about the effects of his action and about the factual nature of the initial situation, but who made no purely *ethical* mistake, would do, if he acted with the intention of giving the other party his material rights in the situation as he sees it. It is important, as compared with a subjectively obligatory act, for the following reason. A formally obligatory act may lead to great wrongs being inflicted, and people who are better informed than the agent about the facts of the situation and the laws of nature may see that it will do so. But this would not indicate any *moral* defect in the agent. On the other hand, a subjectively obligatory act may, and often does, inflict great wrongs simply because the agent is morally stupid or morally crazy. Often there is no reason to think that the doer of a subjectively obligatory act had made any serious mistake about the facts of the situation or the effects of his actions. If such an act then inflicts great wrongs, it can only be because the agent was a moral imbecile or a moral maniac. Of course actions done by conscientious persons which

inflict great wrongs often arise from a combination of factual error or ignorance with moral stupidity or craziness. I should suppose that both factors entered into many of the subjectively obligatory acts of material wrong-doing done by conscientious and fanatical pedants like Robespierre or Lenin.

1.3315. Ambiguities of "right action". We can now deal with some of the ambiguities of the word "right" in the phrase "right act in a given situation". Whatever further ambiguities we may discover later, there is no doubt that there are at least the following three. It hovers about between the three notions which I have called "materially obligatory", "formally obligatory", and "subjectively obligatory". When we are thinking mainly, not of the agent, but of the other party and what the situation demands for him, we tend to use "right" in the sense of materially obligatory. The right act is then thought of as the act which will give this party his material *rights*. When we are thinking mainly of the agent and his possible errors and ignorance about the facts of the situation and the effects of his acts, we tend to use "right" in the sense of formally obligatory. The right act is then thought of as follows. It is the act which *would* give to the other party the rights which *would* be his if the situation were as the agent takes it to be, and if his acts would have the effects which he thinks they will have. When we reflect further, we see that the agent may be ethically mistaken, and yet, in a certain sense, be under an obligation to act on his principles, however false and pernicious they may be. When we have this in mind we tend to use "right" to mean subjectively obligatory.

It is often said that any person, however ignorant, can *know* what is right, and however weak, can *do* what is right. This is true if and only if "right" is used in the sense of subjectively obligatory. Any person, however ignorant, can form some opinion, however absurd, about the nature of the situation in which he stands, about the sort of change which such a situation would demand, and about the sort of action on his part which would bring this about. When he has formed an opinion on these points he of course *knows* what it is. Therefore he knows his subjective obligation. For this is to make that effort which he thinks will produce the change which he thinks is demanded by what he thinks to be the situation. And, in one very important sense of "can", a person always *can* make such an effort.

1.332. Plurality of claims

We must now remove the artificial supposition that the agent has only one claim on him at a time. In fact he always has several, and they may conflict with each other. Two claims may be said to conflict if any action which would tend to fulfil one of them would tend either directly or indirectly to frustrate

or to prevent the fulfilment of the other. We will give some typical examples of this. (i) *A* may have been told something in confidence by *B* and he may now be asked a certain question by *C*. The question may be such that, if *A* answers it truly or even refuses to answer it at all, he will give away to *C* the information which *B* gave him in confidence. Now *C* as questioner has a right to receive a true answer from *A*. And *B*, as having told *A* something on promise of secrecy, has a right to have his confidence respected. Here we have a perfectly direct conflict between two claims arising out of different relations to different persons. (ii) *A* may have promised to do something for *B*. When the time for fulfilling his promise arrives *A* knows that the fulfilment of this promise would be very harmful to *B*. He points this out, but *B* does not believe him or is recklessly willing to risk the bad consequences and refuses to release *A* from his promise. Now *B* has a claim to have the promise fulfilled, and he also has a claim not to be injured unnecessarily. Whether *A* keeps his promise or breaks it he will frustrate one or other of these claims. Here we have a direct conflict arising out of two different facts concerning the same person. One is the fact that a promise has been given and that the promisee refuses to release the promisor. The other is the fact that the promisee is a sentient and conative being capable of feeling pain and sorrow and of suffering intellectual or moral damage through the action of the promisor. (iii) *A* is a man with a widowed mother not very well off and several children. If he makes an allowance to his mother it will make her declining years comfortable but he will not be able to send his highly intelligent son to the university. If he sends his son to the university he will not be able to do anything appreciable for the comfort of his widowed mother. Now his mother has a claim on him, based on the maternal relationship and on past efforts and sacrifices made by her for his sake. His son has a claim on him based on the fact that he has brought him into the world and that he needs to be fed and clothed and educated until he can fend for himself. Owing to *A*'s limited means neither claim can be fully met without very largely evading the other. This is a typical case where the conflict arises, not through anything in the nature of the claims as such, but from the fact that the agent's resources are limited. (iv) Conflicts of the same general nature as this would arise even in a society in which aged mothers were supported and children educated by public funds and not by individuals. Any person has only a limited amount of time and energy at his disposal. If he spends too much of this in attempting to fulfil one obligation he will not have enough left to fulfil other obligations which may coexist with this one or may be going to arise in the future. E.g. the average don is under an obligation to prepare lectures, to supervise certain undergraduates, to carry on research in his subject, and to help in the administration of his college and the university. There is no intrinsic opposition between these claims. The first three activities to some extent fit in with each other; and, if

he happens to be an economist or a political theorist or a psychologist, the fourth may also help him with the other three. But, since they all take up a great deal of his limited time and energy, they *must* conflict to some extent, and they *may* conflict very seriously.

Now it is evident that some claims are morally more urgent than others. E.g. a hostess who has invited a person to a dinner-party which he has accepted has a claim on him to arrive at the house at the appointed day and time and not to disarrange the table by his absence. A patient suddenly stricken with appendicitis has a claim on his doctor to attend to his as quickly as possible. Now the same man may have accepted an invitation to dinner at 7.45 on a certain day and may be a doctor one of whose patients is stricken with appendicitis at 7.15 on that day. No one doubts that the patient has a much stronger claim on this man as his doctor than the hostess has on him as her guest. But in other cases it is not at all clear which of two obligations is the more urgent. Cf., e.g., the man of limited means who has a widowed mother requiring comforts and has an intelligent son who would profit by a university education.

1.3321. Component obligations and resultant obligation. The existence of a plurality of conflicting claims makes it necessary to introduce some further complications into our terminology for dealing with the notions of "right" and "obligation". When several claims conflict, there is a sense in which one's obligation is to make as good a compromise as one can between the various "obligations", in the sense already considered, corresponding to these various claims. And there is a sense of "right action" in which the right action is the one which makes as good a compromise as possible between the various competing claims. We can deal with this by distinguishing between various *component obligations*, each connected with a different claim on the agent, and a single *resultant obligation* to make as satisfactory a compromise as possible between these claims when they cannot all be wholly fulfilled. If one ever were subject to only a single claim, one's resultant obligation would be identical with the component obligation corresponding to that claim. In cases where claims conflict only because of the limited time or means or energy of the agent the compromise may consist in doing *something* towards fulfilling all the claims, but not fulfilling any of them as completely as would be ideally desirable. But in some cases one has simply to ignore the less urgent claims altogether, and act as if only the more urgent one existed. This, however, is an extreme case. You might be inclined to say that the doctor has simply to ignore the claims of his hostess and act as if the only claim on him were that of his patient. But this is not quite true. He may have to do this at the moment; but he is under an obligation to ring up the hostess and explain and apologise at once if he can. And, if he cannot, he is under an obligation to write and explain as soon as he can afterwards.

1.3322. Teleological and ostensibly non-teleological obligations. For the present purpose we may classify component obligations as follows. (1) There is the obligation to produce pleasant and other kinds of good experiences in other people and to prevent unpleasant and other kinds of bad experiences from occurring in them and the obligation to improve their character and dispositions. (2) There are other component obligations which are *prima facie* not reducible to these, e.g. the component obligation to answer questions truly, to keep one's promises, and so on. No doubt in many cases a person will be benefitted by being answered truly, and by being given what he has been promised, and so on. It is also true that he will always suffer the displeasure of disappointed expectation if he finds that he has been told a lie or if a promise to him is broken. But it certainly does not seem that these are the reasons why a person who is asked a question is under a component obligation to answer truly, or why a person who has made a promise is under a component obligation to keep it. It seems as if the reason in the one case were simply the fact of being asked a question, and in the other case the fact of having made a promise. There are cases where one has reason to believe that a true answer will do far more harm than good, and that a certain kind of lie will do more good than harm. And yet one feels that the questioner has a claim on one to a true answer, though this may of course be over-ridden by his claim on one to receive kind treatment. Suppose, e.g., that an officer is the sole survivor of an action in which one of his men displayed disgraceful cowardice. This man's mother writes to the officer and asks him for details of her son's death. If he describes them truly, her whole life will be made miserable and no one will be any the better. If he tells a certain kind of picturesque lie, she will be comforted and no one will be a penny the worse. I think that everyone who was not dying in the last ditch for an over-simple theory would say that the officer was subject to two conflicting component obligations of different origin, viz. an obligation to give a true answer, based simply on the fact that he has been asked a legitimate question; and an obligation to avoid giving useless pain, based on the fact that the questioner is a sensitive and affectionate being.

I propose to call the two kinds of component obligation which we have distinguished *teleological* and ostensibly *non-teleological*, respectively. I call the obligation to keep promises, as such, and so on, *ostensibly* non-teleological and not simply *non-teleological* for the following reason. It is an essential part of utilitarianism that the ground of *all* obligations is really teleological, and I do not want to use expressions which would rule out this theory by definition. It is worth while to notice that a very important sub-division of ostensibly non-teleological obligations is connected with limitations on the teleological obligation. These may be called *distributive obligations*. It is commonly held, e.g., that a person is under a more urgent obligation to do

good and to prevent harm to his parents or his children than to strangers. Again it is held that, among those persons to whom he has an equally urgent obligation to do good, e.g. his children, he is under an obligation not to make arbitrary preferences. No one is inclined at first sight to believe that these distributive obligations can be deduced from the teleological obligation to produce as much good and as little evil on the whole as possible. Most people would admit that there might be cases where more good would be produced by breaking one of these distributive obligations than by fulfilling it and yet that you might be under an obligation to keep it. So we must count them as instances of *ostensibly non-teleological* component obligations. The urgency of a purely teleological obligation varies directly with the amount of good which will be produced or of evil which will be averted. Thus, in theory, there is a general principle for comparing teleological component obligations with each other. But there seems to be no general principle for comparing the urgency of two ostensibly non-teleological obligations, e.g. truth-telling and promise-keeping, with each other. And there seems to be no general principle for comparing the urgency of a teleological obligation (e.g. not to give needless pain to a person) with an ostensibly non-teleological one (e.g. truth-telling). I think that it is this circumstance which make utilitarianism so attractive to many people. According to it, the ostensibly non-teleological obligations are all ultimately derived from teleological obligations, and so they can be weighed against each other and against admittedly teleological obligations in accordance with a single principle.

1.3323. Most claim-fulfilling acts. A person's resultant obligation in any situation is to make as good a compromise as possible between the various claims which the situation imposes on him. So we come to the important notion of *a most claim-fulfilling act*. We must not talk rashly of *the* most claim-fulfilling act in a given situation. For, granted that all the claims on a person cannot be completely satisfied by him, there might be two or more different compromises which were equally satisfactory and were all more satisfactory than any other which was open to him. Then each of them would be *a* most claim-fulfilling act; but none of them would be *the* most claim-fulfilling act.

In dealing with individual claims and component obligations we saw that eight cases are possible when we allow for factual and ethical ignorance and error. Eventually we chose three of these as being of outstanding importance. In dealing with resultant obligations we see that there are two additional possibilities of error. (i) The agent may make ethical mistakes about the relative urgency of various component obligations. These are in fact much the commonest of ethical mistakes. He may thus be led to think that by doing act A he will fulfil all the claims on him as completely as he can. Yet really act

B, which he knows to be also within his power, might make a better compromise. (ii) He may make factual mistakes about the limits of his own powers. He may think that *B* is out of his power, when really it is not.

Let us start from the completely objective end. The completely objective action is the following: An act which would *in fact* bring about a certain change, which *would* satisfy as fully as any other change *actually* producible by the agent all the claims which the situation *as it really is* does *in fact* impose on him. Such an act might be called a "*materially* most claim-fulfilling act". If we allow for the possibility of error or ignorance at each of the five points which I have underlined, it is evident that 32 alternatives are possible. It would be tedious to enumerate them all, so I will confine myself to the three which are of outstanding importance. At the other extreme we come to the most completely subjective possibility. This is as follows: An act which the agent *thinks* would bring about a certain change, which change he *thinks* would satisfy as fully as any other that he *thinks* he can produce all the claims which he *thinks* that the situation, *as it appears* to him, imposes upon him. This may be called a "*subjectively* most claim-fulfilling act". The only intermediate case that seems to be important is that which allows for every possibility of factual error and ignorance but not for ethical error and ignorance. This may be defined as follows: An act which the agent *thinks* would bring about a certain change, which change *really would* satisfy as fully as any other change which he *thinks* he can produce, all the claims which the situation *really would* impose on him if it were as he *believes* it to be. This may be called a "*formally* most claim-fulfilling act". If a person does an act of the first kind, he discharges his *material resultant obligation*; if he does an act of the second kind, he discharges his *subjective resultant obligation*; and if he does an act of the third kind, he discharges his *formal resultant obligation*.

1.33231. Acts "open to" an agent. In talking of "most claim-fulfilling act" in various senses, I have always added the qualification "*open to* a given agent in a given situation". Before going further we must try to clear up this notion of the acts "*open to* an agent". We can go a long way here without touching on the question of free-will versus determinism.

In the first place, it is evident that some such limitation is needed if the notion of a most claim-fulfilling act is to have any meaning. Whenever we can significantly talk of a minimum or a maximum, e.g. the *shortest* distance between two places, the *greatest* area that can be enclosed by a curve, and so on, certain limiting conditions are expressed or implied. In the case of the shortest distance we should want to know whether the path was to be confined to the earth's surface or whether one was supposed to be capable of burrowing through the earth or flying through the air. And, even when this was settled, we should need to know whether the geometry of the space con-

taining the two points was supposed to be Euclidean or non-Euclidean, and so on. Similarly, there is no clear meaning in the question "What is the most claim-fulfilling act in a given situation?" unless it be asked with reference to an agent whose powers are definitely limited. In certain situations the most claim-fulfilling act for a boxing-blue in full training might be quite different from that for an elderly solicitor with a weak heart. And the most claim-fulfilling act for an angel or a magician might be one which no ordinary human being could accomplish.

When a given agent is in a certain situation and is just about to act he has a certain stock of cognitive and conative powers and dispositions, innate or acquired, and these delimit the acts which are open to him. There will be a certain set of alternative possible acts such that each of them would happen if the agent willed it and would not happen unless he willed it. Now these possible acts may be divided into three classes, viz. (i) those which the agent actually thought of, (ii) those which he did not think of, but which he would have thought of if he had taken enough time and trouble in reflecting, and (iii) those which he would not have thought of even if he had tried as hard as he could. It seems to me that the third class may be excluded for the present purpose. It is true that each member of it would happen if and only if the agent willed it. But, since it is not in his power to think of any of them, it is not in his power to will any of them. So the acts open to a given agent in a given situation are those possible acts which (a) he did think of or would have thought of if he had tried hard enough, and (b) each of which would happen if and only if he willed it.

There is one further point to be noticed here. There might be a certain conceivable act x, which the agent either did think of or would have thought of if he had tried hard enough. And this may be one which would happen if and only if the agent willed it. But he might underestimate his own powers and mistakenly believe that it woud not happen even if he were to will it. Or he might mistakenly believe that it would happen anyhow even if he did not will it. It seems to me that such an act must be counted among those which are open to the agent, although he mistakenly believed that it was not open to him. Still it must be admitted that if a person believes that a certain act is not open to him, in the sense defined above, this opinion does exclude it from the class of alternatives which he seriously considers when he makes his decision. The possibility of mistakes of this kind is allowed for in the notion of a *formally* and a *subjectively* most claim-fulfilling act.

It is evident that a person cannot do a materially most claim-fulfilling act, except by chance, unless all the following conditions are fulfilled. (i) He has completely adequate and correct *factual* information about (a) the nature of the situation, (b) the alternatives which are open to him, and (c) the effects which each alternative action would have. (ii) He has completely adequate

and correct *ethical* information about (a) the claims imposed upon him by the various factors in the situation, (b) the relative urgency of these claims, and (c) what kind of change would satisfy them all as fully as any other change which he could produce. When I say that, unless these conditions are fulfilled, he cannot do a materially most claim-fulfilling act except by chance, what I mean is this. If he sets himself to do such an act and succeeds, his success will be due to some happy combination of circumstances, outside his knowledge and control, by which his various errors and ignorances cancel out.

A person could do a *formally* most claim-fulfilling act without any help from chance, even if all the factual conditions broke down, provided that the conditions of ethical knowledge were fulfilled. What is needed is this. Completely adequate and correct *ethical* information about (a) the claims which would be imposed on him by the various factors which he believes to be present in the situation, (b) the relative urgency of these claims, and (c) what kind of change would satisfy them all as fully as any other change which he thinks he can produce. If these conditions are not fulfilled a person who sets himself to do a materially most claim-fulfilling act may succeed in doing a formally most claim-fulfilling act; but his success will be due to his various ethical ignorances and errors cancelling out.

The only kind of act which a person could be always sure of doing if he chooses, however ignorant and stupid and crazy he may be both factually and ethically, is a subjectively most claim-fulfilling act. For, in order to do this, he has only to set himself to do what he believes to be a materially most claim-fulfilling act and to keep himself set in that direction. Now this kind of resultant obligation can be fully discharged by a person who is grossly ignorant and stupid or crazy in his judgments on both factual and ethical matters. It is therefore not surprising that an act which is "right", in this sense, may inflict the most dreadful wrongs, or that a person who habitually acts "rightly", in this sense, may be a private nuisance or a public calamity.

1.33232. Problems of maximisation. There is not very much that can be said about how to discover what is a most claim-fulfilling act in a given situation. All that we can do is to point out some of the peculiar difficulties and some of the relevant considerations. (1) As I have already said, the various claims which may conflict with each other are *prima facie* extremely heterogeneous. There are first the teleological ones to be weighed against the various ostensibly non-teleological ones. Then there are various ostensibly non-teleological ones to be weighed against each other. And the teleological ones may be qualified by ostensibly non-teleological ones about distribution. I do not think that any general rules can be given, except very vague ones. I suspect that skill and insight in this kind of weighing and balancing of claims

is gradually gained by practice; that some people can acquire it much more readily and fully than others; and that those who have it can do very little to impart it to others by precept. It seems to me to be more like playing some fairly difficult game, in which skill and chance both a play a large part, than like any operation which can be reduced to rules and learnt from a book. (2) One great difficulty is that it is a problem of maximisation in at least two heterogeneous dimensions. In general there are *several* claims and they are of *various* degrees of urgency. One wants to satisfy as many of the claims as possible; but one also wants to satisfy the *more urgent* ones *more completely* than the less urgent ones, if one cannot satisfy them all fully. This suggests that there may often be two kinds of most claim-fulfilling act in a given situation. One fulfils all or most of the claims to some extent, but does not satisfy even the most urgent of them very fully. The other satisfies the most urgent claims fully or nearly so, but leaves many of the less urgent ones completely unfulfilled. I gave an example some time ago of a don who has claims on him to do original work in his subject, to give lectures, to supervise pupils, to take part in the administration of his college and the university, and perhaps to take some part in local or national politics. The two extremes are to specialise on the one activity which one is best at, and almost wholly to neglect the others, at the risk of becoming lopsided and in many ways parasitic on others. The other extreme is to plunge into all these activities, at the risk of doing none of them particularly well, and letting one's special talents run to seed. (3) In trying to find out what is the most claim-fulfilling act in a given situation and at a given moment, a person must not confine his attention to that situation and that moment. Suppose he acts with the intention of producing a certain change in the immediate future. The effects of his actions are most unlikely to be confined to the immediate future. It may cause the situations in which he will be placed for years ahead to be quite different from what they would otherwise have been. These situations, if they arise, may impose claims on him to which he would not otherwise have been subject. And the same act which causes these claim-imposing situations to arise may affect the agent's power to fulfil the claims which they would impose. E.g. suppose that a person, in fulfilment of the claim of gratitude to a benefactor, consents to act as trustee under his will. In the remote future this may involve him in the most complicated legal business and may even expose him to financial loss or ruin if his co-trustee should default. These remote consequences may make it impossible for him in the future to fulfil the claims of his employers on his time and energy, and the claims of his wife and family for support and education.

I think it may be well to put this into symbolic form. Let us denote the state of affairs which exists around the agent at the time when he makes up his mind either to do nothing and let things slide or to do a certain act at once by

F_1. If he decides to do nothing and let the situation develop without interference we shall get a series of successive phases which we can denote by F_1, $F_2 \ldots$, $F_n \ldots$. Each phase in the series will be highly complex. The factors of a typical phase F_n may be denoted by the symbols $f_{n1}, f_{n2} \ldots, f_{nm} \ldots$. These will be interrelated in a characteristic way which we may denote by the formula

$$F_n = R_n (f_{n1}, f_{n2} \ldots, f_{nm} \ldots).$$

I shall denote the series $F_1, F_2 \ldots, F_n \ldots$ by S, and shall call it "the unmodified course of events". Now suppose, instead, that the agent decides to perform a certain act x. *The initial phase* F_1 will contain, as before the factors $f_{11}, f_{12} \ldots$, $f_{1m} \ldots$. But to these will now be added the act x as an additional cause-factor. So the subsequent phases will be different from what they would be in the unmodified series. We will describe the series as modified by x by the symbol $\overset{F}{x}_1 \} F_2^x \ldots, F_n^x \ldots$. Consider a typical subsequent phase such as F_n^x. Here the elements and their relations will in general be modified, though of course some of them may happen to be the same as they would have been in the unmodified series. We may represent this by writing

$$F_n^x = R_n^x (f^x{}_{n1}, f^x{}_{n2} \ldots, f^x{}_{nm} \ldots).$$

I shall denote the series $\overset{F}{x}_1 \} F_2^x \ldots, F_n^x \ldots$ by S^x. Suppose the agent had done act y instead of act x. Then we should have had the modified series S^y, i.e., the series $\overset{F}{y}_1 \} F_2^y \ldots, F_n^y \ldots$, when $F_n^y = R_n^y (f^y{}_{n1}, f^y{}_{n2} \ldots, f^y{}_{nm})$.
The position then is this. In considering what, if anything, he ought to do now in the initial phase F_1 the agent must take into account the following points. (1) The various factors $f_{11} f_{12} \ldots$etc., in the initial situation and the claims that they now impose on him. (2) The situations which will arise in the future according to whether he does nothing, or does x or does y, i.e. such alternative possibilities as F_n, F_n^x, and F_n^y. He will have to consider the various factors which there would be in each of these alternative possible future developments, and the claims which each would impose on him if it became actual. (3) The effects of doing nothing, or doing x, or doing y on his own future powers and resources. E.g. if he does x, will the change which this will make on his own powers and resources in the remoter future be such that he will be incapacitated for satisfying the claims which the later terms, such as F_n^x, of the modified series S^x, will impose upon him? Thus the problem can be extremely complex. He has to consider the various alternative series of successive external situations which diverge from the present situation according to what he does or leaves undone now. He has to consider the corresponding alternative series of states of himself which diverge from his present state according to what he does or leaves undone now. He must consider the effects

of any attempt to satisfy present actual claims both on future claims and on his powers of satisfying them. And in so doing he must take account of the various contemporary factors which are contained in the initial situation and in each of the later situations which would arise according to which of the alternative possible series he initiates.

1.3324. Notion of an optimific act. There is another maximal notion which we can now consider, viz. the notion of an *optimific act* in a given situation. This is, roughly, an act which will produce as much good or as little evil as any other act open to the agent at the time. The notion is important for two reasons. (i) Even if we have non-teleological obligations, and even if our teleological obligations are limited by various non-teleological distributive obligations, everyone admits that the purely teleological obligation to produce as much good and as little evil as possible exists and may be very urgent. Hence the notion of an optimific act is bound to be important because of its connexion with this important part of our obligations. (ii) If utilitarianism were true, all our obligations reduce in the end to the one unlimited teleological obligation to produce as much good and as little evil as possible. Therefore, on that theory, a most claim-fulfilling act will necessarily be identical with an optimific act. And so the notion of optimific act will be fundamentally important for utilitarians.

1.33241. Utility. We will now define this notion and those which are connected with it. We begin with the notions of *utility* and *disutility*. An act has utility if (a) it produces goods which would not otherwise have existed, or (b) it increases the goodness of goods which would have existed apart from it, or (c) it prevents or reduces a diminution of existing goods which would otherwise have taken place, or (d) it prevents the occurrence of evils which would otherwise have come into existence, or (e) it diminishes the badness of evils which would have existed apart from it, or (f) it prevents or reduces a growth of existing evils which would otherwise have taken place. *Disutility* can be defined by substituting "badness" for "goodness", "evils" for "goods" and vice versa throughout the above definition.

Let us suppose, as before, that a person is in the initial situation F_1 and that he can either let things slide or do act x or do act y. According to which alternative he chooses one or other of the following three alternative series will be actualised.

$$F_1, F_2 \ldots, F_n \ldots \qquad \text{i.e. } S$$

or $$\left.\begin{matrix} F_{x1} \end{matrix}\right\} F_2^x \ldots, F_n^x \ldots \qquad \text{i.e. } S^x$$

or $$\left.\begin{matrix} F_{y1} \end{matrix}\right\} F_2^y \ldots, F_n^y \ldots \qquad \text{i.e. } S^y$$

Now consider a triad of corresponding phases in the three series, e.g. F_n, F_n^x and F_n^y. And then consider another triad at a different position, e.g. F_p, F_p^x and F_p^y. It might be that F_n^x contains a greater balance of good or a less balance of evil than F_n, and a less balance of good or greater balance of evil than F_n^y. We can write this as $F_n^x > F_n$ and $F_n^x < F_n^y$. Now it might also be the case that $F_p^y < F_p^x$ and also $F_p^y < F_p$.

In considering the relative utility of inaction, of x, and of y, we shall have to balance the goods and evils of all the various phases of each alternative series S, S^x, and S^y. E.g. to substitute an efficient dictatorship for a corrupt and inefficient system of representative government, as Signor Mussolini did in Italy, has a fairly obvious nett utility so long as you confine your attention to the more immediate phases of the resulting course of events. But one has also to consider the remoter phases, e.g. those which will ensue when the dictator dies naturally or gets fossilised or assassinated and a successor has to be appointed. It is then not so clear that this act has greater nett utility on the whole than putting up with the muddle and twaddle of representative institutions.

Again, in each phase in the series of events which would ensue on the doing of an action, there will in general be some good features, some bad ones, and some indifferent ones. If a different act were done, the corresponding phase in the series of events which would *then* ensue would often be better in some respects and worse in others. Thus to estimate utility or disutility of a proposed action the various goods and evils at each phase of the series which it would initiate will have to be balanced against each other. There are various distinctions which are important in connexion with utility, and I shall now proceed to draw them.

1.332411. Normal and individual utility. We must distinguish between the two following propositions. (i) Most acts of the kind K would have great utility (or great disutility). (ii) This particular act, which is of the kind K, done in this particular situation, would have great utility (or great disutility). In an ordinary peaceful society most acts of promise-keeping have utility and most acts of promise-breaking have disutility. But it is quite possible that in a particular situation of a peculiar kind an act of promise-breaking might have great utility and an act of promise-keeping might have great disutility. So we must distinguish between the *normal* utility or disutility of certain *kinds* of

act, and the *individual* utility of a particular act of a given kind performed in a particular situation. As to the connexion between the two all that can be said is this. If acts of the kind *K* have great normal utility, there is a *prima facie* case against supposing that a particular act of this kind performed in a particular situation will have great individual disutility; and vice versa. So it will always be sensible to require strong and definite reasons for doing an act of a kind which has great normal disutility and against doing an act of a kind which has great normal utility, if one's intention is to do an optimific act.

1.332412. Collective and singular utility. It may be that few, if any, acts of a given kind, taken one by one, would have very great utility or disutility. But the conjunction of a great many such acts, either simultaneously or in close succession, might have great utility or disutility. It might even be the case that acts of a certain kind would have great utility if they were infrequent and great disutility if they were frequent; or conversely. Most acts of slightly understating one's income in making an income tax return have very little disutility when taken singly. But the concurrence of a great many such acts in any one financial year may have great disutility. Again, deliberately abstaining from having children may have great utility, so long as only a minority of people practise it and provided that they are suitably distributed among the population. But it might have great collective disutility if a majority of people practise it; or if, as is in fact the case, those who do are mainly confined to the most efficient and intelligent sections of the community. So we must distinguish between the *singular* utility or disutility of acts of a given kind, taken one by one; and the *collective* utility or disutility of a combination of many such acts within a restricted region of time or space.

1.332413. Primary and secondary utility. All the cases that we have so far considered may be called instances of *primary* utility or disutility. By this I mean that the good or bad effects of the actions do not depend upon men's beliefs about them but are direct consequences of their own nature. But in human affairs beliefs, particularly if they are widely held, may set up processes which cause them to become true or to become false. A very obvious case is this. Suppose that an originally quite baseless rumour arose that a certain bank was in difficulties. If this was believed by enough of the depositors, a large proportion of them would simultaneously try to withdraw their money. And then the bank actually would be in difficulties. The application to our present problem is as follows. Suppose that a certain act would have considerable primary *utility* in a certain particular situation. Suppose that acts of this kind have great *normal disutility* or great *collective disutility*, i.e. either that in most situations such an act would produce a great deal of evil, or that the concurrent performance of many such acts would produce a great deal of

evil. Suppose finally that most people are strongly inclined to do acts of this kind, e.g. because they satisfy a strong and widespread impulse or because they give a great deal of immediate pleasure to the agent. Then, although the doing of such an act by a certain person in a certain situation might have considerable primary utility, it might have so much secondary disutility as to outweigh its primary utility. This can happen in the following way. Suppose that the act were done by a prominent and respected person, and that it were widely known that he had done it. Then it is extremely likely that many other people would make this an excuse for doing acts of this kind in situations where the very special circumstances which gave this particular act its primary utility were lacking. Now by hypothesis in most situations the doing of such an act would have primary disutility; or, again, the concurrent performance of such acts by many people would have primary disutility. Therefore the primary utility which the act would have, if done by this agent in this situation, may be altogether outweighed by the secondary disutility which would arise in this way.

There are two important special cases which may be treated under this heading. (i) In some kinds of situation which are frequently occurring to many people it is of very great utility that there should be a rule *of some sort* and that it shall be rigidly kept; but following one rule has in itself no more utility than following another rule. An obvious example is the rule of the road. It is of the utmost utility that everyone in the same country should follow one and the same rule about overtaking and passing traffic. But it is a matter of complete indifference whether the English left-hand rule or the Continental right-hand rule is adopted. Once a rule has been set up, any breach of it has great disutility; not because of any special merit in that particular rule, but simply because it is a breach of *the* rule which is commonly accepted by one's neighbours. (ii) The open breach of any law of the land or any widely accepted moral rule tends to have considerable secondary disutility even when the law is absurd or the moral rule is groundless. For the existence of a general system of law and morality, which most people obey, even when it goes against the grain, is of enormous utility; since it is a necessary condition of nearly every good that an individual can hope to attain in this life. Now such a system will always contain many anomalies; it will press hardly on some people all the time and on most people at some time. It can be maintained in the face of these disruptive influences only if it is regarded by most people who are affected by it with sentiments of quasi-religious awe which it would be impossible to justify to a sceptical stranger. And once these sentiments are destroyed it is extremely difficult to build up an organised society again or to find any principle of cohesion except naked brute force and terror to put in their place. Now any open breach of a law of the land or a widely accepted moral rule, particularly if it is likely to be

imitated by many, tends to bring the whole system of law or morality into contempt. Therefore at certain critical periods even one such act may have very great secondary disutility, and a conjuction of a number of them may be disastrous. The history of any South American republic and of Germany under the Weimar Constitution is full of instances of this fact. On the other hand, if bad laws were never publicly broken and absurd moral rules never publicly flouted, a society would either petrify or putrify. Thus the utility or disutility of such acts depends largely on the general stability of the social systems at the time when they are done. In a very stable society, such as that which existed in England between 1860 and 1880, they may have considerable utility. In a very unstable society, such as that of Germany between 1918 and 1935, they may have considerable disutility.

1.332414. Average-changing utility and distributive utility. Suppose that several people will be affected by an action, and that the action will produce some good states and some bad states in the persons whom it affects. The easiest example to think of is where the good states are pleasant experiences and the bad ones are unpleasant experiences; but I shall not assume that pleasantness is the only good-making characteristic and unpleasantness the only bad-making characteristic of experiences. Now good and bad experiences may be distributed in various alternative ways among the same people. Suppose it is possible to talk, as utilitarians do, of the net amount of pleasure in the experiences of a group of persons. Then we could imagine the same net amount of pleasure being produced either (a) by giving all the pleasure to A and all the displeasure to B, or (b) by exactly the opposite distribution, or (c) by many intermediate kinds of distribution in each of which both A and B were given some pleasant experiences and some unpleasant ones. Now it would generally be held that a given net amount of distributable goods and evils, distributed in one way among a certain group of persons, would constitute a better *total state of affairs* than the same net amount distributed in a different way among the same group of people.

A rather similar point arises even when an act affects only one individual. The same amount of pleasant and unpleasant experiences might be distributed throughout his life in such a way that all the pleasant experiences came earlier and all the unpleasant ones later, or in exactly the opposite order, or in many intermediate ways. Most people would hold that the life of this person, taken as a whole, would have a very different value according to the way in which this given amount of pleasure or displeasure was distributed throughout it.

It is evident that we must distinguish between the net value of the successive experiences of an individual and the value of the total course of experience which these together make up. The latter will depend jointly on the

former and on the distribution of those experiences in time. Similarly we must distinguish between the net value of the contemporary experiences of a number of individuals who form a group and the value of that collective state of affairs which these experiences together make up. The latter will depend jointly on the former and on their distribution among the individuals of the group. (Of course this collective state of affairs, made up of the contemporary experiences of the members of a group, is *not* itself an experience. It is desirable to say this firmly, because people so often start by using metaphorical language which suggest that it is, and end by taking their own metaphors literally.)

Now consider two alternative acts A and B, each of which will affect all the members of a certain group G. It might be that A would produce a greater balance of good over bad experiences in the members of this group than B would do. But it might also be that B would cause a better distribution of the good and bad experiences which it produces than A would do. I should express this by saying that A has greater *average-changing* utility than B, but B has greater *distributive* utility than A. If we are going to compare the utility of A with that of B, we must take into account both these kinds of utility and disutility. What we may call the *totalising utility* of such an act is a function of its average-changing and its distributive utility. A will have greater totalising utility than B if the collective state of affairs which it produces in the group of persons affected is better, when account is taken both of the net balance of good and bad experiences in the various individuals and the distribution of these goods and evils among the individuals.

If an act is a factor in producing some good state, e.g. a pleasant experience, it will also be a factor in producing it in some definite person. It will therefore also be a factor in determining the way in which this distributable good will be distributed. So an act would hardly have average-changing utility or disutility without having some degree of distributive utility or disutility. But the converse is not true. An act may be a factor in determining the way in which a certain good or evil shall be distributed without being a factor in producing it. Suppose, e.g., that X has decided to give a treat to some poor child or other, and that Y brings to his notice a particular poor child Z. Then Y's act is a factor in determining that a pleasant experience, which X will in any case produce in *some* poor child, will be produced in the child Z. This act may have distributive utility or disutility, according to whether Z is more or less needy or deserving than certain other candidates. But it has no average-changing utility; that belongs to X's act.

1.332415. Distribution of goods and evils and distribution of means. It is important to notice that what I have been talking about is the distribution of good and bad states, e.g. pleasant or painful experiences, among persons. To

give an experience to a person means to cause that experience to occur in that person's mind. This must be carefully distinguished from the distribution among persons of *material or economic means* to good and bad states, e.g. money or alcohol. To give money or alcohol to a person is to make him the legal owner of that money or that alcohol and enable him to do what he likes with it. An equal distribution of alcohol might produce a very unequal distribution of good and bad experiences, since people's tastes and capacities in that connexion are very different.

Now any distribution of money or material objects among a group of persons will automatically carry with it a certain distribution of these good and bad experiences which depend on the possession or the lack of that money or of those material objects. We must not assume, however, that the distribution of means which would bring about the best distribution of good and bad experiences in the recipients is necessarily the one which would have the best consequences on the whole. For the ownership and transference of property is a typical example of those departments of life in which it is of the utmost utility to have rules which are well-known, readily applicable, and rigidly enforced. And, if property is distributed by transference in accordance with such rules, it will often happen that it is allocated in a way which does not bring about the best possible distribution of good and bad experiences in the persons immediately concerned.

Suppose, e.g., that I were in a situation in which I had to choose between repaying a sum of money which I had borrowed from an undeserving rich man *A* or giving it to a deserving poor man *B*. Suppose I repay *A* and am then unable to give anything to *B*. The undeserving *A*'s happiness will be increased to a trivial extent, and the deserving *B* will continue to endure the evil experiences which arise from poverty. Suppose, on the other hand, that I evade paying *A* and give the money thus owed to *B*. The happiness of the undeserving *A* will be diminished to a trivial extent, and the deserving *B* will derive considerable happiness from the relief of his urgent needs.

If, then, we confine our attention to the experiences likely to be produced in the two alternative recipients of the money, there is little doubt that to give the money to *B* has greater totalising utility than to repay it to *A*. For it has greater *average-changing* utility, since *B*'s happiness is greatly increased and *A*'s only slightly diminished by it. And it has greater *distributive* utility, since it is a better state of affairs, from the standpoint of distribution, when deserving persons enjoy happiness and undeserving ones do not than when the opposite allocation exists.

Nevertheless, when we take into account normal, collective, and secondary utility, the balance is almost certainly reversed. It is of the utmost utility that there shall be simple and easily applicable rules about the distribution and transference of money, and that everyone can rely on their being enforced. It

is of the utmost utility that persons who need money and can use it profitably shall be able to borrow it from those who have more than they can use themselves. Now, human nature being what it is, those who have money will not lend it unless they can count on being repaid regardless of their own demerits and of the merits and the needs of other persons. Such questions as whether *A* did or did not lend money to *X*, how much he lent, and under what conditions, are in most cases capable of being answered in a way which will satisfy all impartial persons. It is not difficult to set up tribunals which will in general investigate such questions with reasonable competence and objectivity. But such questions as whether *B* is more deserving than *A*, and, if so, how much more; of how much additional happiness is deserved by so much additional virtues; and so on; are incapable of being answered objectively. No one has the relevant data, and there is no agreement about the relevant principles.

There is a further complication about the distribution of money and other means to good and bad experiences. So far I have considered only the *teleological* obligation to produce as much good and as little evil as possible. I have tried to show that, when all the remote and collateral consequences of the alternatives are taken into account, one is probably nearly always under a *teleological* obligation to pay a debt to an undeserving rich man rather than give the money to a deserving poor man. All this would have to be taken into account by a Utilitarian. But many people would say that there is no need to appeal to these remote and collateral good and bad consequences. They would say that, even if it were certain that the consequences of giving the money to the deserving poor man *B* would be on the whole better than those of repaying it to the undeserving rich man *A*, it would still be one's duty to repay *A* and not give the money to *B*. They would say that the money is owed *to A* and not to *B*. It was lent to me by *A* on promise of repayment, or I have taken goods or services from *A* on the understanding that I will pay this sum of money for them at a certain date in the future. The common-sense non-utilitarian view is that these relations between me and my past acts, on the one hand, and *A* and his past acts, on the other, suffice to give him a moral claim on me to repay the money to him. The question of the goodness or badness of the consequences is irrelevant to this claim. Now this amounts to saying that the mere existence of the debtor-creditor relationship gives rise to a *non-teleological* obligation on the debtor to pay back what he has borrowed from the creditor and not to use it for other purpose. If repaying the debt happens also to fulfil the teleological obligation to produce as much good and as little evil on the whole as possible, so much the better. But according to non-utilitarians the obligation to repay the debt has an independent basis in the mere existence of the debtor-creditor relationship.

We may sum this up as follows. Whenever we have to distribute money or other means to good or bad experiences we must distinguish between the

distribution of the means and the consequent distribution of the good and bad experiences. Our teleological obligation is concerned directly with the latter; it is concerned with the former only indirectly as a means to the latter. Now one may be in a situation in which one seems to be under a *direct non-teleological* obligation to distribute money or other property in a certain way and an *indirect teleological* obligation to distribute it in a different way. In many cases it is found that this conflict disappears if we take into account the remote and collateral good and bad consequences of the alternative ways of distributing the property. But one cannot be sure that it would disappear in every case. Even when there is no conflict, common sense holds strongly that there are two distinct and independent obligations. One is direct and non-teleological and is based upon some special relationships like that of debtor and creditor. The other is teleological and indirect, and it is based upon the fact that this distribution of means will bring about the best consequences in the long run and on the whole. The utilitarian, of course, denies that there is any *non-teleological* obligation.

1.33242. Definition of an optimific act. We can now define an "optimific act". *X* is an optimific act if it has at least as great totalising utility as any act open to the agent in the situation. This means that, when account is taken of (a) the nett amount of good produced or evil averted by it in the lives of the various persons affected, and (b) the good or bad way in which these goods and evils are distributed among those persons in consequence of it, the total result is no worse than that which would have followed from any other act open to the agent. In making this estimate we must take account, not only of the direct and primary utility and distutility of this act and the alternatives to it. We must also take into account any secondary or indirect utility or disutility which it, or the alternatives to it, might derive from being known about and perhaps widely imitated. It must be noted that the total result of an optimific act need not be positively good. There may be situations in which the consequences of every alternative open to the agent, including that of inaction, will be more bad than good. Even if this is not so, it may be that the situation after a given event will be much worse than it was before, no matter what alternative act the agent may do. It may be that the only question is whether any act, and if so, what act, will most diminish the damage. I think it likely that the English Cabinet was faced with such a situation in August 1914, and almost certain that they were in September 1939. Now many people find it impossible to believe that the consequences of an optimific act may contain a balance of evil. And they find it impossible to face the fact that, in spite of an optimific act being done in a certain situation, the total state of affairs may be worse afterwards than it was before. So they try to persuade themselves that the subsequent state of affairs will contain some great

positive good, and will at any rate be better than the previous state of affairs. Much of the disillusionment which followed the war of 1914 – 18 sprang from this fallacy, and we are now seeing it repeated.

Now this mistake leads to another. Suppose that a certain act X was done under the belief that it was optimific. Suppose that the subsequent state of affairs is found to be more bad than good, or at any rate to be worse than the previous state of affairs. Really this does not show that X was not an optimific act. But anyone who thinks that the sequel to an optimific act must be prositively good, or at any rate better than the previous state of affairs, will conclude that X was not optimific. He will conclude that some other alternative, such as Y or inaction, would have had better consequences. This may happen to be true, but the argument is completely fallacious. And it is a very common fallacy.

1.33243. The three associated notions. The notion of an optimific act, as I have just defined it, is purely objective, like the notion of a *materially* most claim-fulfilling act. A person who intended to do an optimific act in a given situation could be sure of doing so if and only if the following two conditions were fulfilled. (i) If he had complete and correct *factual* information about (a) the nature of the situation, (b) the acts which are open to him, and (c) the consequences which each of these alternative acts would have throughout the whole of future time. (ii) If he had complete and correct *ethical* information about the relative values of the various alternative sets of consequences which would follow from the various alternative acts open to him. It is obvious that no one is ever in this position. Therefore if anyone who intends to do an optimific act ever succeeds, his success is due partly to luck. Let us call this a *materially* optimific act.

We could define a *formally* optimific act as follows. It is an act such that the consequences which the agent *thinks* it would have *really would* be no worse than the consequences which he *thinks* would follow from any other act which he *thinks* is open to him. An agent who intended to do a materially optimific act could count on doing a formally optimific act, however ignorant or mistaken he might be factually, provided he made no mistakes in his judgments of value. But if he were ignorant or mistaken about values, he would succeed in doing a formally optimific act only by luck.

Lastly, we can define a *subjectively* optimific act as follows. It is an act such that the consequences which the agent *thinks* it would have *appear to him* to be no worse than those which he *thinks* would follow from any other act which he *thinks* is open to him. Anyone, however ignorant or crazy about facts and about values, can be sure of doing a subjectively optimific act, provided he sets himself to doing a materially optimific act and persists in his intention.

The notion of *utility* has no reference either to intention or motive, but simply to good or bad consequences. And it is not concerned with whether the consequences are states of the agent himself or of others. In so far as an act will in fact produce good states or prevent or improve bad states in persons *other than* the agent, it may be called a *benefic* act. If it not only has these effects but is also done by the agent with the intention of producing them, it may be called a *beneficent* act. If you substitute "bad" for "good", "good" for "bad", and "worsen" for "improve" in these definitions, you get the definitions of a *malefic* act and a *maleficent* act. These must be distinguished from *benevolent* and *malevolent* acts. The latter introduce the question of the agents' motive, which we are at present ignoring.

1.33244. Utility and claim-fulfilment. It will now be worthwhile to compare and contrast the two notions of utility and claim-fulfilment. An act has utility if it produces results which are on the whole *better* or *less bad* than the results of inaction would have been. An act is claim-fulfilling if it produces a certain change in a certain person's condition, that change being the one to which he has a *moral claim* because of a certain relationship in which he stands to the agent. The following points emerge.
(1) Each notion involves a causal factor, viz. the notion of an effect to be produced by the act. (2) Each involves a factor which is non-causal and which is quite different from any that occurs in any of the positive sciences. In the case of utility this is the evaluatory predicate *good* or *better*. In the case of claim-fulfilment it is the obligatory predicate *having a moral right to*. (3) The latter is essentially a moral notion; the former is not. An act has utility if it is a factor in producing or conserving or increasing any kind of value, e.g. pleasant sensations, states of aesthetic appreciation, states of virtuous volition or rightly directed emotion, and so on. Now pleasant sensations and states of aesthetic appreciation are not, as such, *morally* valuable; though states of virtuous volition or rightly directed emotion are. (4) Having a moral right always depends on some pre-existing relationship between the person who has the right and the person on whom it gives him a claim. But the goodness or badness of a person's experiences, and therefore the utility or disutility of the act which produces them, may be quite independent of any previous relationships between this person and the agent. This is certainly true of many of those goods and evils which consist in pleasant and painful experiences. (5) Although utility itself is not a specifically moral notion, it is of course connected with the specifically moral notions of rights and obligations in the following way. The most elementary moral right which a person has in his dealings with another is not to be unnecessarily hurt or thwarted. This right depends on no special relationship but simply on the fact that he is a sentient and conative being. The corresponding component obligation of

beneficence may, of course, be limited and even over-ridden in any particular case by other component obligations which arise from more special relationships. But it is always there in the background.

1.3325. Notion of an optimising act. There is one more maximal notion which is worth considering beside those of a most claim-fulfilling act and an optimific act. It is what I call the notion of an *optimising act*. Suppose that an agent in a given situation has three alternative acts open to him, viz. x, y, and z. According to which of them he does, we shall get one or other of the alternative series

$$\overset{F_1}{x}\Big\} F_2^x, F_3^x, \ldots \qquad \text{i.e. } S^x$$

or $$\overset{F_1}{y}\Big\} F_2^y, F_3^y \ldots \qquad \text{i.e. } S^y$$

or $$\overset{F_1}{z}\Big\} F_2^z, F_3^z \ldots \qquad \text{i.e. } S^z$$

Now in dealing with optimific acts we took no account of the first terms of such series, since we were concerned only with the *consequences* of actions. We considered only the goodness and badness of the residual series, which begin respectively with the terms F_2^x, F_2^y, and F_2^z. We can call these residual series R^x, R^y, and R^z respectively. x would be an *optimific* act if R^x were at least as good, on the whole, as R^y and as R^z. The notion of an *optimising* act arises when we leave out this restriction and compare the three complete series S^x, S^y, and S^z in respect of goodness and badness. x would be an *optimising* act if S^x were at least as good, on the whole, as S^y and as S^z.

It is evident that the notions of an optimising act and of an optimific act are different. It is also theroetically possible that in a given situation the act which would be optimific and the act which would be optimising would be different acts. Suppose, e.g., that R^x were better on the whole than R^y and than R^z, whilst S^y was better on the whole than S^x and than S^z. Then x would be the optimific act and y would be the optimising act in the situation F_1. Whether this theoretical possibility could ever be realised in practice would depend on the following questions. Do acts themselves have any value or disvalue? Even if they do not, does the whole composed of an initial situation F_1 and an act x ever differ in value from a whole composed of the same situation F_1 and a different act y? It is quite possible that the latter might be the case even if x and y had neither value nor disvalue taken by themselves. For x might harmonise in some way with F_1 whilst y might disharmonise with it. Cf., e.g., three sounds a, b and c. b and c might each be neither pleasant nor unpleasant, but the combination ab might be highly pleasant and the combination ac highly unpleasant. It is only if acts have no values or disvalues in

themselves, and no special harmony or disharmony with the initial situations in which they are done, that we could be sure that an optimising act never differs from an optimific act.

If we allow these possibilities, a person who wants to produce as much good or as little evil as he can in a given situation cannot confine himself to the question "Which of the acts open to me will have the best or the least bad *consequences*?" For any act that he may do may contribute to the amount of good or evil in the world in two quite different ways. (i) It may make an *immediate* contribution, either by its own intrinsic goodness or badness, or by harmonising or disharmonising with the initial situation in which it is done. (ii) It will make a *consequential* contribution, by cooperating as a cause-factor with the initial situation to produce a train of consequences which are good or bad.

When we consider an act apart from its tendency to produce such and such consequences there seem to be four factors which might give it value or disvalue. They are (i) its immediate pleasantness or unpleasantness, (ii) the fact that it was done with such and such an intention, (iii) the fact that it was done from such and such motives and in spite of such and such other motives, (iv) its unintended relations to the initial situation in which it was done.

Now an act would certainly derive value, though not moral value, from being a pleasant experience; and it would derive disvalue, though not moral disvalue, from being an unpleasant experience. Again, it would certainly derive *moral* value from being done from certain motives, and *moral* disvalue from being done from certain other motives. But we are leaving motives out of account for the present. Could an act derive value or disvalue from being done with such and such an intention, apart from any question of the motive with which it was done? E.g. could we say that the mere fact that an action was done with the intention of fulfilling a claim gives it some value, and the mere fact that it was done with the intention of frustrating a claim gives it some disvalue, quite apart from the motive with which it was done? To say that a person acts with the intention of fulfilling a certain claim means simply that he believes that one consequence of his act will be that the other party will get his rights in the matter. This particular part of the total foreseen consequences of his action may be of no interest whatever to the agent or may be positively distateful to him. It is difficult to believe that the mere fact that the act was expected by the agent *inter alia* to bring about the fulfilment of the claim gives any kind of value to it. Now if this kind of intention gives no value to an act, it seems unlikely that any other kind of intention would do so. So I am inclined to think that the intention with which an act is done is never in itself relevant to the value of the act. When it is relevant it is because motive is relevant to the value of an act and motive involves intention. Lastly, I do not see that any relation of the act to the initial situation, which falls altogether

outside the agent's intention, would give either value or disvalue to an act. So I am inclined to think that, if we leave motive out of account, as we are explicitly doing at present, the only value or disvalue which an act itself has consists in its own pleasantness or unpleasantness as an experience of the agent.

It does not follow that apart from motive the only *immediate* contribution which an act can make to the amount of good or evil in the world is through its pleasantness or unpleasantness as an experience. For it might make an immediate contribution by harmonising or disharmonising with the initial situation in which it was done. It seems plausible to say that an act done with the intention of fulfilling a certain claim harmonises with, or is fitting to, the initial situation which imposes the claim. Similarly an act done with the intention of frustrating a claim seems to be in disharmony with, or be unfitting to, the initial situation which imposes the claim. And this fittingness or unfittingness of the act to the initial situation seems to be independent of the motive with which the act was done. Suppose, e.g., that the initial situation consists of a question being put to a person, and that the act is an answer by that person. Then an answer which is intended, with whatever motive, to give true information about the *quantum*, seems to be an appropriate or fitting response to this kind of initial situation. And an answer which is intended, with whatever motive, to give false information, seems to be inappropriate or unfitting to it. So I am inclined to think that an act which is intended to be claim-fulfilling makes an immediate contribution to the value in the world simply by harmonising in this peculiar way with the initial situation which gives rise to the claim. Of course this immediate contribution may be more than counterbalanced by the badness which the act may derive from being done, e.g., from a bad motive, e.g. malice. And, even if the motive is good or indifferent, this immediate contribution to the goodness in the world may be more than counterbalanced by the evil in its consequences.

1.3326. The meaning of "right action" and "wrong action". When we talked about a person subject to only one claim I pointed out that the word "right" in the phrase "right action" hovers about between three meanings, viz. "materially obligatory", "formally obligatory" and "subjectively obligatory". We must now see what it means in the more concrete case where a person is subject to several claims which may conflict with each other. I do not think that anyone would claim to define a right act as an optimific act or as an optimising act. It is true that utilitarians hold that any act which would be right in a given situation would also be optimific and conversely. But they regard this as a synthetic proposition, whether *a priori* or empirical, and not as either being or following from the definition of "right". I think that nearly everyone would agree that "right" is definable in terms of most claim-

fulfilling. Utilitarians hold that right acts are always optimific and vice versa because they believe that one's only ultimate obligation is to produce as much good and as little evil as one can.

There is, however, the usual threefold ambiguity to be noted. A "right action" in a given situation may mean a materially most claim-fulfilling act, or a formally most claim-fulfilling act, or a subjectively most claim-fulfilling act. We may distinguish these three senses of "right" as *materially* right, *formally* right, and *subjectively* right.

There are two points to notice about this. (1) Even the most completely objective of these notions, viz. *materially* right, is relative to the powers and capacities of the agent. For it involves the notion of the range of alternatives which *are open* to that agent. This relativity to the agent is quite explicit in the case of formal and subjective rightness. For in the one case it is relative to his state of factual knowledge and belief at the time. And in the other case it is relative, not only to this, but also to his state of ethical knowledge and belief at the time. (2) On the other hand, even the most completely subjective of these notions, viz. *subjectively* right, is not completely relative to a particular individual. It is true that what is subjectively right for Smith will in general be different from what would be subjectively right for Jones in the same situation. But, if so, this is because there is some relevant difference in their factual or ethical beliefs. It is never merely because they are different persons with different tastes and inclinations. Suppose that Smith's and Jones's factual and ethical beliefs about a given situation, and about their own powers, and so on, happened to agree completely. Then an act which was subjectively right for either would be subjectively right for both. If we want to make all this quite explicit we shall have to use the following expressions. (1) So and so would be a *materially* right act for any person of such and such powers in such and such a situation. (2) So and so would be a *formally* right act for any person of such and such powers and with such and such factual knowledge and beliefs in such and such a situation. (3) So and so would be a *subjectively* right act for any person of such and such powers and with such and such factual and ethical knowledge and beliefs in such and such a situation.

The next point to notice is this. There might be a situation in which no act open to the agent is right. For there is generally the possibility of not acting at all, and in some cases it may be that any possible act would frustrate so many or such urgent claims that it is more claim-fulfilling or less claim-frustrating to let things take their course. In that case we should say that it is right to do nothing, but we can hardly talk of *inaction* as a right *action*. We can get out of this difficulty by using the word "behaving" to cover both action and inaction. We can then define the statement "x behaves rightly in situation S" as follows. It means "Either there are right actions open to x in situation S and

he does one of them; or there are no right actions open to him and he abstains from acting". Next we can define the statement "x behaves wrongly in situation S" as follows. It means "Either there are right actions open to x in situation S and he does some other action or abstains from acting; or there are no right actions open to him and he acts". It would be easy to start with these definitions and to define behaving with *material* rightness, with *formal* rightness, and with *subjective* rightness. And similarly for the three kinds of wrongness.

1.3327. The notion of reasonable belief and conjecture. In connexion with subjective rightness there is a distinction to be drawn which we must now consider. In this case we consider an agent whose knowledge is limited and whose beliefs may be more or less mistaken. Now we constantly talk about a certain man's beliefs and expectations being "reasonable" or "unreasonable". We recognise that a belief may be false, and yet it may be reasonable for a certain person to hold it. If an act is to be subjectively right the agent must believe it to be materially right. It does not matter whether this belief is *true* or *false*. But it may make a considerable difference to our valuation of the act or of the agent according to whether the belief is *reasonable* or *unreasonable*. So it will be worth while to go into the notion of reasonable belief fairly fully at this point.

I shall begin with cases where no ethical considerations are involved and gradually work up to the more complex cases.

(a) Suppose a person has to estimate the amount of paper needed to paper a room and is given a set of measurement. If the measurements given are correct and he makes his calculations correctly, he will reach the true answer. His belief will then be both true and reasonable. If the measurements given are incorrect and he makes his calculations correctly, he will reach a false answer. His belief will be false but reasonable. If the measurements given are either correct or incorrect and he makes mistakes in his calculations, he may happen to reach either a true or a false answer. His belief, whether true or false, will be unreasonable. Thus the belief will be reasonable if and only if he makes his calculations correctly.

(b) Suppose that a person knows that a certain coin has been thrown 1000 times in succession and that there have been 495 heads and 505 tails. Suppose now that it is thrown 6 times in his presence and gives heads every time. It is now about to be thrown a 7th time. If the person strongly expects the result of the 7th throw to be *a tail*, "in order", as we might say, "to bring the average right", his strong expectation will be wholly unreasonable even though a tail should in fact turn up. The reasonable expectation is to expect a head or a tail with almost equal conviction. If any preference is reasonable, it would be a very slight preference in favour of the next being a *head*. For what he knows is

that out of 1006 throws, 495 + 6, i.e. 501 have been heads and 505 have been tails, and that there has been a run of heads in the last 6 throws. This suggests that the coin is practically unbiassed, but that there may be some kind of bias in favour of heads in the present thrower's way of casting the die.

(c) Next suppose that the person does not *know* that the coin has been thrown 1000 times and has given 495 heads and 505 tails. He is told this and he believes it, but the information is in fact false. Really it has given 100 heads and 900 tails. What is the reasonable expectation for this man to have about the result of the next throw? It seems to me that it is just the same as it would be if his false belief had been true. Relative to the *actual facts* about the past results of tossing the coin it is of course much more likely that the next throw will give a tail than a head. For it is almost certain that the coin is strongly biassed in favour of tails. But if the only relevant information which the man has about it is the false information that it has given 495 heads and 505 tails in 1000 throws, it would not be reasonable for him to expect a tail more strongly than a head.

There is a sense, then, in which what it is reasonable for a man to expect strongly is contrary, not only to what in fact turned out to be true, but also to what was *objectively probable*. I must now try to explain the distinction between what I call "objective" and "subjective" probability.

A man's beliefs about a certain subject may be *inadequate*, but *true so far as they go*. Or they may be both inadequate and partly false. Now probability is always relative to *inadequate* data. Presumably there is a complete set of facts from which, together with the laws of motion it follows necessarily that the next throw will be a head or follows necessarily that it will be a tail. This would be an adequate set of data. But no one ever knows more than a selection of these facts; and, relative to such a selection there is a certain characteristic probability that the next throw will be a head.

Now take the case of a man whose relevant information is not only inadequate to settle a question, but also partly incorrect. Here there are two notions to be distinguished.

(a) We might imagine his information being *corrected* where it is false, but being in no way *supplemented*. That is, each relevant false belief would be replaced by a corresponding true belief of exactly the same degree of generality or particularity. We can then consider the probability, relative to this corrected but not supplemented set of data, of the proposition which he is considering. I shall call this the *objective* probability of that proposition relative to the *extent* of his information.

(b) Instead of imagining his false beliefs being adjusted to fit the facts, we can imagine that the facts had been such as to fit his beliefs. As before we will assume that his beliefs are in no way supplemented. But we now suppose that the facts had been different in such a way as to make his relevant false beliefs

true. We can then consider the probability, relative to this "cooked" set of data, of the proposition which he is considering. I will call this the *subjective* probability of that proposition relative to the *state* of his information.

In such a case there are therefore four different things to be considered and contrasted: (1) That which the person *in fact* thinks most likely or guesses with most conviction; (2) that which has the highest *subjective probability* relative to the state of his information; (3) that which has the highest *objective probability* relative to the extent of his information; (4) that which in fact is already or turns out to be *true*.

Now (1) will in general differ from (2) if he makes mistakes in logic or in the principles of probability. (2) will in general differ from (3) if some of his information about relevant matters of fact is false. And either (2) or (3) may differ from (4), since probability is always relative to factual data which are *inadequate* to settle the question under consideration.

We can now sum all this up as follows. (1) Suppose that a man's relevant factual information is inadequate to settle a question and leaves two or more alternative answers possible. Then he is justified only in *conjecture*. His conjectures will be *reasonable* if and only if the degree of conviction with which he conjectures each alternative is proportional to the *subjective* probability of that alternative relative to the state of his information. (2) Suppose that his relevant factual information is adequate to settle a certain question, as in my first example of estimating the amount of paper needed to paper a room when an adequate set of measurements is given. Then he is justified in believing with *certainty*. His certain belief will be reasonable if and only if it is a valid logical consequence of his factual information. The four things to be distinguished here are: (1) The conclusion which he *in fact* draws from his data; (2) the conclusion which is *ligically entailed* by his data; (3) the conclusion which would be logically entailed by a corrected set of data in which any errors in his information had been corrected; (4) that which is in fact true. Here (3) and (4) coincide, because the factual data are adequate to settle the question.

It is plain from the consideration of these examples that the reasonableness or unreasonableness of a belief or a conjecture depends entirely on the formal or *a priori* factor which is involved in reaching it. If the belief or conjecture is such as could be reached from the person's premisses without making any mistake in logic, theory of probability, arithmetic, etc., then it is reasonable. It will be reasonable even if it turns out to be false, and even if it be objectively improbable. If, on the other hand, it is such that it could be reached from the person's premisses only by some breach of the laws of logic or probability or arithmetic, then it is unreasonable. It will be unreasonable even if it be true and even if it be objectively probable.

1.33271. The notion of mathematical expectation. There is one other complication to be considered before we can apply the notion of reasonable belief or conjecture to moral questions. Suppose that A and B are two alternative courses of action open to me. Relative to the state of my information it would be reasonable for me to guess that A would have the consequence α and that B would have the consequence β. But let us suppose that it is reasonable for me to feel much more confident that α will follow if A is done than that β would follow if B is done. Suppose, on the other hand, that β would be a much better state of affairs if it should happen than α would be if it should happen. Thus I have to choose between aiming at a more valuable result which is less likely to be attained and a less valuable result which is more likely to be attained. What is it reasonable for me to choose in such circumstances?

A second complication, closely akin to this, is the following. Very often we can say of a proposed course of action only that it will undoubtedly have one or other of a certain set of alternative possible consequences, and that some of them are much more likely to follow than others. Now it may be that certain of these alternative possible consequences would be very good, certain others very bad, and the rest moderately good or bad. Suppose a person were trying to do the optimific or the optimising action, and one or more of the alternative courses of action which he might choose were of this nature. On what principles would it be reasonable for him to choose?

To deal with these questions it is necessary to introduce something analogous to what is called "mathematical expectation" in the theory of games of chance. This is defined as the product of the probability of an event happening by the sum of money which a player will gain or lose if it should happen. I will give some examples.

(1) Suppose that a fair die is to be thrown and that I am to win a shilling if it gives a 6, and am to lose 6d if it gives any of the other five numbers. Then my expectation of gain is $\frac{1}{6}$ th of 1/- i.e. 2d. My expectation of loss is $\frac{5}{6}$ th of 6d i.e. 5d. So my nett expectation is a loss of 3d per throw. It would be reasonable for me to expect to be paid *at least* 3d per throw to induce me to enter the game, and it would be reasonable for the thrower of the die to offer me *at most* 3d per throw.

(2) Now suppose that a rival game is going on. This consists of a fair roulette-board with one red, two white, three blue sectors, all of equal size. I am to lose 6d if the pointer stops at a red sector; I am to win 2/- if it stops at a white sector; and I am to lose 1/- if it stops at a blue sector. The red alternative gives an expectation of loss measured by $\frac{1}{6}$ th of 6d, i.e. 1d. The white alternative gives an expectation of gain measured by $\frac{2}{6}$ th of 2/-, i.e. 8d. The blue alternative gives an expectation of loss measured by $\frac{3}{6}$ th of 1/-, i.e. 6d. Thus my nett expectation per spin is 8d − 1d − 6d, i.e. a gain of 1d. Suppose I had to enter one game or the other, and suppose I made my choice simply from

the standpoint of monetary gain or loss. Then it would be reasonable for me to choose to enter the second game.

We can sum this up as follows. In examples such as we have been considering we can distinguish the following three notions. (1) The act which would *in fact* produce the greatest nett monetary gain or the least nett monetary loss on a given occasion. This might be called the *most fortunate act*. (2) The act which has, relatively to the extent of the agent's information, the greatest nett objective expectation of gain or the least nett objective expectation of loss. This is the act which it *would be* reasonable for the agent to do if he were aiming at doing the most fortunate act and if all his relevant information were correct so far as it goes, and its only defect were its inadequacy. (3) The act which has, relatively to the agent's state of information, the greatest nett subjective expectation of gain or the least nett subjective expectation of loss. This is the act which it *is* reasonable for the agent to do in the actual state of his knowledge and belief, if he is aiming at doing the most fortunate act. We can now apply these notions to the optimific act, the optimising act, and the most claim-fulfilling act.

1.33272. Application to the optimific act. The principles here are exactly the same as in the case of games of chance. But there are a number of additional complications in detail. (1) Each player is no longer playing only for himself unless we accept ethical egoism. He is playing for a syndicate of which he is one member, viz. humanity present and future. In so far as he confines himself to the *average-changing* utility of his acts, it does not matter whether he or some other member of the syndicate is to win a prize or pay a forfeit. (2) The acts of each player do not only produce so much distributable good or evil. They also affect the distribution of distributable goods and evils among the members of the syndicate. Now a given amount of good and evil, distributed in a certain way, constitutes a better total state of affairs than the same amount distributed differently. The player who aims at doing an optimific act will therefore have to consider both the average-changing utility and the distributive utility of his acts. He will have to consider *who* is likely to win the prizes and *who* is likely to pay the forfeits, and what bearing this will have on the welfare of the syndicate as a collective whole or on partial groups within it. (3) The prizes and the forfeits are of the most varied kinds, and it is doubtful whether their values can be compared and measured in terms of some common standard, like money. Again, the value of a combination of goods and evils is not related in any simple way to the values which each constituent would have in the absence of the rest. (4) In the artificial case of the games we assume that the intending player knows exactly what prize he will win or what forfeit he will lose for each of the alternative possibilities. But when one aims at doing an optimific act one is seldom or never in that

position. Even if you know for certain that doing A would bring about α and doing B would bring about β, one may be uncertain whether α would be on the whole better or worse than β.

All these differences make the problem of making a reasonable guess as to which act would be optimific enormously more complicated than that of making a reasonable guess as to which act would be most fortunate in the artificial case. Nevertheless the artificial case shows clearly in a simplified form what is the criterion of a reasonable act for an agent who aims at doing an optimific act. He must consider, to the best of his ability, the relative probabilities of bringing about various alternative results according to whether he does this, that, or the other act. He must estimate, to the best of his ability, the relative values or disvalues of each alternative possible result. And he must weight or discount the probable value or disvalue of each result by the probability or improbability of bringing it about. He will not be able to give an absolute numerical measure of the probability of bringing about a certain set of consequences if he does a certain act. And he will not be able to give an absolute numerical measure of the value or disvalue which this set of consequences would have if it were brought about. But this will not greatly matter provided that he can see (i) that the probability of bringing about one set of consequences is very much or very little greater than the probability of bringing about another set of consequences; and (ii) that the value of one would be very much or very little greater than the value of another. This may suffice to enable him to reject several alternatives. It may leave him with perhaps two between which he can see no reasonable ground for choosing. It will then be equally reasonable for him to do either of these.

We can sum this up as follows. Suppose a person aims at doing an optimific act in a given situation. Then he will be acting reasonably provided he does any act which has at least as great subjective expectation of nett utility or at least as small subjective expectation of nett disutility as any act which it is reasonable for him to believe to be open to him. Such an act will be reasonable with respect to his actual state of knowledge and belief about the relevant facts and about the relative values of the various alternative sets of consequences. It will be not only subjectively optimific but also reasonable. It is subjective, in so far as it is relative to a certain state of knowledge and belief about facts and values. But it is trans-subjective, in so far as it would be the same for *any* person who was in that state of knowledge and belief.

1.33273. Application to the most claim-fulfilling act. The application of the notion of reasonableness to most claim-fulfilling acts is very similar. So far as concerns the question of the consequences of alternative possible acts and the probability of bringing about those consequences it is precisely similar. The ethical part of it is somewhat different. Instead of considering the relative

values of various alternative trains of consequences we shall have to consider the relative urgency of various claims and the extent to which each of them would be fulfilled or frustrated according to which set of consequences follows. As regards ostensibly non-teleological obligations the calculations may be easier here, because we may not have to attend to remote consequences but only to consequences which are almost immediate. E.g. in considering whether an act will fulfil or frustrate the ostensibly non-teleological obligation of truth-telling or of promise-keeping, we have to consider only whether the act will or will not provide the questioner with true information or give the promisee what has been promised. About this we can often be quite certain, and the remoter consequences are here irrelevant. But in many cases this does not help us much. For we are also subject to the teleological obligation of beneficience, whether it be limited to certain persons or groups of persons or be universal. Unless the ostensibly non-teleological obligations are so urgent as to make it reasonable to neglect all other considerations we shall have to consider what would be the most beneficent act, though we may in the end decide that this is not the right act for us to do. So in many cases a person who aims at doing a most claim-fulfilling act will have to consider seriously what would be an optimific act even if he is not a utilitarian.

We defined a subjectively right act in a given situation as an act which the agent believes to be a most claim-fulfilling act among the alternatives which he believes to be open to him in that situation. We see now that the agent's opinion on this point may be either *reasonable* or *unreasonable* in relation to his state of knowledge and belief about the situation and about the laws of nature and about the relative urgency of the various claims upon him. So we have to distinguish subjectively right acts into those which it was and those which it was not reasonable for the agent to believe to be materially right. Now we said that an act is *morally justifiable* if and only if the agent believes it to be materially right. We see now that an act which is morally justifiable, in this sense, may yet be unreasonable in relation to the agent's state of knowledge and belief.

1.333. Motive

We can now remove the third artificial simplification which we introduced at the beginning. Hitherto we have left out of account the agent's motives in doing an act. We must now take these into consideration.

I will begin by reminding you of some results which we reached in discussing motives from a psychological point of view. (1) A motive-factor is any quality or non-causal relation or causal property which the agent believes, rightly or wrongly, that a possible act of his would have, which either attracts him towards or repels him from doing it and thus constitutes for him a reason for or against doing it. Each such factor gives rise to a component of attraction or a component of repulsion. We say that the agent does

an act *from* or *because of* the components of attraction and *against* or *in spite of* the components of repulsion. (2) When we discussed conscience from a psychological point of view we saw that one fact which attracts most people towards doing an act is the belief that it would be right, and one factor which repeals most people from doing an act is the belief that it would be wrong. We will call this the *conscientious motive-factor*. (3) In discussing purity and mixture of motives we saw that a person's motive in choosing alternative *A* in preference to alternative *B* might be either *homogeneous* or *heterogeneous*. If it is heterogeneous it may be either *monarchic* or *polyarchic* or *cooperative*. And if it is cooperative, it may be either *minimal* or *non-minimal*. All these motives were defined and exemplified. We will now use them to define a *conscientious* action.

1.3331. Conscientious action. An action is conscientious if the following conditions are all fulfilled. (1) The agent has reflected on the situation, the alternative actions open to him, and the probable consequences of the various alternatives if done in the present situation, in order to discover what is the right course. In doing so he has tried his utmost to learn the relevant facts and to give to each its due weight, he has exercised his judgment on them to the best of his ability, and has striven to allow for all sources of bias. (2) He has decided that, on the factual and ehtical information available to him, a certain action is probably as claim-fulfilling or as little claim-frustrating as any of the alternatives which he believes to be open to him. (3) His belief that this action would have this moral characteristic, together with his desire to do what is right as such, was either (a) the *only* motive-component for doing it, or (b) *was* sufficient and was *not* superfluous, in presence of the other motive components, to give a resultant motive for choosing this action in preference to the alternatives to it. In our terminology this amounts to saying that an action is conscientious if either (a) the motive for choosing it was homogeneous and its *only* component was the desire to do what is right as such, or (b) the motive for choosing it was heterogeneous but monarchic, and the *governing* motive-component was the desire to do what is right as such. If the first alternative is fulfilled, we can say that the action was *purely* conscientious. If the second is fulfilled, we can say that it was *predominantly* conscientious.

The following would be an example of a predominantly, but not purely, conscientious action. Suppose that a person, who lives in a country where military service is voluntary, decides after reflexion that the right action for him is to enlist. Suppose that these are his motive components which move him to undertake this action, viz. (i) his belief that it is right plus his desire to do what is right as such, and (ii) his dislike of being thought cowardly by his friends if he does not enlist. Suppose that the components moving him not to

enlist are fear of death and wounds, love of comfort, and so on. Then his action in enlisting is predominantly conscientious if the following two conditions are fulfilled: (a) his belief that it is right plus his desire to do what is right as such *would* have sufficed to overcome his fear, his love of comfort, etc. even in the absence of his dislike of being thought cowardly. And (b) his dislike of being thought cowardly *would not* have sufficed to overcome those anti-components in the absence of his desire to do what is right and his belief that it is right to enlist. In such a case there *is* a non-conscientious component for doing the action which the agent believes to be right, but it is both superfluous and insufficient. It would be absurd to refuse to call the action "conscientious" in such a case.

We come now to the more difficult and doubtful cases. The first is when the motive for choosing the action os polyarchic, and the conscientious component is sufficient but superfluous. This would be illustrated by our first example if we varied it as follows. Suppose now that the agent's dislike of being thought cowardly *would* have sufficed to overcome his fear and his love of comfort and *would* have induced him to choose the course of action which he now believes to be right, even if his desire to do what is right or his belief that it is right had been absent. Here we have two pro-components, one conscientious and the other not. Each is sufficient by itself, and therefore each is superfluous in presence of the other. All that is necessary is that *one or other* of them should be present. The difficulty of this case arises as follows. If you confine your attention to the *sufficiency* of the conscientious component, you will be inclined to say that the action *is* conscientious. If you confine your attention to the *superfluity* of this component you will be inclined to say that the action is *not* conscientious.

The second doubtful case is when the motive for choosing the action is cooperative and minimal and the conscientious component is not superfluous but is also not sufficient. This would be illustrated by the following modification of our original example. Suppose now that neither the conscientious component, in absence of the dislike of being thought cowardly, nor the latter in absence of the former, would have sufficed to overcome the agent's fear and his love of comfort. Only the combination of the two suffices to do this. Each pro-component is now indispensable and neither of them separately is sufficient. The difficulty in this case arises as follows. If you confine your attention to the *indispensability* of the conscientious component, you will be inclined to say that the action *is* conscientious. If you confine your attention to the *insufficiency* of the component, you will be inclined to say that the action is *not* conscientious.

I will group together purely and predominantly conscientious actions, in the sense defined, under the name of *fully conscientious* actions. I will group together the two doubtful cases, which we have just been discussing, under

the name of *semi-conscientious* actions. We can then subdivide the latter into (i) those in which the conscientious component is sufficient but superfluous, and (ii) those in which it is indispensable but inadequate.

Suppose that a person deliberately does an act which he believes to be less claim-fulfilling or more claim-frustrating than some other act which he believes to be open to him at the time. Then he must be acting against his desire to do what is right and to avoid doing what is wrong as such. Any act of this kind may therefore be called *contra-conscientious*.

It is plain that many of our deliberate actions are neither fully conscientious nor semi-conscientious nor contra-conscientious. For one may have decided to do an action without having considered it and the alternatives to it from the standpoint of rightness or wrongness. Such acts may be called *morally unconsidered*. A *morally unconsidered* act may be such that, *if* the agent had considered it and the alternatives to it from a moral standpoint, he *would* have judged it to be at least as claim-fulfilling or as little claim-frustrating as any act which he believed to be open to him. Since he did the act without having made this judgment about it we can say at once that the conscientious motive would have been *superfluous*. But we cannot say for certain whether it would have been *sufficient*. Again it may be that, *if* the agent had considered the act and the alternatives to it from a moral standpoint, he would have judged it to be less claim-fulfilling or more claim-frustrating than some other alternative which he believed to be open to him. And it may be that he would still have decided to do it in spite of this opinion about its moral character. If both these hypothetical conditions were fulfilled we could call this morally unconsidered act *potentially contra-conscientious*. Thus a morally unconsidered act may be potentially contra-conscientious or potentially semi-conscientious, but it cannot be potentially fully conscientious.

It is plain that any conscientious act is subjectively right. But it may be unreasonable and formally wrong and materially wrong. On the other hand an act may be subjectively right without being in any sense conscientious. No doubt the agent's belief that it is materially right will attract him to some extent unless he is a moral lunatic. But this conscientious component of attraction may have been neither a sufficient nor an indispensable factor in moving him to do the act. He may have known or believed, e.g., that the act would injure a person whom he disliked or would benefit his country. The attraction due to one or other or both these factors might have been sufficient to induce him to do this act, and the conscientious component of attraction might not have sufficed, in the absence of the malicious or the patriotic component of attraction.

In that case this subjectively right act was not conscientious, though it was also not contra-conscientious.

There are two other remarks worth making about conscientious action at

this point. (1) A *purely* conscientious action, as distinct from one that is *predominantly* conscientious, must be a very rare event. For a person's motives for and against doing an act of importance in any fairly complex situation are certain to be mixed. E.g. it is hardly credible that either undertaking or refusing to undertake military service during a war should be a purely conscientious act. For everyone fears death and wounds, on the one hand, and everyone dislikes to incur the suspicion of cowardice and selfishness, on the other.

Now the definition of "predominantly conscientious acts" and of the two kinds of "semi-conscientious acts" show that they all have the following peculiarity. They all involve the notion of what *would* have happened if certain conditions had been other than they in fact were. E.g. would the conscientious pro-component have been strong enough to overcome the anti-components if the non-conscientious components which in fact cooperated with it had been absent or had been weaker? The notion of the consequences of unfulfilled conditions always enters whenever the question of sufficiency and dispensability is raised. It follows that an individual can seldom be rationally justified in feeling any strong conviction that an action of his was conscientious. For, in order to decide this question, he has to form an opinion as to how he would have acted in the *absence* of certain motive-components which were in fact *present*. It seems to me that *a fortiori* it must be almost impossible in many cases to decide rationally on whether another person's action is conscientious or not.

The rough and ready test which commonsense applies is this. One feels fairly confident that a man's act was conscientious if one has reason to believe that all the following conditions are fulfilled. (1) If there are very strong and very obvious motives *against* doing it which affect practically all human beings. E.g. if doing it involves the practical certainty of torture and death to oneself and ruin to one's family and friends. (2) If there are no very obvious strong motives *for* doing it except the one which the agent alleges, viz. the belief that it is right and the desire to do what is right as such. Suppose, e.g., that a solitary atheist who disbelieved in human survival were captured by a tribe of fanatical Christians and were told that, if he would profess Christianity, he would be set free, and, if he would not, he would first be tortured and then killed. Suppose it were afterwards discovered that this man never expecting that any account of his action would reach the R.P.A. or the Anti-God League, had refused to profess Christianity on the ground that he thought it wrong to profess what he disbelieved. Then I should be strongly inclined to think that this action had been purely or predominantly conscientious. But in real life such extreme cases are very rare. In most real cases I should find it difficult to entertain any confident opinion about the conscientiousness of an act whether my own or another's.

(2) The other point to notice is this. In my definitions of "conscientious action" I took the components *against* the act as fixed. I then raised questions about what the agent would have done if the non-conscientious pro-components had been absent and the conscientious pro-component present, and vice versa. I think that this is correct. But of course another question can be raised. Let us suppose that a certain act was fully conscientious. We could ask ourselves "How strong was the agent's disposition to act conscientiously?" Here we should have to take the non-conscientious pro-components as fixed and imagine the anti-components increased in number or in intensity. The question would then be "At what point, if any, would the agent begin to act against his conscience?" Suppose, e.g., that the death-penalty were imposed for refusal to undertake military service. Then it is quite certain that some people, who now conscientiously refuse, would contra-conscientiously accept; and it is fairly certain that others would still conscientiously refuse. This would not imply that, under the actual present conditions, the refusal of the first class of persons is not conscientious. It would show only that their disposition to act conscientiously, though strong enough for the actual situation, is weaker than that of the second class of persons.

1.3332. Other motives. Instead of isolating the conscientious motive-factor and discussing it in the way in which I have done, we might take any other important motive-factor and discuss it in a similar way. Take, e.g. malice. We could ask ourselves whether an agent's *only* motive for doing a certain act was his belief that it would injure a certain person and his desire that that person should suffer. If so we could call the act *purely* malevolent. Suppose that other motive-factors were present attracting the agent towards doing this act. Then we could ask ourselves whether he would have done it from malice even in their absence. And we can ask whether he would have avoided doing it if these other pro-components had been present and the component of malice had been absent whilst the anti-components had been the same as before. If both these questions can be answered in the affirmative we should say that the act was *predominantly* malevolent.

1.3333. Ethical bearing of motives. I think it is admitted by everyone that an agent's motives in doing an action have a very important bearing on the *moral goodness or badness* either of the act or of the agent or of both. Some people hold that the agent's motives in acting also have a bearing on the *rightness or wrongness* of his act. But others deny this. It is denied, e.g., by Ross.[1] It is also denied by Mill.[2]. Mill says that motives have a great deal to do with the

1. W.D. Ross, *Foundations of Ethics* (Oxford, 1939).
2. J.S. Mill, *Utilitarianism.* (Reprinted from *Fraser's Magazine*, Oct.-Dec. 1861, London, 1863, and many subsequent reprints.)

goodness or badness of the agent, but nothing to do with the rightness or wrongness of his acts. We will now consider these two points in turn.

1.33331. Motives and moral value. It is quite in accordance with usage to talk of *good* and *bad* motives. People also talk of "right" and "wrong" motives. E.g. it would be quite usual to say that so-and-so did the right thing from wrong motives or the wrong thing from right motives. But I do not think that "right" and "wrong", when used in this way and applied to motives, mean anything different from "morally good" and "morally bad" as applied to motives. So it will be better to confine ourselves to the latter expressions when talking of motives and to keep "right" and "wrong" for acts, as before.

As we have seen there are two aspects about a motive, viz. a cognitive and a conative-emotional. It is a belief that a contemplated possible act would have a certain quality or non-causal relation or would produce a certain kind of result. That is its cognitive aspect. This belief is toned with desire or aversion and possibly with some special kind of emotional tone. That is its conative-emotional aspect. Now what makes a motive morally good or bad is a certain kind of *fittingness* or *unfittingness* between its conative or emotional aspect and the ostensible property of the act or its consequences in respect of which this conation or emotion is directed towards it. It is fitting to feel attraction towards an act in respect of one's belief that it would be right. It is unfitting to feel attraction towards an act in respect of one's belief that it would corrupt another person's character or make him unhappy. That is why we call the conscientious motive good and the malevolent motive evil.

1.333311. The conscientious motive and moral value. The question whether an *act* is morally good or not seems to come down to the question whether it is morally creditable or discreditable to the agent. This in turn seems to come down to the question whether it is the outcome of a good or a bad disposition in the agent. Now, as we have seen, a person's dispositions fall into three distinct but intimately interconnected groups, viz. cognitive, conative, and emotional. Now cognition may be concerned either with mere matters of fact or with moral claims, obligations, etc. and with moral and other values. Again a person may be moved to act or abstain from acting in a certain way either by his beliefs and desires about the non-moral properties of the proposed act, e.g. its pleasantness, or by his beliefs and desires about its moral properties, e.g. its rightness. Lastly, certain of his emotions may be directed on to persons or actions in respect of their supposed non-moral properties, e.g. the supposed beauty or wit of a person, the supposed cleverness of an action, etc. Or they may be directed on to persons or actions in respect of their supposed moral properties, e.g. the supposed honesty of a person, the supposed rightness of an act, etc. Thus cognitive, conative, and

emotional dispositions must each be subdivided into *moral* and *non-moral*.

Now conscientious acts must be distinguished into those which are *reasonable* and those which are *unreasonable* in respect of the agent's state of knowledge and belief at the time. And among those which are unreasonable we must distinguish between those which are *ethically* unreasonable and those which are *only factually* unreasonable.

An act which is fully conscientious but ethically unreasonable is creditable to one and only one part of the agent's moral nature, viz. to his moral *conative-emotional* dispositions. It is not creditable to the whole of his moral nature; for it is discreditable to his moral *cognitive* powers and dispositions that he should misjudge, e.g., the relative urgency of various moral claims or the relative value of various possible states of affairs. We may say that such an act is a sign of *moral good-will*, and is so far morally creditable; but it is also a sign of moral stupidity or moral delusion, and is so far morally discreditable. A fully conscientious act which is ethically reasonable but factually unreasonable is wholly creditable to the agent's moral nature. It is discreditable only to his *non-moral* cognitive powers and dispositions.

A person who is either ethically or non-ethically unreasonable to a high degree may be much more harmful to his fellows if he is extremely conscientious than if he is not. For his cognitive stupidity or craziness is likely to lead him, either through ethical or factual mistakes, to believe certain acts to be right which will really inflict serious wrongs. And his strong desire to do what is right is likely to make him carry his mistaken beliefs into action in face of all difficulties, where a less conscientious person with the same mistaken beliefs would be content to do nothing or to do what he believed to be wrong.

It is therefore evident that it may be *materially* right for other individuals or the authorities in a society to prevent a conscientious person from doing what he believes to be right or to try induce or force him to do what he believes to be wrong. Undoubtedly this is in itself a bad thing, but it may be the lesser of two evils. Moreover, if other individuals or the authorities in a society honestly believe that it is materially right to treat a certain conscientious individual in this way, then it is *subjectively right* and *morally justifiable* for them to do so, even if their belief is false or unreasonable. And if they persecute him from the belief that it is right to do so and the desire to do what is right as such they are acting *conscientiously*. And in that case their act is *morally creditable* in precisely the same sense and for precisely the same reason as the act of the individual whom they are persecuting. What is sauce for the conscientious goose is sauce for the conscientious ganders who are his neighbours or his rulers.

This fact is often obscured by the following causes. Many people inadvertantly or dishonestly confine their attention to certain historical cases, such as the trial and execution of Socrates or of Christ, which have two

peculiarities. Here later generations have held (1) that the individual was not only conscientious but also *correct* in his ethical opinions, and (2) that the tribunal which condemned him was either ethically unreasonable or was not acting conscientiously. It is useful to take as a corrective example the case of a high-minded Indian official conscientiously securing the capture and execution of a high-minded Thug for conscientiously practising murder.

An act which is fully conscientious and both factually and ethically reasonable is creditable to the agent in every respect. It may not be materially right. But, if it is not, this is through no defect, either voluntary or involuntary, in the agent. Moreover, it is plausible to hold that a person who habitually acted both conscientiously and reasonably would be more likely to do materially right acts than an otherwise similar person who acted on any other plan or on no plan at all. For, by hypothesis, such a person always has to find out what is materially right, always comes to the conclusions which are reasonable on the data at his disposal, and then always does what he believes to be most probably right. So, provided that a person is both factually and ethically reasonable, the more conscientious he is the more likely he is on the whole to give to his fellow men their material rights.

I will now say something about the moral disvalue of contra-conscientious acts. If a person does what he believes to be wrong or wittingly fails to do what he believes to be right, it shows that other motive-components were stronger than the conscientious one. So far this is always morally discreditable. But the degree of moral discredit depends very much on the nature and the strength of the motives which overcame the conscientious one. The maximum discredit is when the conscientious component was overcome by others which are themselves bad, e.g. by a malevolent motive. Next come cases where the motive which overcomes the conscientious one is in itself neither good nor bad but is a comparatively weak morally indifferent motive. For this shows that the conscientious component must have been very weak. The kind of case which I have in mind is where a person fails to do what he believes to be right merely because it would involve a little exertion or some slight inconvenience, such as writing a letter or missing an enjoyable dinner-party. The minimum moral discredit is when the motive which overcomes the conscientious one is both intense and morally respectable.

An example would be the case of a judge or a general who believed it to be right to inflict the death-penalty on his son who had incurred it by an act which was in some ways highly to his credit. If this judge or general failed to inflict the death-penalty, he would certainly be acting contra-conscientiously, and this would be morally discreditable to him. But we should not be inclined to condemn him severely (though it might be right to punish him severely) in view of the kind of motive which had overcome his conscience. Next comes the case where the motive which overcomes the conscientious one is neither

good nor bad but is very intense and is common to nearly all men. What I have in mind here is the case of a man who acts against his conscience when exposed to threats of ruin, torture, and death. We greatly admire a person who acts conscientiously in spite of such motive-components against doing so; but, if we have any imagination and power of introspection, we do not severely condemn a person who fails to do so.

1.333312. Other motives and moral value. We come now to other motives. It seems clear that some motives are morally bad and that they make any act done from them morally discreditable even if it is not contra-conscientious. E.g. any act which is done purely or predominantly from the desire to corrupt or to give pain to another person is morally discreditable. Such an act may be materially right. It may even be subjectively right. If it is subjectively right it will not be contra-conscientious. But it is also not conscientious. E.g. a master may believe that it is right to beat a certain boy and he may be correct in that opinion, but his predominant motive for beating the boy may be a cruel desire to see him suffer. If so, his act is morally discreditable; though it is both materially and subjectively right and is not contra-conscientious.

Some motives are in themselves neither good nor bad. But they may lead a person to do acts which are contra-conscientious and therefore morally bad. E.g. desire for one's own happiness and safety is certainly not directly evil, like desire to injure another person. In itself it is morally indifferent. But, if it is strong, it is very liable to prevent a person doing what he believes to be right if he thinks that the right action is likely to be painful or dangerous. Can we say that any motive except the conscientious one is good without qualification? We might be inclined to say this of benevolence, i.e. the desire to benefit others. But, unless utilitarianism is true, there are other component obligations, such as promise-keeping, which are not reducible to beneficence and may conflict with it. Even if utilitarianism is true, it is certain that many people do not accept it. Now a person who is not a utilitarian may be in a position in which the act which he believes to be most claim-fulfilling and the act which he believes to be most beneficent are different. If such a person in such a situation does the latter act from benevolence he does a contra-conscientious act. I do not think that we could say that this was morally creditable to him, though it would be much less discreditable than if he had acted against his conscience through fear or laziness or malice. Suppose, on the other hand, that the act which he thinks right and the act which he thinks most beneficent happen to coincide, and that he would have done it from the conscientious motive alone. Then I think that the presence of the benevolent motive makes the conscientious act more *amiable*, though I doubt whether it makes it morally better. A person who is intentionally benefitted by another would generally prefer that the agent should be moved wholly or partly by affection

for him than wholly by a sense of duty towards him. And a spectator might find the former kind of act more attractive to contemplate than the latter. (Cf., e.g., the expression "cold as charity".) But neither of these considerations seems to be relevant to the *moral values* of the two acts, i.e. roughly to the credit which they do to the agent as a moral being.

We may sum up as follows. Although the conscientious motive is not the only one which is good and which can make an act done from it morally creditable, it does stand out ethically from all other motives. (i) No act can be morally creditable if it is done against this motive. This cannot be said of any other motive. (ii) Although the presence of a strong and persistent desire to do what is right as such is by no means sufficient to constitute a good character, it is of more fundamental importance for that purpose than any one other desire. It is not sufficient to constitute even a *morally* good character; for it may be accompanied by stupidity or craziness about right and wrong, good and evil. And, even when accompanied by moral insight and intelligence, it is not sufficient to constitute a good *all-round* personality. For this requires other powers and dispositions in addition to the specifically moral ones. But it is true that, unless a person has a fairly strong and persistent desire to do what is right as such, he is very unlikely to be a good person or to become a better one. And he is almost certain not to remain a good person if his circumstances should become unfavourable to virtue. I do not think that this can be said of any one other desire.

1.33332. Motives and rightness. I have no doubt that the very ambiguous word "right" is sometimes used in such a way that an act would not be called "right" if it were done from a bad motive. It may even be sometimes used in such a way that an act would not be called "right" unless it were purely or predominantly conscientious. But it is certain that it often is used in other ways. For, in the first place, it is quite sensible to say that a person did a right act from a bad motive, or a wrong act from a good motive. Secondly you cannot define "conscientious" without introducing "right" in a sense which does not involve any reference to motive. For a conscientious act is one in which the agent's only motive or his predominant motive for doing the act was his belief that it is right and his desire to do what is right as such. Now the sense of "right" which enters into this definition of "conscientious" cannot in turn involve a reference to the conscientious motive. Moreover it is clear that, when a person considers what is the right act for him to do in a certain situation, he does not as a rule consider his own motives at all. He considers the probable consequences of the various alternatives open to him, the various claims upon him, and the extent to which these various consequences would bring about the fulfilment or frustration of these claims. So I think that there is no doubt that there is an important sense of "right" and

"wrong" act, in which the question of motive does not enter; and this sense is required in the definition of a conscientious or a contra-conscientious act. It would seem to be desirable to keep the words "right" and "wrong" quite free from all reference to motive. We can substitute "morally creditable" for "right" and "morally discreditable" for "wrong" when the latter words are used in a sense which does involve a reference to motives.

Apart from the reason which I have given for this there is another reason which has been given by Ross.[1] I propose to state it in my own way.

When we defined "right action" we said that right actions are a sub-class of a certain wider class, viz. of actions which are open to the agent at the time when he has to act. We defined this class as a set of alternative possible actions such that any member of it would become actual if and only if the agent decided to enact that one. Now suppose that the property of being done from a certain motive were made part of the definition of a right action. Then right actions would fall *outside* the class of actions which are open to the agent, in the sense just defined. For it is not within the power of a person at any given moment to determine by a mere act of will whether he shall be moved by this, that, or the other motive. Suppose, e.g., I am asked a question and have to do something about it immediately. Then the alternatives open to me are to refuse to answer it, to answer truly, or to tell one or other of a number of different lies. These are all open to me in the sense that any one of them will become an actual event if and only if I choose to enact that one. Now consider any one of them, e.g. giving a true answer. No doubt there is a sense of "might" in which it is quite correct to say that I "might" give a true answer from various motives, e.g. from the conscientious motive, from malevolence, from fear of being caught out in a lie and punished, and so on. But this sense of "might" is different from the sense in which I "might" give a true answer, or refuse to answer, or tell one or another of several lies. And the alternatives of doing an act from this, that, or the other motive are not open to me in the sense in which the alternatives of doing this, that, or the other act are open to me. This is obvious when one goes into detail. To say that I tell the truth from malevolence means that I believe that doing so will hurt another person, and that this belief attracts me so much that it overcomes the repulsion due to any other beliefs that I may have about the act of truth-telling in this situation. To say that I tell the truth from fear of detection and punishment means that I think I am likely to be caught and punished if I lie, and that this belief repels me so much from lying that it overcomes any aversion to truth-telling which may arise from other beliefs which I have about its effects in this situation. Now it cannot be said that each of these beliefs will arise if and only if I decide to have it. At any given moment one cannot give oneself this or that belief at will. Again, suppose that either or

1. Ross, *op. cit.*

both of these beliefs already exist in me. Then it cannot be said that I can make which of them I choose predominantly attractive or repulsive to me by any act of will that I can here and now perform.

Suppose we were to include among our alternatives, not only different possible intentional acts, but also the same intentional act done from different possible motives. Then we should have a set of alternatives which are not all open to the agent, in the sense in which the different possible acts, regardless of motive, are open to him. Now there is no doubt that "right" and "wrong" are commonly used in such a way that the following two conditions hold. (a) Any act that is either right or wrong must be an act open to the agent. (b) In any situation there is at least one right way of behaving (including inaction under this head) open to the agent. Now, if being done from a certain motive be made part of the definition of right action, this will break down. Alternatives which are not open to the agent might be right or might be wrong. And all the alternatives open to an agent might be wrong. Since this is so, it is plainly undesirable to use "right" and "wrong" in such a way that they involve a reference to motive. It is also quite unnecessary, since we can use the words "morally creditable" and "morally discreditable" when we want to bring in this reference.

There are several points to be considered before leaving this subject. (1) I think that there are cases where one takes into account the motive from which one would be acting if one were to enact a certain alternative, and when this seems to have a bearing on which alternative it is right for one to choose. Suppose I had to decide whether I would or would not prosecute a certain person for a crime. Suppose he were a man whom I disliked; or that he was a professional rival of mine, so that I should benefit from his downfall. I should certainly have to reflect that, if I decided to prosecute him, one motive-factor for my action will be desire to gratify personal hatred or to ruin a professional rival. Another factor may be a disinterested desire that the criminal shall be punished for the welfare of the community and for his own reformation. I might come to the conclusion that it is very doubtful whether this latter motive would be strong enough by itself to induce me to prosecute. If so, I should have to admit that it is very doubtful whether the act of prosecuting, if I should decide on it, would not be predominantly the outcome of bad motives in me. I might very well decide, on that ground, not to prosecute. Would this be right? And does it conflict with the results of our previous argument?

I do not think that it upsets the argument, though it shows that one of the premisses needs to be carefully stated. We must distinguish between first-order motives and second-order motives. The second-order motive here is the disinclination to indulge first-order desires which I judge to be morally bad. In so far as I have second-order desires, like this, it is to some extent within

my power to choose between acting from this, that, or the other first-order motive. It is still true that, just before I make my decision, my motives will be what they have by then become, and that neither their presence or absence nor their relative strength can be affected by any voluntary decision that I could then and there make about them. But it is true that their relative strength may by then have become very different from what it was when I began to reflect. And it is true that an important cause-factor may have been my own reflexions on the moral value and disvalue of the first-order motives which were going to be involved, and the appeal which these reflexions made to certain of my second-order desires.

This settles the psychological question. What about the ethical question? In considering what it is right to do in such a case it would be ethically relevant to consider the effect on one's own character of doing one alternative or another. It might be that, if this consideration were left out of account, alternative x would be preferable to alternative y. But suppose that the doing of x would indulge a morally bad conative disposition, whilst the doing of y would not. Then one consequence of doing it might be to strengthen this bad disposition, whilst the doing of y would not have this kind of bad consequence. In that case, if x were in other respects only a little more claim-fulfilling than y, this difference in their effect on the agent's character might suffice to tip the balance and make y right and x wrong for this agent to do. I think that this is the only way in which motive might be relevant to rightness.

Much the most important reason for hesitating to do the act which seems right in such cases, e.g. prosecuting a criminal whom one personally dislikes, is epistemological. Can one be sure that one's dislike of him has not exercised an irrational influence on the cognitive process by which one reached the conclusion that it was right to prosecute him? The point here is not that you may be acting from a bad motive if you prosecute him. It is that desires and emotions are aroused which may lead you to believe that prosecution is the right act though really this is an unreasonable conclusion on the evidence available to you. So, even if you should be acting conscientiously in prosecuting him, there is a danger that your conscientious act may be ethically or factually unreasonable.

1.34. Summary

This completes the analysis of rightness and the notions bound up with it which we began a long time ago. It may be summarised and concluded as follows. We can imagine what might be called an *ideal act* for an individual in a given situation. This would combine the following properties. (1) It would be materially right, i.e. it would *in fact* bring about a certain change which *would* satisfy as fully as any other change *actually* producible by the agent all the claims which the situation *as it really is does in fact* impose on him. (2) It

would be *subjectively right,* i.e. the agent would believe it to be materially right. (3) This belief of the agent's would be *reasonable* on the factual and ethical data at his disposal. (4) The act would be *conscientious*, i.e. the agent's only motive or his predominant motive for doing it would be his belief that it is right together with his desire to do what is right as such.

Now there are certain reasons which make it unlikely in many cases that any act performed by an agent could fulfil all these conditions. The fundamental difficulty is to combine conditions (1), (2) and (3). Suppose that the agent's knowledge is inadequate and that some of his beliefs are mistaken. Then, if he draws the conclusions from them which are reasonable, it is extremely unlikely that an act which he *believes* to be materially right will *in fact* be so. It is of course *not impossible* that one and the same act should be materially right and should be reasonably believed to be so by an agent with limited knowledge and partly mistaken beliefs. But it is a bit of remarkable luck if it should be so.

Let us now consider the ideal act and the various possible departures from it, from the standpoint of the agent himself and from the standpoint of the persons affected by the action. (a) So long as the persons affected by the act confine their attention to this particular situation, all that is important to them is that the act shall be materially right. If and only if it is materially right it will bring about the satisfaction of their claims on the agent as fully as any act open to him could do. From this point of view it does not matter what the intention or the motive of the agent may have been. (b) If the persons affected look beyond the present situation and the present act, they cannot take this simple view. They have got to coexist with the agent and they or their friends may expect to be involved in other transactions with him in future. From this point of view the important point is the indications which the present act gives of the agent's character and dispositions. What mainly matters is that it should be conscientious and reasonable. If it has these properties, even though it be not materially right, it indicates a character and dispositions which are likely to lead to materially right acts on other occasions. (c) Failure to do a materially right act need involve no kind of discredit to the agent, since it may arise from lack of knowledge which no human being could have had. Provided that the act is reasonably believed by the agent to be materially right on the information available to him, and is done wholly or predominantly from the desire to do what is right, it is wholly creditable to his present character and dispositions. (d) If the agent unreasonably believes the act to be materially right and does it wholly or predominantly from the desire to do what is right, it is creditable to his moral conative dispositions, but it is discreditable to his cognitive powers and dispositions. If ethical unreasonableness has been involved, it is discreditable to his moral insight and power of moral discrimination. (e) If the agent believes the act to be materially

wrong, it must be contra-conscientious and therefore discreditable to his moral conative dispositions. (f) Even if the agent reasonably believes the act to be materially right, it may still be discreditable to his moral conative dispositions; for his only motive or his predominant motive for doing it may have been an evil one, such as malevolence. (g) When the agent has made up his mind as to which of the alternative acts open to him are right and which are wrong, it depends on nothing but his volition at the moment whether he does one that he thinks right or one that he thinks wrong. If and only if he sets himself to do one of the former he will do it; if and only if he sets himself to do one of the latter he will do it. In this sense it is always within a person's power to do or to leave undone what he believes to be right. (h) In this sense it is *not* within a person's power at any given moment to act from this, that, or the other motive. The motives which move him now may have been in part determined by the voluntary decisions which he made in the remoter past. E.g. the present strength of his desire for alcohol may be in part determined by his having chosen to indulge it to excess in the past. But at the moment when he chooses to do a certain action he cannot also choose the motives from which he will do it. These motives are the *causes* of his present choice of that action and they cannot be the effect of any volition that he can make at the time. (i) On the other hand he may in the course of his deliberations reflect that if he chooses a certain alternative he will be indulging a certain desire. He may believe that this desire is morally bad in itself, or that to indulge it would weaken his character, or he may know that indulging it in the past has led to consequences which he has regretted. These reflexions may cause him to desire not to indulge that desire. And it may be that it would not be right on the whole to choose the alternative which would involve indulging that desire. Thus in the course of his deliberation second-order motives may supervene and they may make his final decision different from what it would have been if he had not reflected on the first-order motives to which the various alternatives appeal.

1.4. Theories of right and wrong

Prima facie there seem to be a large number of different kinds of circumstances which impose moral claims on a person and they do not seem to be reducible to any one principle. Naturally moral philosophers have not been willing to rest content with this. They have tried to show that all the different kinds of obligation can be reduced to a single fundamental *principle*. Again, plain men have wanted some *criteria* by which they could judge what is right or wrong in doubtful and complex cases.

I will begin by pointing out the difference between a unifying principle and a criterion. The notion of criterion is essentially practical. We say that X is a

satisfactory criterion for *Y* if and only if the following conditions are fulfilled. (1) If and only if *X* is present then *Y* is present also. (2) The presence or absence of *X* is considerably easier to detect than the presence or absence of *Y*. A good criterion of *Y* may be a rather superficial and unimportant characteristic. E.g. in chemistry the criterion for the presence of an acid in a solution is the rather trivial circumstance that a certain complicated organic substance (*litmus*) turns red when put in contact with it. Conversely an important and fundamental characteristic which is present when and only when *Y* is present may be quite useless as a criterion for *Y*. For its presence or absence may be even harder to recognise than that of *Y*.

It is important to bear this distinction in mind in what follows. Moralists who put forward a monistic theory about what makes right acts right are often not clear whether this is supposed also to supply a handy criterion for deciding what is right or wrong in particular cases. If it is put forward as a criterion it may be open to criticisms which would be irrelevant if it is not.

The most important attempts to provide a monistic theory of the grounds of moral obligation or the criterion for moral obligation are utilitarianism and Kant's theory. I will now consider them in turn.

1.41. Utilitarianism

We must first notice that utilitarianism might take several different forms. It might be either *analytic* or *synthetic*. Then synthetic utilitarianism might be held either as an *a priori* or as an *empirical* proposition. Lastly any of these three forms of utilitarianism might be combined with either a *hedonistic* or a *non-hedonistic* theory of good and evil. So there would be six possible varieties in all. I will now say something about these divisions.

(1) *Analytic utilitarianism* is the doctrine that the sentence '*X* is a right act in the situation *S*' *means* the same as the sentence '*X* is an act whose consequences will be at least as good as those of any other act open to the agent in the situation *S*'. Similarly it holds that the sentence '*X* is a wrong act in the situation *S*' *means* the same as the sentence '*X* is an act whose consequences will be less good than those of some other act open to the agent in the situation *S*'. I have already said that "right" and "wrong" are very vague and ambiguous terms. But I think it is quite safe to say that in ordinary life we *never* mean this by them. Nobody finds any *contradiction* or *absurdity* in the supposition that in a certain situation the right act may be to tell the truth whilst the most benefic act would be to tell a certain lie. In particular, if the evil results of telling the truth and the good results of telling a certain lie will wholly or mainly affect the agent himself, the plain man thinks it paradoxical to say that it is obviously wrong to tell the truth and right to tell this lie. It may be *proved* to us that an act which is right is always optimific, and that an act which is optimific is always right; just as it may be proved to us that a triangle

which is equilateral is always equiangular and that a triangle which is equiangular is always equilateral. But it is as certain that we do not mean the same by the two words "right" and "optimific" as it is that we do not mean the same by the words "equilaterally triangular" and "equiangularly triangular". So we may reject analytic utilitarianism at once, and confine our attention to synthetic utilitarianism.

(2) *Synthetic utilitarianism* holds that, although the words "right" and "optimific" are not just two names for one and the same characteristic, yet a right act in any situation is always an optimific act, and an optimific act in any situation is always a right act. Now this doctrine might take two different forms. (a) It might hold that the mutual implication of these two characteristics is a *necessary* fact like the mutual implication of equilateral and equiangular triangularity. If so, we presumably see it directly to be necessary when we reflect on the two characteristics, or else we deduce it from other propositions which we can see directly to be necessary. This form of utilitarianism may be called *a priori synthetic utilitarianism*. (b) It might be held that the mutual implication of the two characteristics is, so far as we can tell, a *contingent* fact, like the mutual implication of cloven-footedness and chewing-the-cud. If so, we reach our belief in it by problematic induction performed on the results of a great number of observations. In that case the doctrine can never be more than a highly probable empirical generalisation. This forms of the theory may be called *empirical synthetic utilitarianism*. Hume was certainly a utilitarian of this kind. I am inclined to think that Sidgwick was a utilitarian of the *a priori synthetic* kind. It is true that he appeals, in Book IV of the *Methods of Ethics*, very largely to empirical facts about the actual moral judgments of men.[1] But I think that his object is to convine non-utilitarians, by means of examples, that men really assume the utilitarian principle and judge in accordance with it even when they seem not to.

(3) Utilitarianism has generally been combined with *ethical hedonism*, i.e. the doctrine that nothing is intrinsically good or evil but experiences, and that the only good-making or bad-making characteristic of an experience is its pleasantness or unpleasantness respectively. The name "utilitarianism" is often used in such a way that it connotes ethical hedonism. This is, however, very inconvenient. Utilitarianism is a theory about right and wrong, whilst ethical hedonism is a theory about good and evil. And it is logically possible to combine a utilitarian theory about right and wrong with a non-hedonistic theory about good and evil. I shall therefore distinguish between hedonistic and non-hedonistic utilitarianism. The latter is sometimes called "ideal utilitarianism", but this name suggests a surreptitious attempt to recommend the theory to virtuous people.

1. H. Sidgwick, *The Methods of Ethics* (1st edition, London 1874; 7th edition, London, 1907).

1.411. Argument for utilitarianism

I will now state what seem to me to be the essential points in the argument for utilitarianism. I think that the argument falls into two parts. The first part tries to show that the utilitarian theory has certain positive advantages, viz. that it has a certain *prima facie* plausibility, and that it would introduce a unity and coherence into ethics which are lacking on non-utilitarian views, provided that it can account for all the admitted facts. The second part tries to deal with the facts which appear to favour a non-utilitarian theory and to be difficult to reconcile with utilitarianism. It tries to show that these facts can be explained in terms of the utilitarian theory, together with certain admitted facts about human psychology and certain very plausible hypotheses about primitive societies. Taken as a whole, the argument for utilitarianism may be compared to the argument for the heliocentric theory and against the geocentric theory in astronomy. The facts about our moral judgments correspond to the observed positions of the planets night after night. The utilitarian claims to provide a single simple hypothesis which will account for all the admitted facts.

The argument may be put as follows. (a) If I am told that a certain act would be right or would be wrong in a certain situation, it is always reasonable to ask "*Why* is it right, or *why* is it wrong?" "What *makes* it right or what *makes* it wrong?" This would be admitted by non-utilitarian moralists, like Ross. (b) One answer which would often be given is of the following kind. You will be told that the act is right because, by doing it, you will be producing as much good or as little evil as you can under the circumstances. Or you will be told that the act is wrong because, by doing it, you will be producing less good than you might or more evil than you need under the circumstances. (c) Even non-utilitarian moralists, like Ross, admit that this kind of answer is, in many cases, correct. But they hold that, in many cases, a different kind of answer must be given, an answer which makes no reference to the production of good or evil. Very often the answer would take the following form. "That act is right because it will bring about the fulfilment of a promise which you have made" or "That act is wrong because it will produce a false belief in the mind of a questioner about the subject on which he has asked you a question". Utilitarians must admit that such answers to such questions very often are given, and that they often are regarded as quite satisfactory. They must admit that we often decide that an act would be wrong because it would be an act of intentional deception or of promise-breaking, without considering whether the consequences in the given case will be good or bad. And they must admit that we sometimes decide that such an act would be wrong even when, so far as we can see, the results of doing it would be better than those of any alternative open to us in the situation. (d) At this stage the utilitarian does two things. (α) He tries to show that these

other kinds of answer are not ultimate, and that they all depend on the first kind of answer. And (β) he tries to show how the apparent exceptions to his theory can be explained in terms of his theory. We will take these two points in turn. (α) The first kind of answer to the question "What makes this act right, or what makes it wrong?" seems much more ultimate and intellectually satisfactory than any of the other kinds of answer. One does not feel inclined to raise the question "*Why* should it be wrong to produce less good than I might or more evil than I need?" or "*Why* should it be right to produce the *most* good or the *least* evil that I can under the circumstances?" It seems self-evident that the first kind of act would be wrong and the second kind right. But it does not always seem reasonable to rest content with answers of the other kind. One is inclined to ask "*Why* should the mere keeping of a promise, as such, be right, regardless of whether the promisee and others are the better or the worse for his getting what he has been promised?"; "Why should the mere producing of a false belief in the mind of a questioner be, as such, wrong, regardless of whether the questioner and others are better or worse for his having a false belief about the *quaesitum*?" The position then is this. There is one kind of answer to the question "What makes this act right or makes it wrong?" which is admitted by everyone to be *in many cases* the correct answer. This kind of answer seems to be ultimate and intellectually satisfactory, and to raise no further questions. There are also a number of other kinds of answer. Between them they refer to a whole litter of right-making and wrong-making characteristics, which seem to constitute no coherent system. And these answers do not seem to be ultimate or intellectually satisfactory; they seem to be merely preliminary answers which raise further questions. The utilitarian concludes that probably the first kind of answer is the fundamental one, and that all the others can be reduced to it, in so far as they are valid. Moreover, in reducing the other answers to the one fundamental kind, we shall see the interconnexion of the former, and the limits within which they are valid. We shall thus introduce order and coherence into the mere litter of rights and duties and claims which the non-teleological theory presents to us as ultimate.

(β) This brings us to the utilitarian attempt to reduce all other kinds of answer to this kind. The main line of argument here is as follows. (i) We must first recall certain important distinctions which we drew when we discussed the notion of utility in connexion with optimific acts. We distinguished utility into *primary* and *secondary*. Secondary utility or disutility is that which arises from the general knowledge or belief that a certain act has been done in certain circumstances and the fact that it may be widely imitated. Under the head of primary utility we distinguished between the *normal* utility of acts of a given kind, e.g. returning borrowed property, and the *individual* utility or disutility possessed by a particular act of that kind performed in particular

circumstances, e.g. returning a borrowed revolver to a person who has gone mad since the loan was made. We also distinguished between the *singular* utility of acts of a certain kind taken one by one, e.g. of any one act of trespass on a field of wheat, and the *collective* utility or disutility of a combination of many such acts within a limited area or a limited period, e.g. a hundred people trespassing in a field just before harvest time. (ii) Some rules are much more fundamental than others from the standpoint of utility. Unless we can count on agreements being generally kept, even when it is inconvenient to one of the parties to keep them, no one will make agreements. Now organised society would be impossible without many agreements being made and on the whole being kept. And the existence of a reasonably stable and complex society is a necessary condition without which most kinds of good are impossible. Thus the rule that agreements shall be kept, even when it is inconvenient to one of the parties to keep them, is an extremely fundamental rule from the point of view of utility. (iii) Suppose that an exceptional case arises in which it would really be more advantageous to all parties concerned that a certain moral rule should be broken than that it should be kept, provided that the general public did not know of the breach of the rule. It still does not follow on utilitarian principles that the rule should be broken even in this case. For either the fact that it has been broken will become generally known or it will not. If it becomes generally known, other people will be inclined to overlook the special circumstances which made a breach of the rule beneficial in this particular case, and they will be encouraged to break it in cases where these exceptional circumstances do not exist. If the rule has great *normal* utility or great *collective* utility, the results of its being often broken will be bad. Again, the known breach of *any* generally accepted rule is liable to undermine respect for the whole system of rules, and this is likely to lead to bad consequences on the whole. If, on the other hand, the parties concerned conceal the fact that they have broken the rule, the combination of ostensible conformity with real non-conformity, the constant need for caution, and the occasional need for deceit, may have worse consequences than would have been involved in keeping the rule even in the exceptional circumstances in which they were placed. (iv) It is of great utility that the principle of utility should not have to be appealed to in every case where a moral decision has to be made. Most people have neither the time nor the ability nor the freedom from prejudice which would be needed if they are to weigh up carefully the good and bad consequences of all the alternative actions which they might do in every case where they have to make a decision. It is therefore of great utility that the experiences of mankind about what sorts of acts have on the whole led to good results, and what sorts of acts have on the whole led to bad results, in certain frequently recurring types of situation, should be crystallised into rules to which a kind of superstitious sanctity is attached. In

most cases, if you act on these rules without thinking of the reasons for them, you will in fact bring about the best consequences open to you; will save yourself a great deal of time and trouble; and will have the additional motive of superstitious awe for doing what is in fact right or avoiding what is in fact wrong. It is only in specially complex cases that it will be either necessary or desirable, on utilitarian principles, for a man to appeal to the principle of utility. We might compare this to the use of empirical rules of play by bridge-players. In so far as the rules are valid, they could be justified by a mathematician on the principles of probability. But the experienced bridge-player need not consider their logical basis. He will use them automatically in ordirinary situations. And, if he is a good player, he will see for himself when situations arise in which he must not blindly follow them but must appeal to the special circumstances of the moment and to the first principles of the game. (v) It is now easy to see how a utilitarian would deal with Ross's notion of *prima facie* duties, or with what I have called ostensibly non-teleological components of obligation. Certain kinds of action in certain frequently recurring kinds of situation have very great normal utility or very great collective utility. Others have very great normal disutility or very great collective disutility. There are certain ways of acting in certain frequently recurrent kinds of situation which would make an organised society quite impossible if they were to become common rather than exceptional. *Not* to behave in these ways in such situations is therefore a necessary condition for securing all those goods, and avoiding all those evils, which depend on the existence of an organised society for their production or their avoidance respectively. Promise-keeping, truth-speaking, etc. are examples of kinds of action in frequently recurring kinds of situation, without which organised society would be impossible. They are also the most striking instances of ostensibly non-teleological obligations. Now in many cases an individual is under a strong temptation *not* to do such an action when placed in a relevant situation. He is strongly tempted to do an action of a kind which, if it became common, in such situations, would make organised society impossible. Very often the latter kind of action would quite obviously benefit himself, and would not by itself do much, if any, harm to others. It is therefore of great utility that people should have a very strong motive *for* acting in the one way and *against* acting in the other way. Let us imagine two societies, S_1 and S_2. In S_1, from some cause or other, a strong emotion of fear or disgust has become attached to breaking one's promises or telling lies to another member of the society. In S_2 there is no such emotion, and the only motive for not breaking one's promises or lying to one's neighbours when it seems convenient is the recognition that, if this becomes common, the society will break up and all the goods which depend on its continuance will vanish. It is obvious that promises will much more often be kept and that lies will much less often be

told to our neighbours in S_1 than in S_2. It is therefore much more likely that S_1 will become a powerful society than that S_2 will, and that S_1 will have the opportunity to impose its characteristic emotion towards such acts on other people whom it conquers or absorbs. The utilitarian therefore suggests that ostensibly non-teleological obligations are, in the main, rules of conduct which fulfil the following conditions. (a) Situations in which they are relevant are fairly common in the lives of most people. (b) If they were broken in any considerable proportion of the situations in which they are relevant, organised society would be impossible. (c) It would often be to an individual's real or apparent advantage to break them; or there are strong and widespread impulses which, when aroused, are likely to cause them to be broken. On the other hand, the disadvantages of their being frequently broken are remote, and are visible only to a person who reflects calmly and takes a very extensive view. (d) In certain societies in the remote past a direct feeling of obligation has somehow become attached to these rules, and a direct emotion of guilt has become attached to breaches of them. This has provided a strong additional motive to members of such societies for keeping the rules when members of other societies would break them. (e) Once this attachment of a direct emotion of obligation or guilt to such rules has been established, from whatever cause, it will tend to be propagated and to spread. It is propagated by education and tradition from one generation to another, since any society at any moment consists of a majority of grown people and a minority of children. And it tends to spread for the following reasons. A society in which this kind of emotion has become attached to rules which really are essential to stability and organisation will be stronger than societies in which no such emotion has become attached to any rules, or in which it has become attached to rules which are useless or positively detrimental to stability and organisation. Thus a society of the first kind will tend to conquer and absorb societies of other kinds, and to impose its ways of thinking and feeling on them *directly*. And, in so doing, it will acquire prestige, and thus may *indirectly* impose its ways of thinking and feeling on other societies which it does not conquer. Cf., e.g., the effect of the prestige of Greece on republican Rome, and the effect of the prestige of the Roman Empire on the barbarians who lived on its outskirts and eventually destroyed it.

It seems to me that this form of the utilitarian argument is, on the whole, extremely plausible. I will now make some comments on it. (i) It does not assume that primitive men were enlightened utilitarians. It does not assume that *they* saw that the real reason why truth-speaking, promise-keeping, etc. are right is that no stable society is possible unless promises are generally kept and true answers generally given to questions. It does not assume that *they* saw that individuals have such strong motives for lying and promise-breaking that a very strong additional motive on the other side is needed if promises are

to be generally kept and true answers generally given. And it does not assume that *they* deliberately set themselves to provide such additional motives by attaching strong direct emotions by education, etc. to certain kinds of action. If it assumed these propositions, it could safely be rejected. All that the utilitarian needs to assume is that *somehow or other* a strong attracting emotion got directly attached to *some* kinds of act, and a strong repelling emotion to *some* other kinds of act. Once this is granted, it follows that societies in which attracting emotions got attached, from whatever cause, to the kinds of behaviour which in fact are essential to preserve society, and in which repelling emotions got attached, from whatever cause, to the kinds of behaviour which in fact undermine society, will tend to persist and to expand. Societies in which, from whatever cause, emotions of the first kind have got attached to acts of the second kind, and vice versa, will tend to break down. And so, after a time, most societies which still persist and flourish will be those in which the kinds of behaviour which are felt to be directly obligatory are those which in fact are essential to preserve society, and the kinds of behaviour which are felt to be directly wrong are those which would in fact destroy society if they were prevalent. At this stage men may begin to reflect on ethical subjects. They may then see that the kinds of act which they have felt to be directly obligatory are mostly of the first kind, and that the normal and collective utility of such acts is the only reasonable ground for continuing to regard them as obligatory. They may see that the kinds of act which they have felt to be directly wrong are mostly of the second kind, and that the normal and collective disutility of such acts is the only reasonable ground for continuing to regard them as wrong.

(ii) The utilitarian can explain why a good many types of action are held to be right or to be wrong, as such, although reflexion shows that they have no great utility or disutility. It was not because of their utility that the feeling of obligation was originally attached to kinds of behaviour which are in fact useful. And it was not because of their disutility that the feeling of guilt was originally attached to kinds of behaviour which are in fact socially destructive. Hence it is reasonable to suspect that these emotions will have become attached directly to many types of action which are neither socially useful nor socially destructive. It is, indeed, quite compatible with the utilitarian argument that certain types of behaviour which were *always* socially detrimental should be felt to be directly obligatory, and that certain types of behaviour which would *always* have been socially useful should be felt to be directly wrong. A society can swallow and digest without disaster a great deal of what a utilitarian must regard as moral rubbish or moral poison, provided that certain absolutely essential kinds of act are felt to be directly obligatory and certain absolutely destructive kinds of act are accompanied by a direct feeling of guilt.

(iii) Again, it is extremely likely that some kinds of action, which *were* essential to the preservation of society when the feeling of obligation first became attached to them, no longer are so in the very different conditions of a contemporary industrial society. Such actions may now be useless or destructive. Similarly, it is extremely likely that some kinds of action, which *would have been* destructive to society when the feeling of guilt first became attached to them, no longer are so in the very different conditions of a contemporary industrial society. Such actions may now be harmless or useful or even essential to the preservation of society. Yet the feeling of obligation will still be attached to acts of the first kind, and the feeling of guilt will still be attached to acts of the second kind. Contemporary morality will always include a good deal of what a utilitarian must regard as vestigial organs, like the appendix, which were once useful but are now at best useless and at worst actively harmful. In this way the utilitarian can explain quite plausibly, in terms of his theory, the existence of many apparent exceptions to his theory. I compared the argument for utilitarianism to the argument for the heliocentric theory in astronomy. The analogy is now seen to be even closer than it seemed at first sight. The heliocentric theory, together with the theory of gravitation, explains not only the average periodic movements of the planets, but also their detailed irregularities and perturbations, which might seem at first sight to be incompatible with the theory. And the theory is greatly strengthened by this fact. The utilitarian would claim that his theory, together with the admitted laws of human psychology and a highly plausible hypothesis about primitive societies, will account for the facts which seem to be at first sight incompatible with the theory.

(iv) Most utilitarians have also been ethical hedonists. Consequently most attempted refutations of utilitarianism have been occupied at least as much with ethical hedonism as with utilitarianism. We are not at present concerned with the truth or falsity of ethical hedonism. But ethical hedonism is a monistic theory of goodness, and we can raise the question "How far is utilitarianism strengthened and how far is it weakened by being associated with a monistic theory of goodness?" (a) One of the most attractive features of utilitarianism is its claim to introduce order into the chaos of component obligations of various degrees of moral urgency which are involved in such a theory as Ross's. According to utilitarianism there is one and only one question to be considered when there is a conflict between the various claims on one. I have only to consider which of the alternative acts open to me will be most benefic or least malefic on the whole, after allowing for all the evils that may indirectly result from breaking any rules of which the *normal* and the *collective* observance is highly benefic. Now, if there is one and only one kind of good-making or evil-making characteristic, e.g. hedonic tone, this introduces a real unity and coherence into ethics. But, if there is a plurality of irre-

ducible good-making and evil-making characteristics, we have got out of the chaos of irreducibly non-teleological obligations only to fall into the chaos of irreducible intrinsic goods and evils. It seems to me then that, unless utilitarianism can be combined with a monistic theory of goodness, its main positive merit is considerably diminished. It will still introduce *some* simplification; for it reduces all the litter of right-making and wrong-making components to the two characteristics of being benefic and being malefic respectively. And it makes the characteristic of being optimific the necessary and sufficient condition of resultant rightness. But it would plainly introduce much more simplification if it could be combined with a monistic theory of good and evil, such as ethical hedonism. Utilitarianism, combined with ethical hedonism, has no doubt often won supporters for this reason, who would have been much less attracted by utilitarianism combined with a pluralistic theory of good and evil. (b) On the other hand, the difficulties of utilitarianism are considerably increased if it is combined with ethical hedonism or any other monistic theory of good or evil. I will just indicate the main difficulty. (α) People undoubtedly think that acts which would distribute the same net balance of distributable good and evil in different ways may be made right or wrong by the kind of distribution which they bring about. (β) A utilitarian who admits a pluralistic view about good and evil can easily deal with this fact. He will say that the occurrence of pleasant experiences and the occurrence of unpleasant experiences in individuals *are* good-making and evil-making factors respectively, but that they are not *the only* such factors. Certain ways in which these pleasant and unpleasant experiences are distributed among the members of a society are also good-making or evil-making characteristics. The resultant goodness or badness of the total state of this society at any time will depend on both kinds of factor. And the right act will be that which has the greatest net *totalising* utility, when both kinds of factor are taken into account. (γ) A utilitarian who takes a monistic view about good and evil cannot consistently take this line. If he holds, e.g., that hedonic tone is the *only* valifying characteristic, he cannot *also* hold that the way in which pleasant and unpleasant experiences are distributed among members of a society is a valifying characteristic. For he would then be holding that there are at least two different kinds of good-making and bad-making characteristic. (δ) Such a utilitarian will therefore have to take one of two courses.

(A) He may at this point partly desert utilitarianism. He may say that it is *right* to distribute a given amount of distributable good and evil in certain ways and *wrong* to distribute it in certain other ways, although the state of affairs in which it is distributed in one way is neither better nor worse than the state of affairs in which it is distributed in another way, and although no better ulterior consequence will follow from one distribution than from an-

other. He might hold, e.g., that the only intrinsic goods and evils are pleasant experiences and unpleasant experiences respectively. And he might hold that our fundamental duty is to maximise the amount of good and minimise the amount of evil in the world. Nevertheless this duty is subject to certain non-teleological restrictions. It is always wrong, as such, to distribute good and evil in such a way as to conflict with certain very abstract principles of distribution which can be formulated. It will be wrong even if doing so would bring about a greater balance of distributable good over distributable evil in the long run. (B) A really "tough" monistic utilitarian will refuse to desert his utilitarianism at this point, and will proceed as follows. He will say that there is no good or evil but *distributable* good and evil of a *single* kind, e.g. pleasant or unpleasant experience.

Our only duty is to maximise the amount of distributable good and to minimise the amount of distributable evil in the universe, and this duty is subject to no limitations. Any two states of affairs in which there is the same net balance of distributable good and evil are equally good or equally bad, no matter what the distribution may be. And we have no direct non-teleological obligation to bring about one mode of distribution rather than another. But, although one mode of distribution is not, as such, *intrinsically* better or *intrinsically* worse than another, one may be more benefic or more malefic than another. If certain amounts of good and evil be distributed in one way among the members of a society, the most competent of these may be encouraged and the most idle may be stimulated. If the same amounts of good and evil be distributed in a certain other way among the members of this society, the most competent of them may be discouraged and the most idle may be given no incentive to work. The result of the first kind of distribution is that there will be a greater balance of means to distributable good available for distribution in future. The result of the second kind of distribution will be the opposite. We may express this by saying that some modes of distribution have greater *fecundity* than others. (This phrase is due to Bentham.) Now the really tough monistic utilitarian will say that the *only* rational ground for preferring one mode of distribution to another is that the former has greater fecundity than the latter.

Now both the alternatives which are open to a utilitarian who holds a monistic view about good and evil are somewhat unsatisfactory. If he takes the first alternative, it is doubtful whether he can consistently stop just where he wants to. He will hold that certain kinds of act must be ruled out as wrong *merely* because they conflict with certain self-evident principles about distribution, and quite apart from all consideration of the goodness or badness of their consequences. If this is admitted, it seems doubtful whether the exceptions to the utilitarian principle can be confined to the very narrow limits within which a utilitarian would want to confine them. If it is admitted that

there are *any* self-evident limitations on the one fundamental obligation to produce as much good and as little evil as possible, people will begin to claim that many other and more positive non-teleological obligations, such as truth-telling and promise-keeping, are self-evident also. And so we shall be back at the litter of irreducibly non-teleological obligations which utilitarianism claims to abolish. On the other hand, the view that the *only* ground for counting one mode of distribution as right and another as wrong is that the former has greater fecundity than the latter seems extremely paradoxical to common sense. Most people would be inclined to say that some ways of distributing goods and evils are so "unfair" that it is wrong to distribute in this way even if this distribution has greater fecundity than others which are less unfair.

1.412. Sidgwick's form of utilitarianism

I will now say something about Sidgwick's form of utilitarianism.[1] Sidgwick was an ethical hedonist about good and evil, but we shall not be concerned at present with this part of his theory. His doctrine may be summed up as follows.

(i) There are certain necessary synthetic propositions about rightness and goodness, which can be seen directly to be necessary by anyone who carefully reflects on them. (ii) These do not suffice to tell us what is right in any given situation, or even in definite classes of situations, such as being asked a question, being called upon to fulfil a promise, etc. But they do rule out certain kinds of action as wrong in any situation. They may be compared, in this respect, to certain very abstract physical principles, like the conservation of energy or the principle of least action. We know that any physical theory that conflicts with these must be incorrect; but the mere fact that a physical theory is consistent with them does not suffice to prove that it is correct. (iii) The principle of utility is the necessary and sufficient condition for deciding which action or class of actions, among those which are *not* ruled out by these *a priori* axioms about rightness and goodness, is the right action in a given situation or class of situations. But it needs to be supplemented by these axioms. I think Sidgwick would hold that other utilitarians have tacitly assumed these axioms, but have failed to see that they need to be explicitly formulated in addition to the principle of utility. Let us now consider Sidgwick's axioms. They are to be found in Bk. III, Chap. XIII of the *Methods of Ethics*. There are six of them, and they fall fairly definitely into three classes. I will begin with those which are explicitly about rightness, and I will state them in my own way. (1) Suppose that A and B are two agents. Suppose that a certain act, if done by A, would be right; whilst a precisely

1. A much more detailed discussion of Sidgwick's ethics is given by Broad in his *Five Types of Ethical Theory* (London, 1930), ch. 6.

similar act, if done by *B*, would be wrong, or conversely. Then this ethical dissimilarity in the acts must be due to some *qualitative* or *relational* dissimilarity between the agents. It can never be due to the mere numerical otherness of the agents. It must depend on some definite dissimilarity in the powers or qualities or dispositions of *A* and *B*, or on some definite dissimilarity in their situations and their relations to other persons or things. (2) Suppose that there are two alternative acts, β and γ, open to a certain agent *A* at a given moment. β would affect a person *B*, and not *C*. γ would affect *C*, and not *B*. The effect of β on *B* would be precisely similar to the effect of γ on *C*. Suppose that β would be right and γ would be wrong, or conversely. Then this ethical dissimilarity in the acts must be due to some *qualitative* or *relational* dissimilarity in the patients. It can never be due to the mere numerical otherness of the patients. It must depend on some definite dissimilarity in the powers or dispositions or qualities of *B* and *C*, or on some definite dissimilarity in their relations to *A* or to others. (3) Suppose that a person has to administer a law. Then it is always wrong for him to treat differently two people whose position in respect of this law is precisely similar. And it is always wrong for him to treat similarly two people whose positions in respect of this law are dissimilar. He ought to take account of those circumstances which the law contemplates as relevant, and only of those circumstances, in administering the law.

I will now make some comments on these three alleged *a priori* principles about rightness. The first two might be criticised from two opposite points of view. Some people might say that they are true but completely trivial, and others might say that they are not true without exception. Let us consider these two criticisms. (1) The first may be put as follows. No doubt the mere fact that Smith and Brown are different persons will never make it right for Brown to do what it is wrong for Smith to do, or conversely. And no doubt the mere fact that Smith and Brown are different persons will never make it right for Jones to do to Smith what it would be wrong for him to do to Brown, or conversely. But these facts are of hardly any ethical importance. If *I* propose to do what I admit would be wrong for Smith to do, I never excuse myself by merely reflecting that I am not Smith. I always should refer to dissimilarities between my nature or relationships and Smith's nature or relationships. Now there always will be plenty of such dissimilarities, and sometimes they do make an action which would be wrong for one right for the other. So the thing that we want to know is this: What kind of dissimilarity of quality or relationship between *A* and *B* is relevant, and what kind is irrelevant, to the question whether a certain kind of act which would be right for one to do would be wrong for the other? If *A* likes *X* and *B* dislikes *X*, this dissimilarity might make it right or at any rate permissible for *A* to make a proposal of marriage to *X* and wrong for *B* to do so. But it would not make it

right for *A* to recommend *X* for a post for which she is applying and wrong for *B* to do so. Now Sidgwick's first axiom throws no light on questions of this kind, which are the only questions that really arise in such matters. Again, suppose I am thinking of treating Smith in a way in which I admit that it would be wrong for me to treat Brown. I never excuse myself by merely reflecting that Smith is not Brown. I should always refer to dissimilarities between the natures or the relationships of Smith and of Brown. Now there always will be plenty of such dissimilarities, and sometimes they do make it right for me to treat one in a way in which it would be wrong for me to treat the other. So the thing that we want to know is this: What kind of dissimilarity of quality or relationship between *B* and *C* is relevant, and what kind is irrelevant, to the question whether it would be right to do to one of them a certain kind of act which it would be wrong to do to the other? If *B* is *A*'s sister and *C* is his second cousin, this dissimilarity might make it right or at any rate permissible for *A* to marry *C* and wrong for him to marry *B*. But it would not make it right for him to recommend one for a post for which she was applying, and wrong for him to recommend the other. Sidgwick's second axiom throws no light on questions of this kind, which are the only questions that really arise in such matters.

(2) It might be thought that, if Sidgwick's first two axioms are trivial, they are at least obviously true. But this is very doubtful. Let us consider the second axiom. This is, no doubt, true, provided that *A* the agent, is a different person from both *B* and *C*, the two patients. But a person may be both agent and patient, since he can affect himself by his own actions. We might therefore have a case where *A* is the agent, and the two patients are *A* himself and *B*. Will the axiom hold then? Is it not sometimes right for *A* to do to *A* what it would be wrong for him to do to *B*, or conversely? And is not the ground of this ethical dissimilarity in the acts sometimes simply the fact that one patient is *A* himself whilst the other patient is not himself? E.g. is it not often the case that it is right to give a pleasure to another, whilst it is morally indifferent to give a precisely similar pleasure to oneself? And is the ethical dissimilarity not due simply to the fact that in one case the recipient is oneself, and in the other case, is another person?

A very similar difficulty arises about the first axiom. Here we have two agents *A* and *B*. The axiom seems evident if the act affects only some third party *C*. But suppose it is an act which affects either *A* or *B* but not both of them. Let us suppose, e.g., that it affects *A* and not *B*, and that it consists in giving a certain pleasure to *A*. If this is done by *A*, it will be an egoistic act; if it is done by *B* it will be an altruistic act. If done by *A*, it may be morally indifferent or even wrong; and, if done by *B*, it may be right.

I think it is plain from these examples that Sidgwick's first two axioms need to be more carefully stated. The difficulties that I have mentioned involve

some rather interesting logical points. (i) It is easy to give non-ethical examples in which a proposition which is obviously true when *A*, *B* and *C* stand for three different terms, becomes false if *B* or *C* is allowed to be identical with *A*. Take the following geometrical axiom: "If *B* and *C* be two points and *A* is collinear with them, then either *B* is between *A* and *C*, or *A* is between *B* and *C*, or *C* is between *B* and *A*." This is obviously true if *A* is understood to be distinct from both *B* and *C*. But it becomes false if *A* is allowed to coincide with either *B* or *C*. (ii) Propositions in which the phrase "having *R* to itself" occur need to be carefully analysed. Consider the property of "being loved by *A*" where *A* is a proper name of a certain self. Then we must distinguish between the two propositions "*A* is loved by *A*" and "*A* is loved by himself". No doubt they are logically equivalent, i.e. if *either* is true, then *both* are true. But the two properties "being loved by *A*" and "being an object of self-love" are plainly quite different. E.g. suppose that *B* loves *B*. Then *both A* and *B* have the property of being objects of self-love. But *B* need not have the property of being loved by *A*, and *A* need not have the property of being loved by *B*.

We can now state our criticism of Sidgwick's first two axioms as follows. Suppose that an act is going to affect a certain person *B* and him only. Then there will be a certain characteristic dissimilarity according to whether it is done by the person *B* or by any other person. If it is done by *B*, it will be a *self-affecting* act. If it is done by any other person, it will be an *other-affecting* act. Now this kind of dissimilarity in the acts, though it depends merely on the numerical identity or the numerical otherness of the agent-self and the patient-self, may be ethically relevant. If the agent-self and the patient-self are the same, the act may be right; if they are different, it may be indifferent or wrong; or conversely. And the ground for the ethical dissimilarity of the acts may be simply that the one is self-affecting and the other is other-affecting.

In this form the criticism applies primarily to the first axiom. But it can be extended at once to the second axiom. Suppose, in the second axiom, that we allow the possibility that the agent *A* may be identical with one of the patients, e.g. *B*. Then an act done by *A* to *B* who is now identical with *A* will differ in a certain characteristic way from an otherwise similar act done by *A* to *C* or to any other person. If it is done to *B*, who is now identical with *A*, it will be a *self-affecting* act; if it is done to *C* or to any other person, it will be an *other-affecting* act. And this dissimilarity in the acts may be ethically relevant. Undoubtedly common sense thinks that it is highly relevant in many cases. In using Sidgwick's axioms we must then remember that, if an act would be self-affecting when done *by* one agent or done *to* one patient, it must be compared only with an act which would be self-affecting when done by another agent or to another patient. If it is other-affecting when done by

one agent or done to one patient, it must be compared only with an act which would be other-affecting if done by another agent or to another patient. It is useless to compare an act which is self-affecting with an act which would be other-affecting, or vice versa, however alike they may be in all other respects. For the dissimilarity may be the ground of an ehtical dissimilarity between the two acts. Yet this dissimilarity depends simply on *numerical* identity or otherness.

I will not waste time on Sidgwick's third axiom, but will pass at once to two of his axioms about producing and distributing good. I will first state them in his own words. (1) "The good of any individual is of no more importance, from the point of view of the universe, than the good of any other." (2) "It is my duty to aim at good generally, so far as I can bring it about, and not merely at a particular part of it." (*Op. cit.*, p. 382.)

Now these two axioms involve a number of highly obscure and questionable terms. What is meant by the "good of an individual", and what is meant by "the point of view of the universe", in the first axiom? I think that Sidgwick means by "the good of Smith" those good experiences which are Smith's experciences, and by "the evil of Brown" those bad experiences which are Brown's experiences. It would be more accurate to talk, as McTaggart does, of the good *in* Smith and the evil *in* Brown.[1] The net value in Smith would be estimated by balancing his good experiences against his bad experiences. Now any experience will be owned by some one experient. We can therefore distinguish, in regard to any experience, two characteristics. (a) Being an experience of a certain kind, e.g. a twinge of toothache of a certain quality and degree of unpleasantness and duration. (b) Being owned by a certain experient, e.g. by Smith. Now suppose that an experience of a certain perfectly determinate kind could be produced either in *A* or in *B*. *A* will of course always be dissimilar to *B* in many respects. They will have more or less dissimilar dispositions and past experiences, and will stand in more or less dissimilar relationships. In consequence of these dissimilarities the amount of good or evil in the universe might be changed to a very different extent according to whether an experience of a certain perfectly determinate kind were produced in *A* or in *B*. I am sure that Sidgwick did not mean to deny this perfectly obvious fact. I think that his first axiom might be stated as follows: "If the amount of good or evil in the universe would be changed to a different extent according to whether an experience of a certain kind were to occur in *A* or in *B*, then this difference cannot be due to the mere numerical otherness of *A* and *B*. It must always be due to specific dissimilarities in the qualities, or dispositions or relationships or past history of *A* and *B*." If this is what he means, his first axiom about producing and distributing good seems to me to be true, but completely trivial.

1. J.McT.E. McTaggart, *The Nature of Existence*, vol. II (Cambridge, 1927).

We will now consider his second axiom about producing and distributing good. This says that "it is my duty to aim at good generally, so far as I can bring it about, and not merely at any particular part of it". Now we must remember that Sidgwick was a utilitarian about right and wrong and an ethical hedonist about good and evil. As a utilitarian, he believed that one's fundamental duty was to produce as much good and as little evil as one can. As an ethical hedonist, he held that the only things that can be intrinsically good or bad are experiences, and that the only characteristic of an experience which makes it good or bad is its pleasantness or unpleasantness. Now for the present purpose we need not assume the truth either of utilitarianism or of ethical hedonism. It is enough to assume, what nearly everyone would grant, viz. (a) that at any rate *one* important *prima facie* component obligation is to produce as much good and as little evil as one can, and (b) that experiences are an important class of things which can be intrinsically good or bad, even if they are not the *only* such things and even if their pleasantness or unpleasantness is not the only property of them which can make them good or bad. We can then take this axiom to be concerned with those of our actions which are intended to produce good experiences and to avert bad experiences of any kind in any person, i.e. with beneficent actions.

The axiom can now be interpreted as follows. In so far as it is my duty to aim at producing good experiences and averting bad ones, it is my duty to try to produce the greatest possible net balance of good over bad experiences throughout all present and future persons whom my action can affect. If I confine my beneficent efforts to myself, or to my family, or to my class, or to my countrymen, or to my contemporaries, or to any other restricted group of experients, I need to have some positive justification for this restriction. And there is one and only one kind of justification which is valid. The only valid justification for any limitation in the range of my beneficent efforts is that owing to my special limitations or their special relations to me, I can produce most good on the *whole* by confining my beneficent efforts to a *certain restricted part*. A restriction in the range of one's beneficent efforts always needs ethical justification, and the ethical justification must always take this form. This axiom is certainly not trivial, for it would be unhesitatingly rejected by many people. Most people would be inclined to think that I have a more urgent duty to benefit those who stand in certain relations to me than I have to benefit others who do not stand in those relations, and that this special urgency depends directly on these special relations. E.g. it would commonly be held that the mere fact that A is my mother and B is my second cousin makes it my duty to aim at A's happiness rather than at B's, even if I could easily make B happier than I could possibly make A. Sidgwick would have to say that, in view of the actual limitations of each man's powers and sympaties, on the whole a greater balance of good experiences is produced in the

universe by each person concering himself mainly with the welfare of his own parents and near relations, and leaving the welfare of others to be mainly looked after by *their* near relations. And he would have to hold that this is the *only* reason why I have a more urgent duty to aim at producing a balance of good experiences in my mother than to aim at producing such a balance in my second cousin. Now this may be true, but it is certainly not self-evident to me. It seems at least as plausible to hold that certain special relations to me involve in their very nature special claims on my beneficence.

1.413. Ethical egoism, neutralism and altruism

The doctrine which emerges from Sidgwick's axioms may be called *ethical neutralism*, as opposed both to *ethical egoism* and *ethical altruism*. The neutralist theory is that no one has any special duty to himself, *as such*; and that no one has any special duty to others, *as such*. The fundamental duty of each of us is simply to maximise the balance of good over bad experiences in the universe as a whole, so far as he can. If I can increase this balance more by giving another man a good experience, at the cost of foregoing a good experience or suffering a bad experience myself, than I can by any other means, it is my duty to do so. If I can increase this balance more by enjoying a good experience myself, at the cost of depriving another man of a good experience or giving him a bad experience, than I can by any other means, it is my duty to do so.

Ethical egoism is the doctrine that each man has a predominant obligation towards himself, as such. Ethical altruism is the doctrine that each man has a predominant obligation towards others, as such. The extreme form of ethical egoism would hold that each man has an obligation *only* towards himself. The extreme form of ethical altruism would hold that each man has an obligation *only* towards others. According to the former extreme, each man's only duty is to develop *his own* nature and dispositions to the utmost, and to give *himself* the most favourable balance possible of good over bad experiences. He will be concerned with the development and the experiences of other persons only in so far as these may affect, favourably or unfavourably, his own development and his own experiences. The extreme form of ethical altruism would hold that each man's only duty is to develop to the utmost the nature and dispositions of all other men whom he can affect, and to give them the most favourable balance possible of good over bad experiences. He will be concerned with his own development and his own experiences only in so far as these may affect, favourably or unfavourably, the development and the experiences of other persons.

Now the first point to notice is that there is nothing *self-contradictory* in either ethical egoism or ethical altruism, even in their extreme forms. I mention this because Moore professes to show in *Principia Ethica* (pp.

96 – 105) that ethical egoism is self contradictory.[1] He alleges that ethical egoism involves the absurdity that *each* man's good is the *sole* good, although each man's good is different from every other man's good. Really it involves nothing of the kind. Suppose that *A* is an ethical egoist. He can admit that, if a certain experience of his is good, a precisely similar experience of *B*'s would be also and equally good. But he will assert that his duty is not to produce good experiences as such, without regard to the question of who will have them. *A* has an obligation to produce good experiences in *himself*, and no direct obligation to produce such experiences in *B* or anyone else. *B* has an obligation to produce good experiences in himself, and no direct obligation to produce such experiences in *A* or anyone else. And *A* can admit this fact about *B*. This doctrine does not contradict itself in any way. What it contradicts is Sidgwick's second axiom about goodness, viz. that each of us is under a direct obligation simply to produce good experiences as such, without regard to whether they are to occur in himself or in another. Since this axiom is equivalent to neutralism, it is obvious that it will conflict with egoism. But this does not make egoism *self*-contradictory. And unless Sidgwick's axiom is self-evidently true, the inconsistency of egoism with it does not prove that egoism is false.

Similar remarks apply to any argument against ethical altruism on the lines of Moore's argument against ethical egoism. Suppose *A* is an ethical altruist. He can admit that, if a certain experience of *B*'s is good, a precisely similar experience of his own would be also and equally good. But he asserts that his duty is not to produce good experiences, as such, without regard to the question of who will have them. *A* has an obligation to produce good experiences in *B* and other people, and no direct obligation to produce such experiences in himself. *B* has an obligation to produce good experiences in *A* and other people, and no direct obligation to produce such experiences in himself. And *A* can admit this fact about *B*. This doctrine contradicts Sidgwick's second axiom about goodness, but it is in no way self-contradictory.

One way of putting the difference between neutralism and the other two theories is the following. Neutralism assumes that there is a certain *one* state of affairs, viz. maximum balance of good over evil experiences in the universe, at which *everyone* ought to aim as an *ultimate* end. Differences in the proximate ends of different people are to be justified only in so far as the one ultimate end is best secured in practice by people aiming, not directly at it, but at different proximate ends of a more limited kind. The other two theories deny that there is any *one* state of affairs at which *everyone* ought to aim as an ultimate end. There are as many ultimate ends as there are agents. On the egoistic theory the ultimate end at which *A* should aim is the maximum balance of good over evil among *A*'s experiences. The ultimate end at which *B*

1. G.E. Moore, *Principia Ethica* (Cambridge, 1903).

should aim is the maximum balance of good over evil among *B*'s experiences. And so on for *C, D*, etc. On the altruistic theory the ultimate end at which *A* should aim is the maximum balance of good over evil among the experiences of all *others than A*. The ultimate end at which *B* should aim is the maximum balance of good over evil among the experiences of all *others than B*. And so on for *C, D*, etc. On neither theory is there anything that can be called *the* ultimate end at which *everyone* ought to aim. The main difference between the two theories is that for egoism the various ultimate ends are mutually exclusive, whilst for altruism any two of them have a very large field in common. Now there is nothing self-contradictory in the doctrine that, corresponding to each different person, there is a different state of affairs at which he and only he ought to aim as an ultimate end. And there is nothing self-contradictory in the doctrine, which is entailed by this, that there is no one state of affairs at which everyone ought to aim as an ultimate end. Moore simply assumed that there must be something which is *the* ultimate end at which *everyone* ought to aim; showed that ethical egoism is inconsistent with that assumption, and then accused ethical egoism of being *self*-contradictory.

Moore now admits (*Philosophy of G.E.Moore*, p. 613) that the argument in *Principia Ethica* was extremely obscure and confused.[1]. But he says that what he was trying to show was that, if ethical neutralism were true, it would follow that ethical egoism is, not merely false, but *self-contradictory*. He produces a new argument to prove this. I find the argument rather hard to follow, and it seems to me to involve a somewhat subtle logical fallacy. But in a certain sense I should accept the conclusion. Sometimes "self-contradictory" is used to mean "necessarily false". Now any proposition *Q* which is logically inconsistent with another proposition *P*, which is not merely true but *necessarily* true, is not merely false but *necessarily* false. Now if ethical neutralism is true at all, it is presumably axiomatic and self-evident, i.e. necessarily true. And, as we have seen, ethical egoism is logically inconsistent with it. Therefore, if neutralism were true, it would follow that egoism is not merely false but necessarily false, i.e. self-contradictory in one common usage of that phrase. It should be noted, however, that a precisely similar argument could be used to show that, if ethical egoism were true, it would follow that ethical neutralism is not merely false but necessarily false, i.e. self-contradictory in the sense already explained.

Granted that even the extreme forms of ethical egoism and ethical altruism are self-consistent, is there any reason to accept or reject either of them?
(1) If ethical neutralism were true they must both be rejected. Now the following argument can be produced in favour of ethical neutralism. On any theory except this it would sometimes be right for a person to do an act which will obviously produce less good or more evil than some other alternative act

1. P.A. Schilpp (ed.), *The Philosophy of G.E. Moore* (Evanston and Chicago, 1942).

which is open to him at the time. E.g. it is often the case that *A* could either (i) do an act which would add something to his own well-being at the cost of diminishing *B*'s by a certain amount, or (ii) do another act which would increase his own well-being *rather less* at the cost of diminishing *B*'s *very much less*. Plainly *A* would in general be producing more good by doing the latter act than by doing the former. But, if ethical egoism be true, it would be his duty to do the former and avoid the latter. Again, it is often the case that *A* could either (i) do an act which would add something to *B*'s well-being at the cost of diminishing his own by a certain amount, or (ii) do another act which would increase *B*'s well-being *rather less* and diminish his own *very much less*. Plainly *A* would in general be producing more good by doing the latter than by doing the former. But, if ethical altruism be true, the first act would be right and the second would be wrong. I think it is clear then that ethical neutralism is the only one of the three types of theory which can be combined with the doctrine that the right act will always coincide with the *optimific* act. Since utilitarians hold the latter view, they ought to hold the former; and so Sidgwick was right, as a utilitarian, to lay down an axiom which is equivalent to neutralism. I think it might be possible to combine ethical altruism with the doctrine that the right acts always coincides with the *optimising* act. Suppose that *A* can either (i) do an act which will increase *B*'s well-being to some extent at the cost of considerably diminishing his own, or (ii) do an act which would add rather less to *B*'s well-being and diminish his own very much less. Suppose that an act of self-sacrifice has, as such, a certain amount of moral goodness. Then it might be that the *direct* addition which the former act makes to the total goodness in the universe, as an act of self-sacrifice, more than counterbalances the *consequential* diminution which it causes by decressing the agent's own well-being. So the altruistic act might be the optimising act even when it is not the optimific act. Now commonsense does attach considerable positive value to acts of self-sacrifice as such. It is therefore conceivable that the right act, on the extreme altruistic view, might always coincide with the optimising act. But it is not necessary that it should, and it seems very unlikely that it always would. For it seems easy to conceive cases where the most altruistic act possible would increase the well-being of others very slightly and diminish that of the agent very much, whilst some other possible act would increase the well-being of others only a little less and would positively increase that of the agent. In such a situation it is most unlikely that the most altruistic act would be the optimising act, even when its direct contribution to the goodness in the universe, as an act of self-sacrifice, was taken into account.

It is quite plain that no attempt on these lines to reconcile ethical *egoism* with the doctrine that the right act must coincide with the optimising act would be at all plausible. For commonsense attaches no positive value to an

act of sacrificing the welfare of others for one's own benefit, as such. Therefore, when what would be the right act on the extreme egoistic view fails to coincide with the *optimific* act, it is impossible that it should coincide with the *optimising* act.

The upshot of the matter is this. Anyone who finds it self-evident that the right act in any conceivable situation *must* coincide with the optimific act, or that it *must* coincide with the optimising act, could safely reject both ethical egoism and altruism. He would have to accept ethical neutralism as the only principle of distribution which is compatible with his axiom.

(2) Let us now consider ethical egoism and altruism directly. The following remarks are worth making:

(i) Sidgwick, who was an exceptionally clear-headed and honest man, was in the uncomfortable position of finding *both* ethical neutralism and a form of ethical egoism self-evident, and seeing that they are inconsistent with each other. I think that what seemed to Sidgwick self-evident in ethical egoism is well put in a famous sentence of Butler, though Butler states it as a concession for the sake of argument and does not say that he accepts it himself. "...Though virtue...does indeed consist in affection to and pursuit of what is right and good as such, yet...when we sit down in a cool hour we can neither justify to ourselves this or any other pursuit till we are convinced that it will be for our happiness or at least not contrary to it."

I think that this needs a certain amount of clarification. Neither the phrase "justify to oneself" nor "be for one's happiness or at least not contrary to it" is perfectly clear. I suggest as a first amendment: "If a person reflects calmly, he cannot regard any act of his as reasonable unless he is convinced that it will either make him more happy or less unhappy, or at any rate will not make him less happy or more unhappy, than if he had not done it". Now the amended statement still contains one ambiguous word which I have intentionally introduced. That is the word "reasonable" as applied to actions. I think that this sometimes means "right, whether prudent or not", and sometimes "prudent, whether right or not". If "reasonable" be interpreted as "prudent", the statement is little more than a tautology; and it does not conflict with ethical neutralism, which is about what is right as distinct from what is prudent. If "reasonable" be interpreted as "right", the statement is not a tautology and it does conflict with ethical neutralism. For it practically amounts to saying that no act can be right unless it is prudent. Let us then substitute "right" for "reasonable" and consider the statement in the following form: "An act cannot be right unless it will either make the agent more happy or less unhappy, or at any rate not make him less happy or more unhappy, than if he had not done it."

I cannot find the least trace of self-evidence in this, and it is plainly in conflict with many moral judgments of common sense. Such an act is often held to be highly praiseworthy if done for some end which is considered to be

valuable in itself or if done for the sake of persons to whom the agent stands in certain special relationships. Even when we are not prepared to say that such an act is a duty, this is often not because we think it wrong but because we think that a person who does it is doing something which is creditable but is more than the maximum which duty demands. But there are plenty of cases where one would say that it is not only right but a duty to do such an act. E.g. this might be said of certain acts of this kind done by a mother for her child or by a son or daughter for an aged and infirm parent.

The principle would be much more plausible if it were stated, not in terms of happiness or even other forms of desirable experiences, but in terms of improvement or injury to the agent's personality. Suppose we substitute the following: "An act cannot be right unless it will make the agent a better person or prevent him from becoming a worse one, or at any rate will not make him a worse one or prevent him from becoming a better one, than if he had not done it." This is more plausible; but is it self-evident? There are two comments to be made.

(a) It is commonly held to be right and indeed admirable for a person deliberately to sacrifice his life if certain very valuable results for others can be secured in that way and in no other. Cf., e.g., the case of an officer deciding to blow up a certain bridge, where he will undoubtedly perish in the explosion but may save his country from invasion. It is even held to be a duty for a person to sacrifice his life if he stands in certain relationships to others, quite regardless of whether the results will be good or bad. E.g. it is held to be the duty of the captain and crew of a sinking ship to sacrifice their lives, if necessary, in order to save the women and children, quite regardless of calculations about the relative values of the lives and personalities of the two parties. Such cases might be covered by introducing the qualification "if he survives" after the word "agent".

(b) The phrase "better person" and "worse person" are ambiguous. They may mean better or worse *morally*, or better or worse in a sense which need not include specifically moral qualities. One becomes a better person in this not specifically moral sense if one's table-manners, one's golf-handicap, or one's powers of appreciating or playing classical music are improved. Now I do not think that it is at all obvious that an act is never right if it makes the agent a worse person or prevents him becoming a better person in a non-moral sense. Common sense regards it as always regrettable but often right and sometimes a duty for an agent to do an act which will involve cramping his personality and foregoing many possible and desirable developments of it. Any intelligent person who decides to devote his life to living in the slums and working for the poor inevitably does this, and we do not regard all such acts as wrong.

The case is strongest if the act will hamper a person's *moral* improvement

or cause a positive deterioration in his *moral* character. I think that common sense would be very uncomfortable in saying that such an act of self-sacrifice could ever be right. Yet it is difficult to be sure that some acts which common sense approves, or at any rate does not condemn, do not involve such consequences to the agent. A daughter who gives up her life to tending a peevish and selfish invalid mother, instead of marrying and having children, certainly foregoes many possibilities of *moral* development and is very likely to develop certain *moral* defects. No doubt her moral character will be improved in some directions, but it seems very doubtful whether on the whole the moral gain outweighs the moral loss and damage. Yet common sense hesitates to say that such an act is wrong. (ii) Even if the principle, in this very attenuated form, be accepted, it can hardly still be called "ethical egoism". It would best be described as an unconditional limitation on permissible self-sacrifice. One naturally asks whether there is any similar limitation on the sacrifices which it is permissible to impose on *others*, which could plausibly be held to be unconditional and self-evident. There are several remarks to be made on this.

(a) Common sense regards it as permissible for an individual or a community to sacrifice the *life* of a person under certain circumstances. An individual may do it if he is attacked and has reason to believe that he cannot save himself from death or serious injury at any less cost. A soldier not only may do it, but is under an obligation to do it, to a member of an opposing army who refuses to surrender. A community may do it, through its authorised agent, to a person who has been convicted of murder and sentenced to death by due process of law; and it is not merely permissible but a duty for the executioner to carry out the sentence.

(b) Common sense holds that it may be right for *A* to sacrifice *B*'s life when it would be wrong for *B* to sacrifice his own life. Thus, e.g. it is right for the executioner to take the life of the condemned murderer, but it is held to be wrong for the condemned murderer to commit suicide. On the other hand common sense holds that it may be right and even praiseworthy for a person voluntarily to make sacrifices, which it would be wrong for anyone else to impose on him. E.g. a medical research-worker with no one dependent on him would be admired if he voluntarily subjected himself to some process of treatment which might injure him permanently or kill him but which might lead to a valuable discovery. But it would be thought monstrously wrong to subject anyone against his will, or even, I think, to try to persuade him to be subjected, to such a process of treatment.

(c) Kant enunciated the principle that it is always wrong to treat a person as a mere means and always one's duty to treat him as an end.[1] This principle is

1. *Fundamental Principles of the Metaphysic of Morals.* Eng. trans. in T.K. Abbott, *Kant's Critique of Practical Reason and Other Works on the Theory of Ethics* (London, 1873), 6th ed. (London, 1909). A more recent translation is *The Moral Law or Kant's Groundwork of the Metaphysic of Morals*, trans. H.J. Paton (London, n.d.).

very vague. I think it can be interpreted in such a way that no one would be inclined to quarrel with it; but, when so interpreted, it does not give one very definite guidance. The minimal interpretation is this. It is always wrong to regard a *person* as if he were a mere animal and still more as if he were a mere inanimate object. For a person is a being who not only has sensations which may be painful and impulses which may be thwarted, like an animal. He also has the power of rational cognition, the power of reflexive cognition, ideas of right and wrong, good and evil, and all the emotional and conative peculiarities which depend upon these facts. In considering how to treat a person it can never be right to ignore these features which distinguish him from an animal and from an inanimate thing. When thus interpreted the principle is obviously true and it is no doubt highly important. But it does not follow that it is never right, when one *has* taken into account the features which distinguish a person from an animal or a thing, to treat him in certain respects as if he were an animal or a thing. E.g. it is not certain that it is never right to compel a person to do what he believes to be wrong or restrain him from doing what he believes to be right. For, although he is a person, he is not the only person; and there may be situations in which unless you treat a certain person as a dangerous animal he will infringe the rights and liberties and conscience of many other persons.

In view of the facts that common sense approves of capital punishment for murderers, disapproves of suicide, and admires voluntary sacrifices which it thinks it wrong to impose, it would be extremely difficult to formulate any unconditional principle of limitation on the sacrifices which it is permissible to impose on others. I suspect that the principle would have to contain so many qualifications that it would not be plausible to claim that it was self-evident.

(3) I will now make some remarks on the attitude of common sense towards pure egoism, pure altruism, and neutralism.

(i) Common sense would reject pure ethical egoism out of hand as grossly immoral. It is, I think, doubtful whether anyone would accept *ethical* egoism unless, like Spinoza, he had already accepted *psychological* egoism. If a person is persuaded that it is psychologically impossible for anyone to act non-egoistically, he will have to hold that each man's duties are confined within the sphere which that psychological impossibility marks out. But we have seen that there is no reason to accept psychological egoism.

(ii) The attitude of common sense, in countries where there is a Christian tradition, towards pure ethical altruism is different. It would be inclined to described the doctrine as quixotic or impracticable but hardly as immoral. There is a sound practical reason for this attitude. We realise that most people are far more likely to err on the egoistic than the altruistic side; that in a world

where so many people are too egoistic it is undesirable to discourage altruism; and that there is something heroic in the power to sacrifice one's well-being for the good of others. We therefore hesitate to condemn publicly even exhibitions of altruism which we privately regard as excessive.

(iii) Although common sense rejects pure egoism and does not really accept pure altruism, I do not think that it is prepared to accept neutralism without a struggle. It would regard neutralism as in some directions immorally selfish and in other directions as immorally indiscriminate. It undoubtedly holds that each of us has a more urgent obligation to benefit persons who are specially related to him in certain ways, e.g. his parents, children, fellow-countrymen, benefactors, etc., than to benefit others who are not so related to him. And it would hold that the special urgency of these obligations is founded *directly* on these special relations.

(iv) The ideal of common sense is therefore neither pure egoism, nor pure altruism, nor neutralism. I think it may best be described as "self-referential altruism". I will now explain what I mean by this. Each of us is born as a member of a certain family, a citizen of a certain country, and so on. In the course of his life he voluntarily or involuntarily becomes a member of many other social groups, e.g. a school, a college, a church, a trades-union, etc. Also he gets into special relations of love, friendship, gratitude, etc. with certain individuals who are not blood relations of his. Now the view of common sense is roughly as follows.

(a) Each of us has a certain obligation to himself as such. I do not think that common sense holds that a person is under *any* obligation to make himself *happy*, i.e. to "give himself a good time". Possibly that is because most people have so strong a natural tendency to aim at prolonging and getting experiences which they like and cutting short and avoiding those which they dislike. The obligation to develop one's own powers and capacities to the utmost and to organise one's various dispositions with a good all round personality is felt to be strong. This kind of action often goes very much against the grain, since it may conflict with natural laziness and a natural tendency to aim at the easier and more passive kind of good experience. The obligation to make others happy and to prevent them from being unhappy varies in urgency according to the nature of the relation of these others to oneself. It is weakest when the others stand in no relation to oneself except that of being fellow sentient beings. It is strongest when the others are one's parents, or one's children, or non-relations whom one loves and by whom one is loved, or persons from whom one has received special benefits. My obligation towards A is more urgent than that towards B if it would be right for me to aim at the well-being of A before considering that of B, and only to begin to consider that of B after I have secured a certain minimum for A. The greater this minimum is, the greater is the relative urgency of my obligation towards A as

compared with my obligation toward *B*.

Now common sense holds that it is my duty to be prepared to sacrifice a considerable amount of my own well-being to secure a quite moderate addition to the net well-being of my parents or my children or my benefactors, if this is the only way in which I can secure it. But it is not my duty to sacrifice much of my own well-being in order to secure even a considerable addition to the net well-being of other persons who stand in no specially intimate relation to me.

The obligation to develop one's own powers and capabilities to the utmost and to organise one's dispositions into a good personality is held to be strong; whilst the obligation to make oneself happy, if it exists at all, is extremely weak. Hence it is felt to be doubtful how far one ought to sacrifice self-development and self-culture in order to add to the well-being of others. It is only when the claim is very urgent, as in the case of aged and infirm parents on a son or daughter, that common sense approves of this kind of self-sacrifice; and even then it feels considerable hesitation. Apart from such cases, I think that common sense is rather embarrassed. It realises that it is a good thing on the whole that a certain proportion of people should voluntarily forego the development of a great many aspects of their personality in order to live in the slums and add to the well-being of other persons, who have no urgent claims on them. But, whilst it admires the people who make the sacrifice, it regrets the waste of talent; and it is relieved to think that there is no great danger of many gifted persons following their example. On the whole it favours a kind of ethical "division of labour". A certain minimum of self-sacrifice and of self-culture is demanded of everyone; but, when that minimum has been reached, common sense approves of certain persons specialising in self-culture and others in beneficient self-sacrifice.

Lastly, common sense considers that each of us has direct obligations to certain groups of persons, considered as collective wholes, of which he is a member. The most obvious case is one's nation, considered as a collective whole. It is held that an Englishman, as such, is under an obligation in certain circumstances to sacrifice his happiness, his development, and his life for England and is under no such obligation to Germany; that a German is under an obligation in similar circumstances to make a similar sacrifice for Germany and is under no such obligation to England and so on. It must be noticed that Germans, as well as Englishmen, hold that Englishmen have this peculiar obligation to England; and that Englishmen, as well as Germans, hold that Germans have this peculiar obligation to Germany. And this is clearly recognised by both parties even when the two nations are at war with each other. The fact that an Englishman considers that a German should sacrifice himself for Germany, even when his doing so is detrimental to England, and that a German considers that an Englishman should sacrifice

himself for England, even when his doing so is detrimental to Germany, is of considerable importance. It certainly suggests that we are concerned with a genuine and objective, though limited, obligation; and not with a mere psychological prejudice in favour of one's own group. So far as I can see, opinion has varied from time to time and place to place as to what *kind* of group has the most urgent obligation on its members. At present, among most Western people, the nation is put in this supreme position. Among the Greeks and the Romans it was the city. In mediaeval times the supreme obligation was generally to a lord and not to a group. And it may be that in the near future it will be to a class rather than to a nation or a lord. But it has always been held that there is *some* person or group for which every person who stood in a special relation to it, and only such persons, was bound in certain circumstances to sacrifice his happiness, his chances of culture and development, and his life.

I said that common sense accepts a kind of *self-referential altruism*. My meaning will now be clear. Common sense is altruistic in so far as it considers that each of us is frequently under an obligation to sacrifice his own happiness, and sometimes to sacrifice the development of his personality and even to give up his life for the benefit of other persons or institutions, even when it is uncertain whether more good will be produced by doing so than by not doing so. It tends to admire those acts, as such, even when it regrets the necessity for them and even when it thinks that on the whole they had better not have been done. It has no such admiration for the act of making oneself happy, as such, even when it does no harm to others. It admires acts directed towards the development and improvement of one's personality, as such; though its admiration is not very strong unless they are done in face of great external obstacles (e.g. poverty) or great internal handicaps (e.g. blindness). On the other hand, the altruism of which common sense approves is always limited in scope. It does not hold that any of us has any equally strong obligation to benefit all those whom he could equally affect by his actions. It holds that each of us has specially urgent obligations to benefit certain persons and groups of persons who stand in special relations towards *himself*. And it holds that these special relationships are the ultimate and sufficient ground of these specially urgent claims on his beneficence. According to it, each person may be regarded as a centre of a number of concentric circles. The persons and groups to whom he has the most urgent obligations form the innermost circle. Then comes a circle of persons and groups to whom his obligations are moderately urgent. Finally there is the outermost circle of persons (and animals) whose only claim on his beneficence is what we call the "claim of common humanity". This is what I mean by saying that the altruism which common sense accepts is "self-referential".

(4) If this is a fair account of the beliefs of common sense, what line could a

person take who found ethical neutralism self-evident? And what line could a person take who found it self-evident that the right act must coincide with the optimific or with the optimising act, and was therefore committed to neutralism at the next move? The problem is the same for both of them. He would have to do three things:

(i) He would have to hold that common sense is mistaken in thinking that these specially urgent claims on one's beneficence are founded *directly* on these special relations. (ii) He would have to show that all these special obligations, so far as they are valid at all, are derivable from the one fundamental obligation to maximise the balance of good over evil among all contemporary and subsequent conscious beings as a whole. He will try to do this by pointing out that each of us is limited in his resources, in his powers of helping or harming others, in the range of his natural sympathies and affections, and in his knowledge of the needs of others. He will argue that, in consequence of this, the maximum balance of good over evil among conscious beings as a whole is most likely to be secured if people do not aim directly at it. It is most likely to be secured if each aims primarily at the maximum balance of good over evil in the members of a limited group consisting of himself and those who stand in more or less intimate relation to him. The best that the neutralist could hope to achieve on these lines would be to reach a system of *derived* obligations which agreed roughly, both in scope and in relative urgency, with that system of obligations which common sense mistakenly thinks to be founded *directly* upon various special relationships. In so far as this result was attained he might claim to accept in outline the same set of obligations as common sense does; to correct common sense morality in matters of detail; and to substitute a single coherent system of obligations, deduced from a single self-evident ethical principle and a number of admitted psychological facts, for a mere heap of unrelated obligations. (iii) To complete his case he would have to try to explain, by reference to admitted psychological facts and plausible historical hypotheses, how common sense came to make the fundamental mistake which, according to him, it does make. For common sense rejects the neutralistic principle, which he finds self-evident; and it regards as *ultimate* these special obligations of an individual towards certain persons and groups which he regards as derivative.

How could he set about fulfilling this third task? It seems to me that it might be attempted on the lines which I have already suggested for defending utilitarianism. Any society in which each member was prepared to make sacrifices for the benefit of the group as a collective whole would be more likely to flourish and persist than one whose members were not prepared to make such sacrifices. Now egoistic and anti-social motives are extremely strong in everyone. Suppose, then, that there were a society in which, no matter by what means, there had arisen a strong additional motive (however

mistaken and superstitious) in support of self-sacrifice of the member for the sake of the group. Suppose that this motive were conveyed from one generation to another by example and precept and were supported by the sanctions of social praise and blame. Such a society would be likely to flourish and to overcome other societies in which no such additional motive existed. So its ways of thinking on these subjects and its sentiments of approval and disapproval would tend to spread. They would be propagated directly by conquest, and indirectly through the prestige which the success of this society would give it in the eyes of others.

Suppose next that there were a society in which, no matter how, a strong additional motive for *unlimited* altruism had arisen and had been propagated from one generation to another. A society in which each member was prepared to sacrifice himself just as much for other societies and their members as for his own society and its members would be most unlikely to persist and flourish. Therefore such a society would be likely to go under in conflict with one in which a more restricted *self-referential* altruism was approved and practised. Now suppose a long period of conflict between societies of the various types which I have imagined. It seems likely that the societies which would still be existing and flourishing at the end of such a period would be those in which there had somehow arisen, in the remote past, a strong pro-emotion towards altruism confined within the society and a strong anti-emotion towards extending it beyond those limits. And these are exactly the kind of societies which we do in fact find existing and flourishing in historical times.

It seems therefore that, even if neutralism be true and be self-evident to the philosopher in his study, there are powerful causes which would tend to make certain forms of self-referential altruism *seem* to be true and self-evident to most unreflective persons at all times and even to reflective persons at most times. Therefore the fact that common sense rejects neutralism and accepts as self-evident certain forms of self-referential altruism is not a conclusive objection to the *truth* or even to the *necessary* truth of neutralism.

1.42. Kant's theory

I shall now discuss a type of monistic theory of the grounds of moral obligation which is at the opposite extreme to utilitarianism, viz. Kant's theory. This theory is stated in terms of the notion of "ought" or "duty" rather than that of "right" and "wrong". Moreover Kant makes great use of the notion of what he calls "imperatives". He contrasts *moral* imperatives with others. So I shall begin with an independant discussion of the notion of "ought". And I shall then consider the notion of an "imperative", and discuss the connexion between the two.

1.421. Deontic sentences

I shall call any sentence in which the word "ought" or any obviously equivalent phrase such as "being a duty", "being under an obligation", etc. occurs as the principal verb a *deontic* sentence. Examples are: "I ought to go to the dentist", "You ought not to eat peas with a knife", "He ought to make an allowance to his old nurse", "Persons who have borrowed money ought to repay it at the agreed date", "There ought to be laws against cruelty to animals" and "A fountain-pen ought not to be constantly making blots".

(1) The first point to notice is that these sentences fall into two main classes. Some of them assert of a *person* that he ought or ought not to *do* something. Others assert of a conceivable *state of affairs* that it ought to *be*, or of an actual *state of affairs* that it ought not to *be*. (The sentence about fountain pens falls between the two; for it asserts of a class of *inanimate objects* that they ought not to *behave* in a certain way.) We can thus divide deontic sentences into "ought-to-do" sentences and "ought-to-be" sentences.

(2) In each class of deontic sentences we can distinguish between those in which "ought" occurs in a *specifically moral* sense and those in which it occurs in some other sense. In the sentences "He ought to make an allowance to his old nurse" and "There ought to be laws against cruelty to animals" the word "ought" is used in a specifically moral sense. In the sentences "You ought not to eat peas with a knife" and "A fountain pen ought not to be constantly making blots" it is used in a sense or senses which are not specifically moral.

1.4211. "Ought-to-do" sentences about persons. For the present we will confine ourselves to *ought-to-do* sentences in which the grammatical subject is a *person* or *class of persons*.

(1) The word "ought" in English has certain grammatical peculiarities. (i) It cannot, e.g., be used in the future tense. This is of no philosophical significance, as can be seen by substituting the phrase "to be under an obligation". One can say, e.g., "When you become a parent you *will be* under an obligation to support your children". (ii) When used of the past the words "ought" and "ought not" have certain linguistic implications which are also of no philosophical significance. If you say that *X ought* to have done so and so, there is a strong suggestion that he omitted to do this and did something else. Yet obviously people sometimes have done the actions which they ought to have done. Similarly, if you say that *X ought not* to have done so-and-so, there is a strong suggestion that he *did* that very action. Yet obviously people sometimes have refrained from doing the actions which they ought not to have done. All these irrelevant suggestions are avoided by substituting "to be under an obligation". We can say that *X* was under an obligation to do so-and-so on a certain past occasion without suggestion that he failed to do it;

and we can say that *X* was under an obligation not to do so-and-so on a certain past occasion without suggesting that he did it.

(2) In "ought-to-do" sentences the grammatical complement to the word "ought" or "ought not" is a name or description of what I will call an *agibile*, i.e. a possible act of a certain kind. The *agibile* is supposed to have been or to be now or to be going to be completely in control of the agent's will at the time referred to in the sentence. This means that it is assumed that it would have been or will be enacted if and only if the agent had decided or shall decide to enact it and had set himself or shall set himself to carry out his decision.

(3) Kant says, truly I think, that we use "ought to do" only in reference to agents in whom we conceive there to be an actual or possible conflict of motives in regard to the *agibile* in question. It always suggests that the agent may have to force himself to enact a certain *agibile*, and that, unless he makes a special effort, he will do nothing or will enact some other alternative which is in some way easier or more attractive to him. This brings out the difference between "ought to do" even in its most strictly moral sense, and "morally right". No doubt there is a close connexion between the two. On one view, what a person morally ought to do in any situation is what would *be* morally right for such a person to do in such a situation. On another view, what a person morally ought to do is what he *believes* would be morally right for him to do in the situation as *he believes it to be*. But, on either alternative, it is one thing to say that he morally ought to do so-and-so, and another thing to say that so-and-so would be morally right for him to do. In making the deontic statement we imply that he has a desire to do what is right as such, that he has other desires or inclinations which may conflict with this, and that he may need a special effort in order to do what he believes to be right.

This point may be brought out (as Kant remarks) by noticing that, while we should say that God always acts rightly, we should hesitate to apply the word "ought" to him. For we assume that in God there would be no motives or inclinations which could possibly conflict with the desire to act rightly.

This reference to an actual or possible conflict extends to cases where little if anything specifically moral is involved in "ought-to-do". Take, e.g., the case of a person who has a decayed tooth which occasionally gives him severe pain. He may consider the question simply from the point of view of his own interest in the most narrowly hedonistic sense. He may be quite convinced that it would pay him very well from that point of view to go to the dentist, and perhaps suffer a short bout of severe pain there in the immediate future, in order to secure permanent freedom thereafter from toothache in that tooth. Even so, it is very likely that he will have a considerable internal struggle and will not go to the dentist unless he takes himself in hand and forces himself to do so. A man in that position would be very likely to say to

himself "I ought to go to the dentist", and a friend would be very likely to say to him "You ought to go to the dentist".

I think that there is one circumstance which gives a moral tinge even to such statements as "He ought to go to the dentist". We approve, in ourselves and in others, the capacity and the act of overcoming one's own laziness, fear of pain or unpopularity, and desire for immediate passive satisfactions, in order to carry out one's more fear-reaching desires and purposes. For the possession and the exercize of that power is a necessary condition of *all* serious achievement, whether good or evil. We are thus inclined to feel and to express a kind of qualified moral approval of it even when it issues in acts which are morally indifferent, e.g. acts of far-sighted prudence which cost an effort. We do so, even when it issues in acts which we morally condemn. This is probably what lies at the back of the paradoxical imperative: *Si peccas, pecca fortiter.*

Sentences in which "ought to do" occurs in a specifically moral sense coincide roughly with what Kant called "*categorical* imperatives". Sentences in which it occurs in a not specifically moral sense, though it may have the moral tinge noted above, coincide roughly with what he called "*hypothetical* imperatives" and with what he called "imperatives *of skill*". If we take the two latter together, we may describe the context in which they occur as follows. (i) It is assumed that the agent has a certain fairly strong and persistent desire. This may be either (i) peculiar to himself or to his present circumstances or (ii) such that similar desires exist in most men at most times. An example of the former would be that of a man who had decided to poison his wife. An example of the latter would be the desire for good health, long-life, and prosperity. The former case corresponds to "imperatives of skill" and the latter to "hypothetical imperatives". (2) It is assumed that the act which it is said that he ought to do is either absolutely essential, or at any rate much the most effective means in his power, to carry out his purpose in the situation in which he is placed. (3) It is *not* assumed that the decision to seek this end is one that he morally ought to have made, or that the end is a morally good one. (4) It is assumed that the subject is liable, either through ignorance or through laziness or lack of resolution, not to act in the way required. It is under such circumstances that one might say of a man "He ought to give his wife a dose of arsenic in her tea" or "He ought to smoke less and take more exercise".

I shall at present avoid all references to the phrase "imperatives" and shall proceed to deal with these distinctions in my own way. I shall say that deontic sentences in which "ought to do" occurs express "obligations of activity". I shall then divide these into two classes according as the obligation which they express is *ultimate* or *derivative* for a given individual. An obligation of activity is ultimate for a person if it appears to him on inspection to be self-

evident. The following would be two plausible examples "A person ought to try to produce as much good and as little evil as he can" and "A person ought to proportion the strength of his convictions to the weight of the evidence available to him". The former is specifically moral, the latter perhaps is not. So far as I can see, no reason can be given for them, and they do not seem to need a reason. They seem to me to be evident on inspection. An obligation of activity is derivative for a person when it does not seem evident to him on inspection, but seems to him to require and to be capable of being given a reason. I will now consider the two kinds of obligation in turn.

1.42111. Ultimate obligations of activity. Ultimate obligations of activity can be classified on two independent principles, viz. (i) the nature of the activity with which they are concerned, and (ii) the nature of the obligation asserted. Human activities may be divided into theoretical and practical, and so we have, corresponding to them, *obligations of theoretical activity* and *obligations of practical activity*. Suppose I were to say "You ought to accept the conclusions which follow from the premisses which you accept" or "You ought not to believe mutually inconsistent propositions" or "You ought to proportion the strength of your convictions to the weight of the evidence available to you". These would express obligations, though not perhaps specifically moral ones, which you are under as a *thinking* being engaged in the *theoretical* activity of exercising your intellect. These are obligations of theoretical activity. Suppose I were to say "You ought to try to produce as much good and as little evil as you can" or "You ought to keep your promises" or "You ought not to tell lies". These would express obligations which you are under as a being engaged in the *practical* business of co-operating or struggling with others, and affecting yourself and other persons and things by your actions. These are examples of obligations of *practical* activity. There is a fairly close analogy between the two kinds of obligation. We have seen that the obligations of practical activity presuppose an agent who has the desire to do right, but also has other desires which may conflict with it and may induce him to enact one of the wrong agibilia instead of the right one. Similarly the obligations of theoretical activity presuppose a thinker who has the desire to think reasonably, but also has prejudices and lazinesses which may conflict with it and may induce him to accept one of the propositions under consideration which he is not logically justified in accepting, or to believe one of these propositions more strongly or less strongly than the available evidence logically justifies him in doing. In each case a special effort needs to be made and kept up, if the agent is to do as he ought. The obligations of practical activity presuppose that it is, in some sense, within the agent's power to enact the right agibile, in spite of the inclinations which conflict with his desire to do what is right. Similarly, the

theoretical obligations presuppose that it is, in a similar sense, within the agent's power to suspend judgment when the evidence is inadequate, in spite of his desire to make up his mind. And they presuppose that it is in some sense within the agent's power to proportion the strength of his convictions to the weight of the available evidence, in spite of his prejudices and his intellectual laziness.

We must now consider the classification of ultimate obligations in respect of their own intrinsic nature. In this respect obligations of activity can be subdivided into two fundamentally different classes, viz. *teleological* and *non-teleological*. Suppose that the deontic sentence "You ought to produce as much good and as little evil as you can" expresses an ultimate obligation. Then it would be an example of an ultimate obligation of the *teleological* kind. The peculiarity of such an obligation is this. It contemplates a certain *possible* state of affairs, and it contemplates an action as contributing to bring it about or to avert it. Or it contemplates a certain *actual* state of affairs, and it contemplates an action as contributing to prolong it or to cut it short or to modify it in certain ways. Again, it considers this possible or actual state of affairs and these possible modifications of it primarily from a certain point of view, viz. its *value* or *disvalue*. In so far as it considers other features in it it considers these only as *good-making* or *bad-making* characteristics. The ground alleged for the obligation to do or avoid doing the act is simply that it will produce or prolong or improve a good state of affairs or will avert or cut short or improve a bad state of affairs. Next, suppose that the deontic sentence "You ought to answer truly when asked a question" expressed an ultimate obligation. Then it would be an example of an obligation of the non-teleological kind. This obligation might of course be interpreted in two different ways. (i) "You ought to give what you believe to be the *true answer* whether or not you believe it will *produce a true belief* in the mind of the questioner". Or (ii) "You ought to give the answer which you think will *produce a true belief* in the mind of the questioner, whether or not you believe that answer *to be true*". On the first alternative the obligation is not teleological, because it is not based on the consequences of the action at all, but on its qualities or its non-causal relations to the situation of being asked a question. On the second alternative the act *is* considered in reference to its consequences, but the obligation is still not teleological. For the obligation is not based on the *goodness* of the consequences. All that is considered to be relevant is the *truth* of the belief produced in the questioner's mind; whether it is *good* or *bad* for him to have a true belief on the subject of his question is held to be irrelevant.

By combining the two principles of division which I have just explained we reach the following four-fold division of ultimate obligations of activity.

Ultimate obligations of activity

Theoretical		Practical	
Teleological	Non-teleological	Teleological	Non-teleological

I have already given an example of both kinds of practical obligation. The sentence "You ought to apportion the strenght of your convictions to the weight of the evidence available to you" expresses a non-teleological theoretical obligation, if it expresses an ultimate obligation. The sentence "You ought to try to increase the amount of your knowledge and true belief and to reduce the amount of your ignorance and false belief" expresses a teleological theoretical obligation, if it be assumed that knowledge and true belief have positive value and are to be sought for that reason.

Before passing on to derived obligations we must notice one other important distinction. An obligation of activity may be either *restricted* or *unrestricted* in its range of applications. A restricted obligation is concerned with a certain specific kind of situation, e.g. that of being asked a question. The deontic sentence here asserts that any person who is acting in response to such a situation ought to act in a certain specific way. An example would be "Whenever a person is asked a question he ought to return a true answer to it". An unrestricted obligation is supposed to apply equally in *any* situation in which a voluntary action is to be performed. An example would be "Whenever a person acts he ought to try to produce as much good and as little evil as he can". Another example would be "A person ought never to treat others in a way in which he would not be willing to be treated by others". The first of these expresses an unrestricted teleological obligation. The second would express an unrestricted non-teleological obligation. It is evident that all unrestricted obligations will be extremely abstract, and that, taken by themselves, they will give a person very little positive guidance as to what he ought to do in any particular situation.

1.42112. Derived obligations of activity. We can now turn to *derived* obligations. I think it is a true general principle that no obligation can be derived unless some other obligation is already presupposed. You cannot legitimately derive a deontic proposition from *nothing but* non-deontic ones, though you can and perhaps must use non-deontic propositions *in conjunction with* deontic ones in your derivation. I think it is also a true general principle that a *specifically moral* obligation can be legitimately derived only from premisses which include some deontic proposition which asserts a *specifically moral* obligation.

Let us begin with the derivation of specifically moral obligations. To

illustrate the most important types of derivation let us take as an example the specifically moral deontic proposition "A person ought never to give an answer which he believes to be false to a question which is put to him". Now this might appear to some people to be self-evident on inspection. For such persons it will count as an *ultimate* non-teleological obligation of limited range. But there are many people who are not in this position, and yet would accept it as a *derived* obligation. There seem to be two typical alternative possible ways in which it might be derived.

(1) Suppose that an individual finds self-evident the following unrestricted non-teleological deontic proposition, viz. "A person ought never to treat others in a way in which he would not be willing to be treated by others". Suppose further that he knows or believes that no one is willing to be told a lie in answer to a question. These two propositions together entail that a person ought never to give an answer which he believes to be false to a question which is put to him. On the two suppositions which I have made, this proposition would assert a *derived non-teleological obligation* of limited range.

(2) Let us next suppose that a person finds self-evident the following unrestricted teleological deontic proposition, viz. "In all one's actions we ought to try to produce as much good and as little evil as possible". Suppose, further, that he knows or believes that telling lies in answers to questions always produces less good or more evil in the long run than telling the truth or refusing to answer. These two propositions together entail that one ought never to give an answer which one believes to be false to a question which is put to one. On the two suppositions which I have now made, this proposition would assert a *derived teleological* obligation of limited range. Conversely one and the same person might accept both pairs of premisses. In that case one and the same sentence would express for him an obligation of limited range which could be derived both teleologically and non-teleologically.

Under the head of derived obligations of activity we can consider next what Kant called "hypothetical imperatives" and "imperatives of skill". The former presuppose a desire which is assumed to be common to practically all men at all times, and therefore does not need to be specifically mentioned, e.g. desire for good health, long life, happiness, etc. The latter presuppose a desire which is peculiar to a particular person or a particular situation, and therefore has to be specifically mentioned, e.g. a desire in Mr. Jones to kill his wife. Two things seem plain about these deontic sentences. One is that, if they express obligations at all, these are not specifically moral ones. The other is that they are in some way concerned with which I will call "obligations of *consistency*".

An obligation of consistency may be either practical or theoretical, and in neither case is it specifically moral. I would formulate the obligation of practical consistency as follows "A person who intends a certain end ought

either to cease intending it *or* to take the most efficient means open to him to attain it. He ought not *both* to go on intending it *and* to do acts which would make it impossible for him to attain it." I think it is important to formulate it in this way. For this makes it plain that the "ought" and "ought not" here is concerned with consistency or inconsistency between, on the one hand, continuing to intend a certain end, and, on the other, acting or failing to act in certain ways which are relevant to the attainment or non-attainment of that end.

Let us now consider the derivation of a hypothetical imperative in Kant's sense of the word. We can take as an example "A person ought to take exercise and not habitually to overeat". It is assumed as a factual premiss that taking exercise is a necessary condition for keeping in good health, and that habitual over-eating is a sufficient condition for failing to do so. From this factual premiss and the obligation of practical consistency we can infer the following proposition "A person who intends to keep in good health ought *either* to give up that intention *or* to take exercise, and he ought not *both* to go on intending to keep in good health *and* habitually to overeat". Now it is assumed that *all* men intend to keep in good health. On that assumption we can substitute the phrase "a person" for the phrase "a person who intends to keep in good health". It is further assumed that the intention to enjoy good health is a *standing* intention, which a person cannot or will not abandon. In that case the only way to be practically consistent is to take exercise and not to overeat. So we reach the conclusion that a person ought, *in order to be practically consistent*, to take exercise and not habitually overeat.

Let us next take what Kant would call an "imperative of skill". We will suppose that Mr. Jones intends to kill his wife, and that to give her a dose of arsenic in her tea is much the most efficient way open to him for securing that end. From these premisses and the obligation of practical consistency we can infer the proposition "Mr. Jones ought, in order to be practically consistent, *either* to give up his intention to kill his wife *or* give her a dose of arsenic". If he cannot or will not give up this intention, you can say that he ought, in order to be consistent, to give his wife a dose of arsenic. But it is important to add explicitly the qualification "in order to be consistent". For it is *only* in this sense that he "ought" to do this. In the moral and the legal senses of "ought" he ought *not* to do it. He ought morally to give up the intention.

It is sometimes alleged that what Kant calls hypothetical imperatives and imperatives of skill do not really express deontic propositions at all. It is alleged that they are simply equivalent to non-deontic sentences expressing causal propositions of a certain kind. Thus, e.g., it would be alleged that the sentence "You ought, unless you give up your intention to kill your wife, to give her a dose of arsenic" is simply equivalent to "The most efficient means available to you for carrying out your intention to kill your wife is to give her

a dose of arsenic". This seems to me to be a mistake, though I am quite willing to admit that such deontic sentences *may* sometimes be used to mean no more than this. But in general I think that the causal proposition is merely the *factual ground* for the derived deontic proposition. The latter has also a *deontic* ground, viz. the ultimate, though not specifically moral, obligation of practical consistency. I take it that Kant would have agreed at least with the negative part of my statement. For he calls such propositions "imperatives", and surely a mere causal proposition would not be an imperative in any sense of the word.

1.4212. "Ought-to-do" sentences about things. There is a sense of "ought-to-do" in which we apply it even to inanimate objects. It would be quite proper to say, e.g., "A car ought to get from London to Cambridge in less than three hours" or "A fountain-pen ought not to be constantly making blots". So far as I can see, what we mean primarily is this. A car which did habitually take more than three hours would be a poor specimen of car or else in a bad state of repair. Similar remarks apply *mutatis mutandis* to a fountain-pen which constantly makes blots. We are comparing the performance of a certain car or fountain-pen with the average standard of achievement of cars or fountain-pens respectively, in regard to its performance of its specific functions. We are certainly not suggesting that *this* car or *this* pen, in its present state of repair, could go faster or could avoid making blots. Sometimes when we make such judgments we are comparing a thing's performance, not with that of the *average* member of its species, but with that of a conceived *ideal* member. When "ought" is used in the present sense in may be called the "average-comparative" or the "ideal-comparative" ought. In this sense it applies almost exclusively to the performance of their characteristic functions, either by artificial objects, designed and constructed by human beings, or by animals or plants of definite species in which human beings are practically interested.

Now "ought" and "ought not", in this sense, can be applied to human voluntary actions. They can also be applied to men's non-voluntary actions and to their conations, emotions, and dispositions. But in these cases there is a further complication. A man, unlike a car or a pen, has the power of *cognition*. And, unlike a horse or a dog, he has the power of *reflexive* cognition and of moral appraisal. He can, and generally does, have an idea, more or less definite, of an average man and an ideal man. He can, and often does, compare his own performances with those of the average man or the ideal man, as conceived by him. Moreover, he will generally have a desire, more or less strong and persistent, to approximate to the ideal man and not to fall below the average man.

Now it is part of the notion of an ideal man that he would have a high ideal

of human nature and would desire strongly and persistently to approximate to his ideal. Obviously it is not part of the notion of an ideal car or an ideal horse that it would have a high ideal of cars or of horses, and a strong and persistent desire to live up to it. Suppose we say, e.g., that a man ought not to feel pleasure at the thought of another person's pain or disappointment. What we mean to assert is often the two following propositions. (1) That the average decent man does not do this; and that anyone who does so is, at any rate on that occasion and in that respect, falling below the average. (2) That a man who habitually has such feelings in such circumstances must either have a low ideal of human nature or a weak and unstable desire to live up to the ideal which he has, so that in this further respect he falls below the average. Neither of these judgments implies that a particular person, who felt a malicious emotion on a particular occasion, *could* then and there by any act of volition have had a different emotion or a higher ideal of human nature or a stronger and more persistent desire to live up to his ideal.

When we use "ought not" in such contexts we have often the following thought at the back of our minds. We know that a man's character and dispositions are to a large extent moulded by his own past choices. We believe that it is further modifiable in future by his later volitions, within limits which are unknown either to himself or to others. Now, when we use "ought" about the performance of a car, we often have at the back of our minds the thought of it as a product of human workmanship in accordance with human design. I suggest that we often have a somewhat similar thought at the back of our minds when we use "ought" and "ought not" about those manifestations of a man's character which are not dependent, as his deliberate actions are, on his *immediately* precedent volitions. But, for the reasons which I have given, the situation is enormously more complex in the case of a human being than in that of an irrational animal or an inanimate artefact. For in the human case the designer and the constructor and the product are the same person at various times and in various aspects. And the materials out of which he has built his present self, in accordance with his ideas of what he would wish to be, are his own innate dispositions, as modified by his own experiences and by his own past actions and failures to act.

1.4213. "Ought-to-be" sentences. The general formula for "ought-to-be" sentences is "So-and-so ought (or ought not) to exist" or "There ought (or ought not) to be so-and-so". We may take as examples "A state of affairs ought to exist in which the happiness experienced by each person is proportional to his moral goodness", and "There ought to be laws against cruelty to animals".

It is usual to predicate "ought to be" of a *possible* state of affairs with regard to which we either know that it is not realized or do not know whether

it is realized or not. It is usual to predicate "ought not to be" of a state of affairs which we know or believe to be *actual*. But I think that this grammatical usage is of no philosophical significance. One could certainly say "Such and such a state of affairs does exist, and it ought to do so". And we could certainly say "Such and such a state of affairs does not and ought not to exist".

Now it seems to me that, when we say "So and so ought to exist", the following conditions have to be fulfilled.
(1) We have a description of a certain state of affairs which we believe to be *possible*. We may either know that it is actual or know that it is not actual or be uncertain about its actuality. (2) We judge that any state of affairs answering to that description would be on the whole *good*, in some sense of that word. (3) For that reason we hold that anyone who had it in his power to contribute towards bringing such a state of affairs into existence (if it does not exist) or towards keeping it in existence (if it does exist) would have at least a *prima facie* duty to do so. This *prima facie* duty might of course be overridden by others which were more urgent. If that is so, it would seem that "ought-to-be" always involves a reference to "ought-to-do", on the part of some person, actual or imagined. If that reference were altogether excluded, I think we should confine ourselves to saying that such a state of affairs *would be* good if it did exist, and *is* good if it does exist.

1.422. Deontic sentences and imperatives
Kant gives the name "imperatives" to what is expressed by "ought-to-do" sentences about persons. If this is taken literally it would mean that such a sentence in the indicative expresses neither more nor less than what would be naturally expressed by a corresponding sentence in the imperative. E.g. the sentence "You ought not to eat peas with a knife" would express and convey exactly what is expressed and conveyed by the sentence "Don't eat peas with a knife!", i.e. an order issued by one person and received by another. Another possible view would be the following. A deontic sentence in the indicative does express and convey something which is not expressed or conveyed by any sentence in the imperative, viz. some kind of *knowledge* or *belief*, so that it can be significantly described as "true" or "false". But it *also* expresses and conveys a command, as an ordinary sentence in the imperative would do. And its specifically *deontic* character, which distinguishes it from other kinds of sentences in the indicative, e.g. "Peas tend to roll off a knife", is bound up with this imperative function. A third alternative would be a more cautious modification of the second. Instead of saying that a deontic sentence in the indicative derives its specifically deontic character from expressing and conveying a command, it would say that such sentences function in *certain respects* like ordinary imperative sentences. There are unlikenesses as well as like-

nesses, but it is illuminating to dwell on the likenesses. Perhaps this is all that Kant wished to imply by using the word "imperatives".

I will now make some comments. (1) I think that the first alternative can be rejected at once, at any rate as regards specifically moral deontic sentences. In the case of a *literal* imperative, e.g. "Form fours", there is no sense in asking whether what it expresses is *true* or *false*. The only sensible questions that can be raised about a literal imperative are these. Was it actually uttered, and, if so, was it meant seriously? Granted that that is the case, is there any doubt as to precisely what was commanded? Granted that there is no doubt, was the person who uttered the imperative sentence *entitled* to issue orders on this subject to the persons whom he addressed? Now in the case of a moral deontic sentence in the indicative, e.g. "You ought not to give false answers to questions", it seems plain that we can sensibly raise a question which does not fall under any of these headings, viz. "Is it in fact *true* or *false* that a person ought never to act in that way in such situations?" Conversely, it seems that certain of the questions that can be asked about literal imperatives do not arise in regard to moral deontic sentences in the indicative. A person may believe that he ought not to give false answers to questions, whilst he denies that anyone has actually forbidden him to do so, or denies that anyone is entitled to issue orders to him on this subject.

(2) Among literal imperatives we must distinguish two different kinds, which may be called "violent" and "legitimate". The imperative "Stand and deliver", issued by a highwayman at the point of his pistol to a traveller, is an example of the former. The imperative "Form fours", issued by an officer to a company of his own men whom he is drilling, is an example of the latter. Now there seems to be very little analogy between a *violent* imperative and what is expressed and conveyed by a specifically moral deontic indicative. On the other hand, analogies between *legitimate* imperatives and deontic sentences seem to involve something like a logical circle. A legitimate imperative is issued by a person, who has a *right* to give orders about a certain matter, to a person who is under an *obligation* to obey him in such matters. An officer stands to his men in a certain relationship which gives him a right to command them in certain of their actions, and places them under an obligation to obey his commands in regard to those actions. But that is itself a deontic statement, and it cannot in the end be reduced to a literal imperative.

(3) There are at least two causes which make it seem plausible to assimilate deontic indicatives to literal imperatives.

(i) There is at least one genuine likeness between the situation in which a person finds himself when he literally receives a command and that in which he finds himself when he feels under an obligation to behave in a certain way. In both cases the act is not one which he would do simply because he likes doing it, as he might dance a jig because he felt so inclined. Again, it is not one

which he would do as an obvious means to some immediate satisfaction, as he might act if he felt hungry or if he were offered some food whose taste he knows to be pleasant to him. On the contrary, the act commanded and the act felt to be obligatory are often alike in being unpleasant or boring in themselves. They are often alike in that they involve forgoing some immediate satisfaction, or bringing upon oneself some pain or loss, or incurring *some* danger. They tend to be acts which, we say, "go against the grain". And the more they do so, the more fully does the agent feel that he is being commanded, in the one case, and that he is under an *obligation* in the other. This is certainly an important analogy.

(ii) For Christians, Mahometans, and Jews, at any rate, some of the most important negative obligation, i.e. duties of forbearance, are formulated in the so-called Ten Commandments as *literal* imperatives, issued by God and promulgated on his behalf by his prophet Moses. This no doubt makes it easy for those brought up in any of these religions to *identify* what is expressed by a deontic indicative, e.g. "A person ought not to steal" with what is expressed by a literal imperative, e.g. "Thou shalt not steal" or "Do not steal". But, even if we were to accept the story of the events on Mount Sinai in the most literal sense, the inference would be invalid. At most it might be alleged that the only *ground* for the proposition "A person ought not to steal" is the fact that God issued the command "Thou shalt not steal". Now that fact is not itself an imperative, but is the *historical* fact that a certain command has been uttered on a certain occasion by a certain person. Moreover, it is not really possible to hold that that fact *alone* is the ground for the deontic proposition "A person ought not to steal". If a similar command had been uttered by Moses on his own authority, no one would suppose for an instant that the fact that he had uttered it would be a ground for the corresponding deontic proposition. An essential premiss would be that the command was issued by *God*, and that we, as his creatures, stand in such a relation to him that we have a *duty* to obey his orders. Otherwise the Ten Commandments would be nothing but violent imperatives. In that case, although it might be prudent to obey God's orders, there would be no more question of *moral obligation* than there is in handing over one's purse to a highwayman.

(4) We may note the following *differences* between literal imperatives and what is expressed by deontic sentences. (i) A person does not literally issue orders to *himself*. But it is just as intelligible for him to say "*I* ought to make an allowance to my old nurse" as to say "*You* ought to make an allowance to your old nurse". Attempts are sometimes made to evade this difference by representing statements of the form "I ought to do so-and-so" as expressing commands issued by a man's conscience or his higher self to his lower self. This mode of speaking involves personifying one's conscience or higher self and treating one's lower self as another person. It is plainly most artificial and cannot be taken literally.

(ii) A person does not literally command or forbid an action which he knows or believes to have been already done or left undone. But it is quite common to say "*A* did *X*, but he ought not to have done it". It is also quite common to say "*A* failed to do *X*, but he ought to have done it".
(5) It seems *prima facie* that certain deontic sentences, so far from *expressing* commands, state the ground for certain commands. Suppose I utter to someone the literal imperative "Pay me £2.19.4d immediately" Let us assume that I am not an armed robber issuing a violent command. The other man may say "Why should I?" Then it would seem that I am giving a reasonable ground for my demand if I can truly say "You promised to pay me that sum at this date and time, and you know very well that you *ought* to keep your promises".

On the whole, then, it seems to me that the differences between what is expressed by deontic indicatives and by literal imperatives are at least as important and striking as the resemblances.

1.423. Kant's views about deontic propositions
We can now deal briefly with Kant's views about what he calls "imperatives" and we call obligations of practical activity. In terms of the distinctions which we have drawn his doctrine can be stated as follows:
(1) He denied that there are any ultimate *teleological* obligations. I shall not go into his reasons. It seems to me that they are invalid, and that his conclusion is almost certainly false. For the obligation to produce as much good and as little evil as one can is certainly teleological, and it has as good claims to be counted as self-evident, and therefore ultimate, as any non-teleological obligation.
(2) He seems to have denied that there are any *restricted* ultimate non-teleological obligations. He seems to have held that all such restricted obligations as that of telling the truth when asked a question are non-teleological but *derived*.
(3) He seems to have accepted two and only two ultimate obligations, both of which are *non-teleological* and *unrestricted* in range. One is the obligation of practical consistency. This, when combined with psychological facts about human desires, gives rise to derived obligations of practical consistency which Kant calls "hypothetical imperatives". The other may be stated as follows "In any circumstances a person ought to behave in such a way and only in such a way as he could consistently will that *everyone* should behave in similar circumstances". This is what Kant calls "*the* Categorical Imperative" or "*the* Moral Law". He gives a number of other propositions which he says are equivalent to this, e.g. the principle that one should always treat oneself and others as ends and never as mere means. I cannot see that they are equivalent to it, and I shall not consider them further here.

(4) He holds that all restricted moral obligations, such as the duties of truth-telling, promise-keeping etc. can be derived from the one ultimate moral obligation stated above. Perhaps it would be more correct to say that he thinks that this principle furnishes a necessary and sufficient criterion, by which we can decide whether any proposed rule of conduct for situations of a specific type is morally obligatory or morally objectionable. The principle might be compared with the rules of the syllogism, and the specific maxims of conduct which we use it to test might be compared with particular arguments in syllogistic form. If and only if a particular syllogism accords with all the rules, any rational being is *logically bound* to accept it. Similarly, if and only if a maxim of conduct answers to Kant's fundamental principle, any rational being is *morally bound* to act in accordance with it. The test, according to Kant, is always of the same kind. Suppose that a person is inclined to act in a certain way in a certain situation, e.g. to tell a lie when asked an embarrassing question. Then he should ask himself the following question "Could I consistently will that any person whatsoever, when placed in a situation similar in all relevant respects to this one, should act in the way in which I am now inclined to act?" If the answer is "No", then such an act would be wrong. If the answer is "Yes" then the act is not wrong; it is either innocent or obligatory. (I am here stating Kant's doctrine in a rather charitable way. (i) He does not in fact explicitly add the qualification "when placed in a situation similar in all *relevant respects* to this one". But this is essential. For any two situations are unlike in innumerable respects. Unless we add this qualification, Kant's test could always be evaded. On the other hand, if we do add it, we see that the test presupposes certain moral judgments, viz. judgments as to what kinds of dissimilarity between situations are morally relevant and what are morally irrelevant. (ii) I have said that, according to Kant, if the proposed act answers to this test, then the act is *either* innocent *or* obligatory. I am not clear whether he would say this or would say that the act is *obligatory*. If so, his doctrine would seem to presuppose that every voluntary act which is not morally forbidden is morally obligatory. This seems plainly false. Suppose, on the other hand, that he would admit that an act which is not morally forbidden may also not be morally obligatory. Then some further test would be needed for distinguishing, among acts which are not morally forbidden, between those which are morally innocent and those which are morally obligatory.)

(5) Kant assumes that, in deciding whether one could consistently will that everyone should act in the way in which one is inclined to act, we do not need to use premisses about the actual desires, inclinations, and cognitive limitations of human beings. This seems to me to be a mistake. E.g. it might seem plausible to say that one could not consistently will that everyone who has made a promise should break it if he finds it inconvenient to keep it. But if

one asks why it would be inconsistent to will this, the answer seems to be as follows. In the first place, one would dislike to have other persons break the promises which they had made to *oneself* whenever they found it inconvenient to fulfil them. But this presupposes a certain aversion in oneself, which one no doubt correctly believes to exist *mutatis mutandis* in others. Secondly, one has good reason to believe that, human nature being what it is, promises would not be accepted if it was known that everyone felt free to break them whenever he found it inconvenient to keep them. And one desires that promises should be made and accepted, because this is a necessary condition of many other things which one desires. But all this presupposes specific desires and aversions and beliefs in oneself and in other human beings. It seems to me that the only case where we can talk of an inconsistant desire, without a tacit reference to the existence of certain other desires, is where the *desideratum* is complex and the realization of the various elements in it together would involve some logical or causal impossibility. An example would be desiring to be in London and in Cambridge at the same time. Such inconsistent desires can occur as actual experiences only in so far as a person fails to see the logical or causal incompatibility between the various elements in his desideratum. It should be noted here that to be desiring *A & B* involves desiring *A* and desiring *B*; but the converse does not hold. A person may be desiring *A* and at the same time desiring *B* without desiring *A & B*. For to desire *A & B* involves thinking of them together as combined in a certain way to form a single complex desideratum.

1.424. Comparison of Kant's views with certain others

In terms of the distinction which I have drawn the essential peculiarity of Kant's view may be stated as follows. He holds that all obligations of activity can be divided exhaustively into the two classes which I should call (a) *non-teleological* obligations (ultimate and derived) and (b) derived obligations of *mere practical consistency*. On my view this division is not exhaustive. There is a third class, viz. (c) teleological obligations (ultimate and derived). It will be worth while to compare Kant's views on this point with those of the strict utilitarians and with those of Ross.

A. The utilitarians. The strict utilitarian would hold that there are no ultimate *non-teleological* obligations, and therefore no derived non-teleological obligations. He would therefore wipe out Kant's class of categorical imperatives altogether. But he would hold, as against Kant, that there is one ultimate *teleological* obligation, viz. to produce as much good and as little evil as one can. From this can be derived a number of specific teleological obligations. Each of these, however, is only a *prima facie* obligation, which might break down in very special circumstances. And in each case the derivation would

require certain psychological and sociological premisses about actual human desires and limitations.

B. Ross. In terms of the distinctions which I have drawn Ross's position is as follows.[1] (1) He holds that there are *both* teleological and non-teleological obligations of activity. For he holds that we have a *prima facie* obligation to produce as much good and as little evil as we can. And he also holds that we have other *prima facie* obligations which cannot be derived from this, e.g. the duties of truth-telling, promise-keeping, etc. (2) He holds, as against Kant, that there are *ultimate* non-teleological obligations which are restricted in *range*. E.g. the obligation to give true answers is certainly restricted in application to occasions on which we are asked questions. And Ross regards this obligation as ultimate and non-teleological. (3) He holds, as against Kant, that the non-teleological *prima facie* duties may conflict with each other and with the teleological *prima facie* duty of general beneficence. That is why he refuses to call any of them "duties", without qualification, and calls them all "*prima facie* duties". Each is an obligation in the strict sense if and only if the situation is such that it can be carried out without breaking any of the other *prima facie* obligations. (4) The only obligation to which there can be no exception is "You ought to do that action which is on the whole most claim-fulfilling or least claim-frustrating when all the various aspects which tend to make it claim-fulfilling or claim-frustrating have been considered and weighed against each other in respect of urgency". (5) Since this is the only obligation without exception, it bears a certain analogy to the one ultimate teleological obligation of utilitarianism and the one ultimate non-teleological obligation of Kant's theory. But the analogy does not go very far. Nothing can be deduced from Ross's principle. It presupposes a whole mass of ultimate *prima facie* obligations; and orders the agent to weigh these against each other, without offering him any general principle by which he can measure their relative urgency. The utilitarian gives us a single ultimate teleological obligation, from which, together with special circumstances, all the more special rules can be *deduced*, and by which their limits can be determined when they conflict with each other. The Kantian professes to give a single ultimate non-teleological obligation, by reference to which any proposed maxim of conduct can be *tested* and shown conclusively to be right or wrong. He does not claim that special obligations can be deduced from his one ultimate obligation as a premiss, any more than particular syllogisms can be deduced from the *dictum de omni et nullo* by which the validity of any proposed syllogism can be tested. Ross disclaims the possibility of finding either a single ultimate obligation from which all other obligations can be

1. W.D. Ross, *The Right and the Good* (Oxford, 1930). Cf. also W.D. Ross, *Foundations of Ethics* (Oxford, 1939).

deduced, or a single ultimate obligation by which the claims of all other alleged maxims of conduct can be tested. His theory of ethics is essentially pluralistic. The utilitarian theory and the Kantian theory are essentially monistic, though they try to introduce unity in two quite different ways.

Now it seems fairly clear that the Kantian idea of a single self-evident ultimate principle by which all proposed rules of conduct could be tested and shown to be right or wrong, as syllogisms can be tested by the rules and proved valid or invalid, will not work. We are therefore left with two alternatives. One is the irreducible pluralism of Ross. The other is the theory of a single ultimate self-evident obligation from which, together with various special circumstances, all the various special obligations can be deduced. Now it is conceivable that this ultimate obligation might be non-teleological. But I do not know of any non-teleological obligation which could plausibly be suggested for the purpose. On the other hand, the utilitarian teleological obligation has strong claims to be considered self-evident. And it is plausible to hold that it might be used as a premiss from which, together with various special circumstances, all the various special obligations could be deduced. Therefore, in practice, the choice seems to lie between the irreducible pluralism of Ross and some form of utilitarianism.

1.425. Application to ethical scepticism

The distinctions which we have drawn between the various kinds of obligation are useful in dealing with the ethical sceptic, who says that he does not recognise that he is under any moral obligations. As Sidgwick points out, such a man may mean one or other of several different things, and he may not be at all clear which of them he means until you distinguish them for him. (1) He may merely mean that he does not recognise any *non-teleological* obligation of practical activity as binding on him. He does not see that there is anything in the nature of an act of promise-keeping, as such, to make it obligatory on him. He does not see that there is anything in the nature of an act of intentional deception, as such, to make it his duty to avoid it. Such a man may nevertheless recognise the teleological obligation to produce as much good and as little evil as he can. And he may be prepared to admit the other alleged obligations, so far and only so far as they can be shown to follow from this and from special circumstances. Such a man is not an ethical sceptic at all. (2) When this distinction is made clear to the man, he may say that he does not see any obligation on him to produce as much good or as little evil as he can. The next move would be to ask him whether he is only denying that he is under an obligation to consider *equally* the good of *everybody*. Does he admit that he is under an obligation to produce as much good or as little evil as he can, at any rate in a restricted circle, e.g. in his country, or his friends, or his family, or even in himself? If he admits that he is under *any* moral obligation

to produce as much good or as little evil as he can, no matter how restricted is the sphere within which he thinks this obligation is confined, he is not an ethical sceptic. (3) When it was put to him a man might deny that he recognises even the most restricted obligation to produce good rather than evil. Probably at this stage he would take the line of denying that the words "good" and "evil" have any application unless they are taken to mean "what I desire as an end" and "what I shun as an end". At this stage you could still raise the question whether he recognises the obligation of practical consistency. If so he admits certain derived obligations of mere practical consistency, though he admits no ultimate obligation to pursue this end rather than that and therefore admits no derived teleological obligation. (4) When this is put to him the man may finally deny that he recognises any obligation even of practical consistency. He will not admit even that he ought *either* to take what he knows to be the necessary means to the ends which he has decided to pursue *or* to give up pursuing these ends. And he will not admit even that he ought not *both* to do what he knows will prevent him from attaining a certain end *and* to continue to pursue that end. It is only the third and fourth types of man who can be called complete ethical sceptics. The others reject only certain special ethical doctrines. Of the fourth type we can say that, whether or not he is a knave, he is certainly a fool. I do not think that this can be said with certainty of any of the other three. Indeed it seems to me that the first of them is probably, and the second possibly, correct in his opinion.

Chapter 4

ETHICAL PROBLEMS: GOOD AND EVIL

1. Good and evil

The words "good" and "bad" or "evil" are used in a number of different senses. This does not mean that the word "good", e.g., has several completely disconnected meanings as, e.g., the word "post" or the word "plot" has. There is some connexion between all the senses in which the word is used. Some senses can be defined in terms of other senses, e.g. "good" in the instrumental sense can be defined in terms of "good" in one or other of its non-instrumental senses. And the different senses of "good" which are left when all the definition that is possible has been done have at least some kind of analogy to each other. A parallel case would be the senses in which we use the word "sharp" in the phrases "a sharp knife", "a sharp answer", and "a sharp lawyer". Obviously we do not mean precisely the same by "sharp" in these three phrases. But it is equally obvious that there is some real or fancied analogy between them which we feel justifies the use of this adjective in all three.

1.1. Various senses of "good" and "bad"

We can now consider some of the various senses in which the words "good" and "bad" are used.

1.11. Instrumentally efficient

Let us begin with a phrase like "That is a good knife". It is evident that this is not equivalent to the conjunction of the two propositions "That is good and that is a knife". We mean that this is good *as* a knife. It may be bad as a saw or as a razor. Plainly the right analysis here is the following. We are considering the object as an instrument which might be used to produce a certain kind of result in a certain kind of way. And we are saying that it is or is not likely to be efficient when used for this purpose. Our judgment then involves the notion of a possible agent using an object as an instrument to bring about a certain kind of result in a certain kind of way. We pass no judgment on whether the result which the agent wants to achieve is a good one or a bad one in itself or its probable consequences. We simply judge that an agent who uses the object as an instrument for achieving this kind of purpose is likely to get

what he wants through the efficiency of the instrument or is likely to be disappointed through the inefficiency of the instrument. Thus we can say that arsenic is a good poison, although we do not think that the death of the victim is good. There is one and only one respect in which we imply a judgment on the goodness of the result when we say that a thing is good in this sense. When I call an object a good knife, I do imply that a person who uses it for sharpening pencils is likely to get the satisfaction of gaining the end for which he used it. When I call the object a bad razor, I do imply that a person who uses it for shaving is likely to get the dissatisfaction of being frustrated in his purpose. But this is the *only* good or evil feature in the result which I consider when I use "good" and "bad" of a thing in this sense. "Good" and "bad" in this sense are used mainly of artificial objects, or of natural objects with very definite properties which make them specially adapted to be instruments for special purposes. We may substitute for them the terms "instrumentally efficient" and "instrumentally inefficient". We often use the phrases "good for" and "bad for" in this sense, e.g. "good for cutting", "bad for shaving", etc.

1.12. Conducive to the fulfilment or frustration of a widely felt desire

There is another sense of "good" and "bad" which is very closely connected with the above. It is reached by narrowing the first sense in one direction and widening it in another. A certain thing may have such properties that it is a frequent or an invariable cause-factor in the total causes of effects that are generally or universally desired, and is seldom or never a cause-factor in the total causes of effects that are generally or universally shunned. If so, we often express this fact by calling it "good". Such a thing need not be deliberately used by a human being as an instrument to effect a purpose. In this sense we should say that sunshine and ventilation are good, and that under-feeding is bad. Under-feeding is good for producing various kinds of illness; but practically everyone desires to avoid illness, and so we call under-feeding bad. We may substitute the phrases "conducive to the fulfilment of a generally felt desire" and "conducive to the frustration of a generally felt desire" for "good" and "bad" respectively, when used in this sense.

1.13. Average-comparative or ideal-comparative sense

We now pass to a different sense of the words. Take the sentence "That is a good rose". Roses are a natural kind or species of flower. Any individual rose will have a certain determinate form of each of the determinable characteristics which together mark out the rose-species. Individual roses which have these determinables in certain determinate forms are considered to be better specimens of the rose-species than others which have these determinables in certain other determinate forms. We have a rough notion of an "average

rose" and of an "ideal rose". By calling a particular thing a "good rose" we mean that it has the rose-characteristics in such determinate forms that it comes nearer to the ideal rose than the average rose does. By calling a particular thing a bad rose we mean that it has the rose-characteristics in such determinate forms that it is further from the ideal rose than the average rose is. We are not making any judgment as to whether roses are in any sense good things or bad things in themselves or have on the whole good or bad effects.

A given individual may be a member of several natural kinds or species of different orders. Thus a certain creature is a cat, and is a mammal and is an animal and is a living thing. When we talk of an individual as a good or a bad specimen of "*its*" species, we are generally thinking of the lowest species or natural kind of which it is a member. We are thinking of a natural kind, such as cats, whose subdivisions are not themselves natural kinds but are mere classes, such as black cats, tom-cats, etc. Now, just as we can consider an individual in relation to the lowest species of which it is a member, so we can consider a species in relation to the next higher species of which it is a subdivision. Thus I might say of a certain animal "That is a good kangaroo, but kangaroos are a bad kind of mammal, though mammals are a good kind of animal". The sense in which I use "good" and "bad" all through is essentially the same. It might be called the "average comparative" or the "ideal comparative" sense.

1.131. Remarks on the senses so far considered

There are several points to be noticed about the three senses of good which we have now considered, viz. instrumentally efficient, conducive to the fulfilment of a generally felt human desire, and approximating nearer to the ideal member of a species than does the average member. (i) In all of them the word "good" stands for a relational property and not for a quality. (ii) In the first two senses the other term of the relation is quite definitely human beings and their purposes and desires. We may sometimes overlook this fact, however; just as we may overlook the fact that the words "right" and "left" always refer to the body of the person who uses them and to the direction in which he is facing at the time. (iii) In the third sense the other term of the relation is, at the first move, what we call the "average" and the "ideal" members of a species. But we must remember that the phrases "average rose" and "ideal rose", e.g., are not descriptions of actual existents. They stand for concepts which are formed by human beings who have compared a number of actual roses with each other, arranged them mentally in various series, and so on. Now this fact by itself would not suffice to make the ideal-comparative sense of "good" relative to human desires and purposes. It might be that this process of comparison led to concepts of ideal types of plants or animals which are definable without any reference to our desires and purposes, as e.g.

the concepts of perfect circularity or exact straightness in geometry or of a perfect gas in physics are. But in point of fact the notion of an ideal so-and-so generally does contain a good deal of reference to human tastes, desires, and purposes. This is particularly obvious in the case of species which have been deliberately bred and cultivated by men either for use or for pleasure, aesthetic or otherwise, e.g. horses, dogs, roses, etc. Since the horse species, e.g., has been bred and modified by human beings mainly in order to produce animals which will be efficient for hunting, for carting, and so on, part of the notion of an ideal horse will be that it is an animal well-adapted for one or other of these specifically human purposes. This subjective element is at its minimum in the concepts of ideal types of the various species of wild plants and animals and of crystals which botanists, geologists, and mineralogists form in connexion with purely theoretical classifications. (iv) A very special case is where the species in question is the human species, and one is considering whether an individual is or is not nearer to the ideal *man* than the average *man* is. (a) In the first place the human species does stand out objectively in a quite unique way from all other known species. Men are the only creatures known to us who have the power of speaking and writing, of making deductive and inductive inferences, of designing and using tools and machines, of contemplating alternative possibilities and choosing between them and so on. Any man can do many things which no animal can do; and there is nothing that any animal can do which cannot be done and surpassed by men either directly or by means of machinery which they design and construct. There is therefore a perfectly objective sense, quite independent of our desires or tastes or prejudices, in which it can be said that the human species is of unique importance and value as compared with every other species known to us. (b) In consequence of this the property of being conducive to the satisfaction of specifically human desires, though it is relational, does stand out in an objective way from the property of satisfying specifically canine desires or specifically equine desires. You might say "If a dog could consider the matter he would say that the property of satisfying specifically canine desires is of outstanding objective importance". But this really gives up the case; for to suppose that a dog *could* consider such questions is equivalent to supposing that he has the intellectual powers which are characteristic of men and distinguish them from all other known creatures. (c) One of the peculiarities of men is that they have ideas of right and wrong, good and evil, and that they have morally directed emotions, and the desire to do what is right and avoid doing what is wrong. Hence part of the notion of an ideal man has a reference to his specifically moral characteristics. It therefore seems likely that, when a man is called "good" in the average or ideal comparative sense, part of what is meant is definable only in terms of "good" in some specifically moral sense.

Some philosophers, e.g. Spinoza, have held that the only senses in which the word "good" can be intelligently used are in the sense of "instrumentally efficient", or of "conducing to the fulfillment of generally felt human desires", or in the sense of "coming nearer to the ideal man than the average man does". They have held that the notion of an ideal man is simply that of a man who has the specifically human powers to the maximum possible degree. And they have held that these powers are objectively outstanding because they exist in no other known species and because they enable men to do all and more than all that the members of any other species can do. I think that this is *prima facie* a consistent view. But I think that it may come to grief in the long run over the fact that part of the notion of an ideal man involves a reference to specifically *moral* qualities, and that these in turn involve a reference to "good" and "evil" in a sense which cannot be brought under any of the three headings so far considered.

1.14. "A good singer but a bad man"

Next we must consider statements like "He is a good singer, but a bad man". The class of singers is a sub-class of the species man, but it is not itself a species in the sense of a natural kind. Now the members of any natural kind have a very large number of determinable characteristics in common, and each member will have each of these determinables in a certain special determinate form. In trying to form the idea of an average member or an ideal member of a species and to group the actual members around it we shall have to consider all these determinables. *A* may be above the average, *B* below it in respect of one of them, and *B* may be above the average, *A* below it in respect of another of them. When we say that *A* is a good singer but a bad man we might mean that, taking him all round, *A* falls well below the average member of his species; but that, in respect of a certain determinable characteristic, viz. the power of singing agreeably, he comes well above the average. Very often, however, we mean something different. We mean that, in respect of his *specifically moral* characteristics, *A* falls well below the average member of his species; but that, in respect of his power of singing agreeably, he comes well above the average. In English the phrase "a good man" or "a bad man" without qualification, generally means a *morally* good or a *morally* bad man. If we want to remove this restriction we generally use the phrase "a good all-round man" or "a poor specimen, on the whole, of a man".

1.15. Specifically moral sense

This brings us to the distinction between "good" and "bad" in a specifically moral sense and in other senses. We apply the adjectives "morally good" and "morally bad" or "wicked" to rational beings, to certain of their dispositions, to certain of their acts, and to certain of their experiences.

(a) Experiences. An experience can be good or bad, e.g. a pleasant or a painful sensation, without being morally good or bad. But certain experiences, viz. certain desires and emotions, can be morally good or bad. E.g. a malicious emotion is a morally bad experience. But it may be pleasant, and, in that respect, good in a hedonic sense. A benevolent desire is a morally good experience. But the experient may realise that he cannot carry it out; and so the desire may be unpleasant, and therefore bad in a hedonic sense. It must be noticed that a desire to produce *any* kind of good result, no matter whether the goodness of the result is moral goodness or not, tends as such to be a *morally* good desire. Neither beautiful objects nor pleasant experiences are, as such, *morally* good. But a desire to produce beautiful objects or to produce pleasant experiences (at any rate in others) is, as such, a morally good desire.

(b) Acts. An act is morally good if it is done from one or other of certain motives. It is, e.g., morally good if a sufficient motive-factor for doing it is the agent's belief that it is right, and his desire to do what is right. It is morally bad if the agent believes it to be wrong, and nevertheless does it because his aversion to doing what is wrong is overcome by other desires. Of course the degree of its moral badness in this case will vary very much with the nature of the other desires which overcome the desire to do what is right. The act will be much worse if, e.g. the desire to do right is overcome by malice than if it is overcome by personal affection.

(c) Dispositions. A morally good disposition is a disposition to have morally good experiences or to do morally good actions. We distinguish *morally* good and bad dispositions from others, which are good or bad in a non-moral sense, by the words *virtuous* and *vicious*. Thus a disposition to have pleasantly toned experiences or to make perfect approach shots at golf is a good, but not a virtuous, disposition.

(d) Persons. A morally good person is one who has strong and well-organised moral dispositions, which control and organise his non-moral dispositions, keeping the bad ones in check and allowing the good ones to be exercised and developed.

If we reflect on the above examples we notice the following important facts. (i) In order that an experience shall be morally good or bad it is necessary that it shall have an object, real or imaginary. A mere undirected feeling may be hedonically good or bad, but it cannot be morally so. (ii) This, however, is not sufficient. A pure cognition, with no emotional or conative tone, if such there could be, would be neither morally good nor morally bad. At most it might have the kind of value or disvalue which is derived from being correct or delusive, well-founded or ill-founded. An experience cannot

be morally good or bad unless it is a cognition of an object (real or imaginary) toned with desire or aversion or with some kind of emotional tone. (iii) The moral goodness or badness of such an experience depends jointly on the psychological quality of the experience and the nature of its epistemological object, whether real or imaginary. What makes an experience morally good or morally bad is always the fittingness or unfittingness of its emotional or conative quality in kind or degree to its epistemological object. E.g. desire or aversion, as such, is neither good nor bad. It is morally good to desire certain kinds of object and to feel aversion to certain other kinds; and it is morally bad to desire certain kinds of object and to feel aversion to certain other kinds. Again, it is morally good to desire a certain kind of object with an intensity that falls within certain limits. But to desire it with an intensity which falls outside those limits may be inordinate and unfitting, and therefore a morally bad experience. Similarly for emotion. There is probably no kind of emotional tone which will *suffice* to make an experience which it tones morally good or morally bad regardless of what its object may be. Certain kinds of emotional tone are appropriate to the cognition of certain kinds of object, and certain kinds are inappropriate to the cognition of certain kinds of object. And, again, even the kind of emotion appropriate to a certain kind of object may be felt towards it with appropriate or inappropriate intensity. (iv) It therefore looks as if the fundamental ethical notion were that of *fitting*ness or *unfitting*ness, in a very wide sense. We have already seen that it is involved in the notion of rightness and wrongness of actions. We now see that it is an essential factor in the notion of morally good and morally bad experiences.

1.16. Applications to continuants

The first three senses in which "good" and "bad" are used apply primarily to continuants, viz. things and persons. For "good", in the sense of "instrumentally efficient" and in the sense of conducive to the fulfilment of a generally felt desire, presupposes a thing or a person with relatively permanent dispositions and powers. And "good", in the sense of "good of its kind" presupposes a species of things or persons classified and arranged in an order according to determinate values of their determinable dispositional properties. In discussing the distinction between "morally good or bad" and other senses of "good" and "bad" we had to consider instances of the application of the words "good" and "bad" to occurrents. For experiences and actions are occurrents. Now it might be held that "good" and "bad" as applied to occurrents must be more fundamental than "good" and "bad" as applied to continuants. For, it might be said, continuants are called "good" or "bad" only in respect of their dispositional properties. Now dispositions are tendencies to have or to produce such and such occurrent states under

such and such conditions. Therefore dispositional properties must be definable in terms of the characteristics of occurrents. Therefore "good" and "bad", as applied to continuants, must be definable in terms of "good" and "bad" as applied to occurrents.

The suggestion is, e.g., that by calling a *person* "good" you *mean* simply and solely that he has such dispositions so organised that under most circumstances he will do good acts or have good experiences. If so, the sense in which you apply "good" to him *is* definable in terms of the sense in which you apply "good" to his acts and to his experiences. The latter will then be a more fundamental sense of "good" than the former. But there is one possibility which this view fails to notice. It is possible that there is another sense of "good" as applied to persons. It is possible that when I call a person "good" my *ground* for doing so is that he has dispositions to do good acts and to have good experiences, but that this is not what I *mean* by calling him "good". On this view, the property of having such dispositions is a good-*making* characteristic of a person, just as the property of being a desire for another man's happiness is a good-*making* characteristic of an experience. But to call a person who has such dispositions "good" is not merely to say that he has such dispositions; just as to call an experience which has this property "good" is not merely to say that it has this property. If this view be accepted there is a sense of "good" which applies to persons and is just as fundamental as the sense of "good" which applies to experiences and acts. The former cannot be *defined* in terms of the latter. But the characteristic which makes a person "good", in the former sense, is the property of having such dispositions as tend to make him have experiences and do actions which are "good" in the latter sense.

1.17. Senses in which an occurrent is called "good"

An occurrent is often called "good" or "bad" in certain derivative senses which depend on its being a cause-factor in the total cause of certain kinds of effect. These senses are analogous to the first two senses in which "good" and "bad" are used of continuants, viz. "instrumentally efficient or inefficient" and "conducive to the fulfilment (or frustration) of a generally felt desire". When I call an event "good" I often mean only that it was an indispensable cause-factor in the total cause of some result that I wanted. When I call it "bad" I often mean that it was an indispensable factor in producing some result which I wanted to avoid, or that it was a sufficient factor in frustrating some desire of mine. This sense is analogous to "instrumentally efficient" as applied to continuants. Sometimes when I call an event "good" I mean that it was an indispensable cause-factor in the total cause of a result of a kind which is generally or universally desired. Sometimes I mean that events of this kind usually are cause-factors in total causes which produce

results that are generally desired and seldom are cause-factors in total causes which produce results that are generally viewed with aversion. In the former case I could express what I mean by saying that this event was "fortunate" or "unfortunate". In the latter case I could express what I mean by saying that this event is "of a kind which is normally fortunate", even though it may actually have been unfortunate itself.

1.18. Extrinsic and intrinsic goodness

The next distinction to be considered is that between *intrinsic* and *extrinsic* value. In discussing this I shall have to deal with some matters which really belong to metaphysics. I cannot discuss them as fully as I should wish to do, and what I shall say about them may be described as skating rather gingerly over rather thin ice.

I think we must begin by trying to distinguish between the intrinsic and the extrinsic properties of a particular. I define the intrinsic properties of a particular as those which it is logically possible for it to have had even if nothing had existed except itself and its own parts if it has parts. The extrinsic properties of a particular are those which it is logically impossible for ti to have had if nothing had existed except itself and its own parts. E.g. it is logically possible that an object should be round even if nothing but itself and its parts had existed; but it is logically impossible that a person should be a father if no one else had existed. We must distinguish between logical and merely causal impossibility. It is causally impossible that I should have existed unless my parents had existed, and I certainly inherited some if not all of my properties from them. But it is not logically impossible, i.e. it does not involve any contradiction and does not conflict with any *a priori* truth, that a person should have existed without parents and should have had without inheritance the same properties as those which I in fact inherited from my parents. In distinguishing intrinsic and extrinsic properties it is only logical, and not causal, impossibility that we have to consider.

Now it seems that the intrinsic properties of a thing or a person fall into three groups. (1) *Pure qualities*. (2) *Structural properties*, i.e. the spatial and temporal relations between its parts. It is a structural property of a picture to consist of patches of various colours and shapes juxtaposed in a certain way in space to constitute a certain variegated patterned expanse. It is a structural property of a symphony to consist of sounds of various qualities, pitches, and intensities, some of which occur simultaneously and some in a certain order of succession. (3) *Dispositional properties*. It is a dispositional property of arsenic to be poisonous and of copper to be soluble in nitric acid. You might say that these cannot be intrinsic, because they involve in their definition a reference to other persons or things, e.g. to an animal organism or to nitric

acid. But I do not think that this is valid. It seems to me that arsenic would be poisonous and copper soluble in nitric acid even if there had never been any animal organisms for arsenic to poison or any nitric acid for copper to dissolve in. In the case of these dispositional properties the reference to other things, e.g. animal organisms or nitric acid, is purely conditional, not categorical. To say that arsenic is poisonous means that *if* a living organism were to take in arsenic it *would* die. To say that copper is soluble in nitric acid means that *if* a bit of copper were to be put into nitric acid, it *would* dissolve. Plainly it is logically possible that these conditional propositions should be true even though the conditions were never fulfilled. What is true is that these dispositions would remain for ever latent if arsenic or copper were the only things that had ever existed. It is also true that no-one could have known the fact that arsenic or copper had these properties unless they had been manifested. Finally, it is true that no-one could from a conception of these properties unless he had the conceptions of animal organisms or of nitric acid, as the case might be, and that he could not conceive of such things unless they had existed and been perceived. But I do not think that any of these facts is incompatible with the statement that arsenic would have been poisonous and copper soluble in nitric acid even if there had never been anything but arsenic or if there had never been anything but copper. On the other hand, the property of poisoning Mr. Jones at a certain time would be an extrinsic property of a bit of arsenic; for it is logically impossible that it should have had that property unless Mr. Jones had existed and had swallowed it.

I think that the extrinsic properties of a particular consist of its actual relationships to particulars other than its own parts. These relationships may be causal, e.g. the property of poisoning Mr. Jones at a certain time. Or they may be non-causal, e.g. the property of being in Mr. Jones's stomach at a certain time.

There is one other point to be noticed. It is held that many of the dispositional properties of physical objects, at any rate, depend upon their structural properties. Thus it is a dispositional property of metallic gold to be yellow. This means that any person with normal eyesight who looked at a bit of gold in ordinary daylight would see it as yellow. But this is generally held to depend on the molecular structure of gold which causes it selectively to reflect a certain constituent in ordinary daylight and to absorb the other constituents. Now, even if you refuse to admit that the dispositional property itself is intrinsic, you can hardly deny that the structural property on which it depends is intrinsic.

Now I think that, when people talk of the "intrinsic goodness or badness" of anything, they mean the goodness or badness which it derives from its intrinsic properties. Similarly its extrinsic goodness or badness means the goodness or badness which it derives from its extrinsic properties. If this is

what they mean, it would be better to talk of "intrinsically *derived*" and "extrinsically *derived*" goodness or badness. For the ordinary phrases suggest that we have to do with two different *senses* of "good" and "bad"; but it seems to me that we are concerned with goodness or badness in the same sense but of different origins. The following analogy may make this plain. The blueness of the sky and the blueness of a Trinity undergraduate's gown arise in quite different ways. In the former case it is due to the scattering of light from very small particles; in the latter it is due to the presence of a dye which selectively reflects light of a certain wave-length. But we *mean* precisely the same by "blue" when we call the sky "blue" and when we call a Trinity gown "blue".

1.19. Contributory sense

We are now in a position to deal with what Ross calls the "contributory sense of 'good' and 'bad'".[1] This is closely connected with what Moore calls the *principle of organic unities*.[2] So I will begin by explaining and discussing this principle.

1.191. Principle of organic unities

I think that the essential point of this principle may be stated as follows. Any whole W consists of certain elements A, B, C, ... etc., intimately related in a certain order by a certain relation R. Thus, $W = R\ (A, B, C, \ldots)$. Now, in general, the intrinsically determined value of such a whole will depend jointly on the qualities of the elements, on the relation between them, and on the order in which they are related within the whole by this relation. I do not think that there is any sense in talking about each of these three factors making its separate contribution, and in talking of the value or disvalue of the whole as the algebraical sum of the values or disvalues contributed by the qualities of the elements, by the relation between them, and by the order in which they are interrelated, respectively. What is possible in some cases is the following. Sometimes we can compare a whole $W = R\ (A, B, C, \ldots)$ with another $W_1 = R\ (B, A, C, \ldots)$ where nothing is dissimilar but the order in which the elements are related. And we can compare it with another whole $W_2 = R(\alpha, \beta, \gamma, \ldots)$ where nothing is dissimilar except the qualities of the corresponding elements. And we can compare it with another whole $W_3 = R'(A, B, C, \ldots)$, where nothing is dissimilar except the relation which interrelates the elements. If we found that the intrinsically determined values of W_3, W_2, W_1 and W were different, we could ascribe the variation in value entirely to a variation in the relation, in the quality of the elements, and in the arrange-

1. Ross, *The Right and the Good*, ch. 3.

2. G.E. Moore, *Principia Ethica* (Cambridge, 1903). See especially pp. 27ff. Cf. also, G.E. Moore, *Ethics* (London, 1912), Ch. 7.

ment of the elements, respectively in the three cases. An example would be the sort of whole which is composed of a set of sounds sounded simultaneously and in succession within the experience of a given hearer. The beauty or ugliness of the whole sound-complex depends jointly on the pitch, loudness, and tone-quality of the various sounds; on their occurring simultaneously or in continual succession with partial overlapping in the experience of a single hearer; and on the order in which they happen.

Obviously there is no reason to expect that there will be any simple relation between the value or disvalue of a complex whole $W = R(A, B, C, \ldots)$, and the values or disvalues which its elements would have in isolation from each other or in other wholes. The value of an isolated element can depend only on its qualities or its dispositions or its internal structure. The value of an element in a whole W will depend in part on the relationships in which it stands to the other members of this whole. Thus it is quite unreasonable to expect that two elements which are precisely alike in their intrinsic properties will have the same value when one was in isolation and the other was an element of a complex whole. And it is quite unreasonable to expect that they will have the same value as elements of differently constituted complex wholes. For value depends both on extrinsic and on intrinsic properties. If an element were completely isolated, it could have nothing but intrinsic properties. If an intrinsically similar element were part of a complex whole it would *ipso facto* have extrinsic as well as intrinsic properties. And, if two intrinsically similar elements were parts of differently constituted complex wholes, they would *ipso facto* have different extrinsic properties. Now of course some extrinsic properties may be neither good-making nor bad-making. And it may sometimes be the case that different extrinsic properties are equally good-making or equally bad-making. But we have never any right to assume this.

Two very important consequences follow at once. (a) It is never safe to infer the value or disvalue of a complex whole from the known values or disvalues which its elements would have in isolation or as elements in other wholes. For this leaves out of account the value which such a whole may derive from its elements being interrelated in a certain way within it. (b) It is never safe to use arguments of the following kind, although such arguments are very common. The argument would take the following form. We know that a certain whole $W = R(A, B)$ is good or that it is evil. Suppose we also know that A in isolation is indifferent. We are then very liable to conclude that, if W is good, B in isolation must be good; and that, if W is bad, B in isolation must be bad. We now see that any such argument is fallacious even if its conclusion should happen to be true. It is quite possible, e.g., that neither A nor B in isolation has any value or any disvalue and yet that W is good or is bad. For, when we say that A in isolation has no value or disvalue we mean that its *intrinsic* characteristics are neither good-making nor bad-making.

When we say that B in isolation has no value or disvalue we mean that its *intrinsic* characteristics are neither good-making nor bad-making. But the whole W has the structural characteristic of consisting of A and B interrelated in a certain order by R. And this may obviously be a good-making or a bad-making property of W. Again, when A is an element in W, it has, in addition to its intrinsic properties, the extrinsic property of standing in the relation R to B. And this may be a good-making or a bad-making characteristic of A. Similarly, when B is an element in W, it has, in addition to its intrinsic properties, the extrinsic property of standing in the converse of the relation R to A. And this may be a good-making or a bad-making characteristic of B.

The principle of organic unities is simply the statement of these facts. When it is thus stated it is quite obvious and not in the least paradoxical. And it is important because people so often do use arguments which conflict with the principle and are therefore invalid.

The following consequences of the principle are worth noticing. (1) (a) The substitution of one element for another in a whole may alter the value of the whole even though both the original element and the one that is substituted for it are intrinsically indifferent. (b) The substitution of an intrinsically good element for one that is intrinsically indifferent or bad may *diminish* the value of the whole or change it from being good to being bad. (c) The substitution of an intrinsically bad element for one that is intrinsically indifferent or good may *increase* the value of the whole or change it from being bad to being good. (2) Suppose there is a whole $W = R\ (A, B, C)$. Suppose that we can bring an element D, which was not before in any close relation with the elements of W, into a certain close relation; so that we get a new whole $W' = R'\ (A, B, C, D)$. Then there is no reason to expect that there will be any simple relation between the values of W and W', on the one hand, and the intrinsic value of D on the other. This may be summed up rather roughly as follows. "The addition of an element to a whole may give rise to a new whole of different value, though the element be intrinsically indifferent. The addition of an intrinsically good element to an indifferent or good whole may give rise to a whole which is less good or even positively bad. The addition of an intrinsically bad element to an indifferent or bad whole may give rise to a whole that is less bad or is even positively good."

I will give a few examples. (a) If two tunes, each of which by itself was pleasing, were played together in the hearing of a single listener, the resulting series of sounds might be hideous. (b) Certain condiments, e.g. pepper, produce sensations of taste which are unpleasant in isolation. But, if this taste be produced in conjunction with the rather insipid taste of some kind of food, the resulting complex set of taste-sensations may be much better. (c) The following would sometimes be given as examples. Consider the following four kinds of experience: (i) cognising with pleasure another's pleasant experi-

ences, (ii) cognising with displeasure another's pleasant experiences, (iii) cognising with pleasure another's unpleasant experiences, and (iv) cognising with displeasure another's unpleasant experiences. The last is a morally good experience, viz. sympathetic sorrow. Now it might be said that here we have a whole composed of two factors, viz. my sorrowfully toned cognition and your unpleasant experience. Each factor has a quality, viz. unpleasantness, which is bad-making. Yet the whole is morally good. The second and third are morally bad experiences, for the second is a case of envy and the third is a case of malice. Now it might be said that here we have a whole composed of two factors, viz. my cognition and your experience. One factor has the good-making characteristic of pleasantness, and one has the bad-making characteristic of unpleasantness. And the whole is in each case morally bad.

Now I am fairly certain that this kind of analysis is wrong, and that these are not really examples of the principle of organic unities. My reason is this. Suppose that I mistakenly believe that you are having a pleasant experience, and suppose that this mistaken belief of mine is pleasantly toned. Then my experience has just the same moral value as if you really were having a pleasant experience, even though you are in fact asleep or in pain. Similar remarks apply, *mutatis mutandis*, to the other three cases. We may sum up the situation as follows. The moral goodness or badness of any such experience is determined jointly by two of its characteristics, viz. (a) its own hedonic tone and (b) its property of being a belief that another person is having an experience with such and such an hedonic tone. Since the moral goodness or badness of the experience will be exactly the same whether this belief be true or false, there can be no question here of a whole in which the other person's experience is an element. Therefore these experiences of sympathy, malice, envy, etc. cannot be examples of the principle of organic unities, as defined by us. We have here a single experience with two characteristics; not a whole composed of two experiences interrelated in a peculiar way.

Although these are not examples of the principle of organic unities, they are examples of another principle which is rather like it and is equally important. I will call this the *principle of resultant value*. It may be put as follows. From the fact that two characteristics X and Y together would give a certain value or disvalue to anything that had them *both* nothing can be inferred as to the value or disvalue which *either of them separately* would give to anything that had it. And from the fact that X without Y would give a certain value or disvalue to anything that had it, and the fact that Y without X would give a certain value or disvalue to anything that had it, nothing can be inferred as to the value or disvalue which X and Y together would give to anything that had them both. E.g., an experience of mine is not rendered either good or bad by the mere fact that it is a belief that you are in pain. Also an experience of mine is not made morally good or morally bad by the mere fact that it is pleasant or

that it is unpleasant. But an experience of mine *would* be made morally bad by the conjunction of the two properties of being pleasant and being a belief that you are in pain. For it would then be a malicious emotion.

1.192. Definition of the contributory sense of "good" and "evil"
It is now easy to explain and define the contributory sense of "good" and "evil". Sometimes when a thing is called "good" all that is meant is the following. (a) All or most wholes in which this thing, or a thing of this kind, is an element have considerable positive value. (b) In each case, if the thing is removed from such a whole, the residue has much less value or is actually bad. When these conditions are fulfilled we say that this thing, or things of this kind, are *contributively good.* A similar definition could be given of being *contributively bad.* It is obvious from the principle of organic unities that a thing which is contributively good may be intrinsically indifferent or bad and that a thing which is contributively bad may be intrinsically indifferent or good. When an *event* is said to be good in the contributive sense this must mean that all or most wholes in which an event of this kind is an element have considerable positive value, and that the residue which would be left if such an event were omitted would in each case have much less value.

1.193. Collective and distributive goodness
There is one other notion connected with the principle of organic unities. When a whole is composed of a set of parts, every one of which would be good in isolation, it may be called "distributively good". When it is composed of a set of parts, every one of which would be bad in isolation, it may be called "distributively bad". If a whole is both collectively and distributively good, it may be called *good throughout.* If it is both collectively and distributively bad, it may be called *bad throughout.* It follows from the principle of organic unities that a whole might be collectively good and distributively bad or collectively bad and distributively good. One other case remains. A whole may be composed of a set of parts which are not all intrinsically good and are not all intrinsically bad and are not all intrinsically indifferent. It may then be called "distributively mixed". Let us take an example from Kant's ethical theory. According to him pleasure is intrinsically indifferent, and willing what is right as such is intrinsically good. But a total state of affairs consisting of right willing accompanied by the happiness which it deserves is the *Summum Bonum.* Such a whole is collectively good; and is intrinsically better than its only good constituent, viz. right willing. But it is not *good throughout*; for it is distributively mixed, since the constituent of happiness is intrinsically indifferent. The constituent of happiness contributes to the goodness of the Summum Bonum without itself having any intrinsic goodness. The constituent of right willing not only contributes so the good-

ness of the Summum Bonum but also is intrinsically good. On Kant's view right willing is the only thing in the universe which is both intrinsically good and has no elements that are not intrinsically good. So, although Kant never explicitly formulated the principle of organic unities, it is evidently presupposed in his theory of the Summum Bonum.

1.2. "Good" and "good-inclining"

When I talked about right and wrong I said that we must distinguish between rightness and wrongness themselves, and certain properties of acts, such as being the keeping of a promise or being an intentionally misleading answer to a question, which make an act right or make it wrong. I called these at first "right-*making*" and "wrong-*making*" characteristics. When we looked more closely into the matter we saw that it was more correct to call them "right-*inclining*" and "wrong-*inclining*" characteristics. Right and wrong are ethical characteristics, but right-inclining and wrong-inclining characteristics can be described in purely non-ethical terms. Now precisely similar remarks apply to good and bad. If a thing or person or experience or act is said to be good or to be bad, it is always reasonable to ask what makes it so. And the answer will always consist in mentioning some quality or dispositional property or relational property of the entity in question. E.g., painfulness is a bad-making quality of sensation; being a pleasantly toned belief that another is in pain is a bad-making quality of a cognition; and so on. Here also it is bad to talk of "good-*inclining*" and "bad-*inclining*" characteristics. For we have just seen that the presence of X in the absence of Y, might, e.g., make a thing good, whilst the presence of X together with Y might make it indifferent or bad. We might define a "good-inclining" characteristic as one which, if present alone or in most combinations, tends to confer goodness on anything which it characterizes. And we could define a "bad-inclining" characteristic in a similar way.

Now I think it is plain that we often use the words "good" or "bad" when we really mean "good-inclining" or "bad-inclining" respectively. It would, e.g., by quite sensible to say "Happiness is good" or "Envy is bad". But I think it is plain that these are elliptical statements. When we say that happiness is good what we mean is that a person's state of mind tends to be good if and in so far as it is one of happiness. When we say that envy is bad we mean primarily that an experience tends to be bad if and in so far as it is an unpleasantly toned cognition of the success of a rival. We may mean also that a person tends to be bad in so far as he has a disposition to have experiences of envy. It seems, then, that whenever a characteristic, i.e., the sort of entity which is denoted by an abstract noun or an adjective, is called "good" or "bad", what is meant is that it is a good-inclining or a bad-inclining char-

acteristic. Things or persons or experiences or other events can literally be *good* or *bad*; characteristics can only be good-*inclining* or bad-*inclining*.

Is there anything beside things and persons and experiences and other events that can literally be good or bad? The only other entities that seem to be possible subjects for these predicates are what might be called "facts" or "states of affairs". Take, e.g., such statements as "It was a good thing that the R.A.F. won the Battle of Britain"; "It is a good state of affairs when happiness and unhappiness are distributed according to people's deserts"; and so on. I think it is plain that "good" is not used here in the sense of good-inclining. It might be said perhaps that the winning of the Battle of Britain was a complex process composed of a number of simultaneous and successive events interrelated in certain ways. It might not be so easy to regard a state of society in which happiness is distributed in accordance with virtue as a process composed of interrelated events. So on the whole I think it is safest to say that certain facts or states of affairs can be literally good or bad in addition to things, persons, experiences, and other events.

1.3. Are "good" and "bad" definable?

There has been a great deal of discussion as to whether "good" and "bad" stand for characteristics which are logically analysable into simpler terms. If they are, the words "good" and "bad" are definable, in the sense in which words like "square" are definable. If they are not, these words are indefinable, in the sense in which, so far as we know, the words "black" and "white" are indefinable.

1.31. Moore's theory

I think that the best plan will be for me to state in my own way what I understand to have been Moore's theory at the time when he wrote *Principia Ethica*, and to discuss it and the arguments for and against it. The theory may be stated as follows. (1) When we use a sentence like "That experience is good" we are often, if not always, expressing a judgment in which we ascribe a certain characteristic to the experience in question. So the word "good" is often, if not always, used as the name of a characteristic. (2) As we have pointed out, the word "good" is highly ambiguous. In some senses in which it is used it undoubtedly stands for a complex characteristic which can be analysed. When used in these senses the word can be defined. And some other word such as "benefic" or "contributively good" can be substituted for it. (3) But some of these definable senses of the word presuppose another sense of it, which we will call the *primary sense*. In this sense, it stands for a characteristic which is simple and therefore unanalysable. Therefore the word "good", in this primary sense, cannot be defined. (4) It follows at once that

the characteristic for which it stands cannot be a relational property, i.e. a characteristic of the form "having *R* to so-and-so". For obviously all relational properties are complex and analysable into a relation and a term. Hence the characteristic must be either a pure quality or a pure relation. (5) The characteristic is in fact a quality and not a relation. (6) The characteristic is of a peculiar kind, which Moore calls "non-natural".

I think that these are the essential points of the theory. They are not all separately stated by Moore. You will be able to judge for yourselves by referring to Chapters I and II of *Principia Ethica*, to the essay on *The Conception of Intrinsic Value* in *Philosophical Studies*,[1] and to Moore's contribution to a symposium called *Is Goodness a Quality*? in the Supplementary Vol. XI of the *Aristotelian Society Proceedings*.[2]

I will now take the six points in my statement of the theory in order.

1.311.Is "good" the name of a characteristic?

(1) Moore always assumes that "good" is used as the name of one or another of several characteristics in sentences like "This experience is good". He assumes that this will be admitted by everyone, and that the only question is as to the nature of this characteristic or these characteristics. Now it has been pointed out by many philosophers in recent years that it is not safe to let this assumption pass without question. Certainly the *sentence* "That is good" is of the same grammatical form as many sentences which undoubtedly do state that a certain thing has a certain characteristic. It is of the same form as "That is round", e.g.; and there is no doubt that a person who uses the latter sentence is intending to convey the belief that a certain particular has a certain characteristic of which "round" is a name. But we must remember that a sentence which is grammatically in the indicative mood may really be in part emotive or imperative. It may be in part the expression of a certain emotion which the speaker is feeling, in which case it may be called *interjectional*. In that case to utter the sentence "That is good" would be equivalent to uttering a purely expository sentence in the indicative, followed by a certain interjection. E.g., it might be equivalent to "That is an act of self-sacrifice. Hurrah!" Similarly to utter the sentence "That is bad" might be equivalent to "That is a deliberately false statement. Blast!" Again, it may be used to evoke a certain emotion in a hearer. In that case to utter the sentence "That is good" would be like uttering a purely expository sentence in a pleasant tone and with a smile. To utter the sentence "That is bad" would be like shouting a purely expository sentence with a frown. Here the utterance of the ethical words "good" or "bad" is merely a stimulus to produce certain emotions in the hearer, as smiling at him or shouting at him might do. In this case the

1. G.E. Moore, *Philosophical Studies* (London, 1922).
2. Reprinted in G.E. Moore, *Philosophical Papers* (London, 1959).

sentences might be called *evocative*. Lastly, such sentences may be used to command or to forbid certain actions in the hearer. To utter the sentence "That is good" might be equivalent to uttering a purely expository sentence in the indicative followed by a sentence in the imperative. E.g., it might be equivalent to "That is an act of self-sacrifice. Imitate it!" And to utter the sentence "That is bad" might be equivalent to "That is a deliberately false statement. Don't do that again!"

On this view, words like "good" and "bad" do not *mean* anything, in the sense in which words like "white" and "square" do. There is no characteristic of which they are names. A person who utters sentences in which they occur as grammatical predicates is not using them to convey the belief that a certain subject has a certain peculiar characteristic of which the grammatical predicate is a name. And a person who hears such sentences and understands them is being exhorted or commanded or emotionally stimulated but is not receiving any special kind of information about the subject of the sentence. If this is so, Moore's theory breaks down at the first move, and so do the theories of most of his opponents.

It has been pointed out that this theory fits in with two very important facts. (i) It explains why all attempts to define ethical words in purely expository terms seem unsatisfactory. Suppose you substitute a sentence which contains none but expository words for one that contains an ethical word. Then the interjectional, evocative, or imperative force, which the original sentence derived from the ethical word in it, has vanished. You feel that something is missing, and you are quite right. Now suppose you take for granted, as Moore did, that ethical words are names of characteristics. Then you will naturally try to explain this feeling of something being missing by saying that the proposed analysis of an ethical characteristic into purely expository characteristics has missed out some essential factor of the ethical characteristic. (ii) Attempts to define one ethical word, e.g. "good", partly in terms of another ethical word, e.g. "right", do not always seem unsatisfactory. E.g., it is not obviously inadequate to define "a good experience" as "an experience which one can *rightly* desire to have". Nor, on the other hand, is it obviously inadequate to define "right" as "conducive to *good* consequences". Now the theory can explain this fact too. Both the original sentence and the proposed equivalent now contain an ethical word. They therefore both have interjectional, evocative, or imperative force. Now it is possible that two difference sentences, both of which have this kind of force, may produce precisely similar effects, as evokers of emotion or as commands, in all people of a certain community who hear them. Now suppose you take for granted, as Moore did, that ethical words are names of characteristics. Then you will think that the more complex of two such equivalent sentences states the analysis of the ethical characteristic which you believe to

be named by the ethical word in the simpler of the two sentences. And so you will think that some ethical characteristics can be analysed in terms of other ethical characteristics and expository characteristics.

I think that this theory may be further supported by reflecting on how we learn ethical words as children. I suspect that for a small child "good" and "right" acts are practically co-extensive with those which its mother or nurse refers to in a certain tone or with a smile, or which she exhorts one to do. And "bad" or "wrong" acts are practically co-extensive with those which its mother or nurse refers to in a certain other tone or with a frown, or which she exhorts one not to do. Very soon the ethical words acquire the same evocative or interjectional or imperative force as the tone of voice or the smile or frown or the actual command or forbidding. I pointed out that many words are amphibious in character, viz. partly expository and partly ethical. Cf., e.g., the sentence "That is a statement made with the intention of producing a false belief" with the sentence "That is a lie". Now it is certain that the second sentence does commonly express or stimulate an emotion which the first does not. And it is plausible to say that this is the *whole* difference between the first sentence, which is purely expository, and the second, which has an amphibious predicate and is partly ethical.

Let us call theories of this kind "non-attributive theories", since they hold that sentences in which the word "good" occurs as grammatical predicate do not in fact ascribe any special attribute to the subject of the sentence. It seems to me then that this kind of theory is quite plausible enough to deserve very serious consideration. It would have to be refuted before we could be sure that the question "Are the characteristics denoted by ethical names analysable or unanalysable?" is a sensible question. If the non-attributive analysis of ethical sentences is sound, the question would be like asking whether the present king of Utopia is intelligent or stupid.

(2) Henceforth we will suppose, for the sake of argument, that words like "good" and "bad" are names of characteristics. We may say that, when "good" is used in the sense of "benefic" or "contributively good", it stands for a characteristic which is complex. And we will assume that, when "good", in these senses, is defined, the definition always involves the word "good" in another sense, which may be called the *primary* one.

1.312. Criteria for a characteristic being unanalysable

(3) The question now is this "Assuming that the word "good", in the primary sense, is the name of a characteristic, is there any reason to believe that this characteristic is unanalysable?" It seems to me quite clear that there is no means of *proving*, with regard to any characteristic, that it is unanalysable. At most we might be able to show that no analysis so far proposed is satisfactory. And even this is not always so easy as one might think. The

question involves several very difficult and fundamental logical points, which I will now try to state.

Suppose a person raises the question whether the characteristic of which a certain word "N" is the name is simple or complex, and whether a certain proposed analysis of this characteristic is correct or not. Plainly, in *some sense* of the phrase, he must 'know what the word "N" means'. For, otherwise, he does not know what he is asking the question about. Equally plainly this cannot be the same as 'knowing the analysis, if any, of the characteristic which "N" stands for'. If he knew this in knowing what "N" means, the question whether the characteristic is simple or complex, and what is its correct analysis if it is complex, could never arise for him. So the question presupposes at least the following three propositions. (a) That there is a certain one characteristic which the person who asks the question is thinking of whenever he uses the name "N" in certain kinds of context. (b) That, whether this characteristic is in fact simple or in fact complex, he can think of it without *ipso facto* knowing that it is simple or knowing that it is complex as the case may be. (c) If it is in fact complex, he can think of it without *ipso facto* knowing its correct analysis. In practice a further assumption is always made, which we will call (d). It is assumed that all or most other people who speak the language correctly are thinking of the same characteristic as the questioner whenever they use the word "N" in the same kinds of context.

Now it might be extremely difficult to justify assumptions (a) and (d) in many cases. Can I be sure that there is any *one* characteristic which I am thinking of whenever I use the word "good" in the primary sense? May there not be a whole lot of characteristics, such that I am sometimes thinking of one and sometimes of another when I use the word "good" in the primary sense? Again, can I be sure that, when other people use the word "good" in certain contexts, they are always or generally thinking of the characteristic which I am thinking of when I use it in such contexts? The only evidence that can be produced is consistency or inconsistency of usage. Do I call similar things "good" sometimes and "bad" at other times? Do other people agree among themselves and with me in the things that they call "good" and the things that they call "bad"? If there is great inconsistency about applying the words "good" and "bad", there is at least a presumption that conditions (a) and (d) are not fulfilled. Now there certainly is a considerable amount of inconsistency.

We will suppose, however, that this difficulty can be overcome, and that we can satisfy ourselves that conditions (a) and (d) are fulfilled. We will now concentrate our attention on conditions (b) and (c). If I were lecturing on logic, I should raise certain difficulties about (b) and (c); but I propose to waive these, and to pass straight to the following question. Suppose you can think of a certain characteristic without *ipso facto* knowing whether it is

simple or complex, and without *ipso facto* knowing its correct analysis if it is complex. How are you to set about answering the question whether a characteristic which you are thinking of is simple or complex? And, if it is complex, how are you to decide whether a proposed analysis of it is right or wrong? Suppose it is suggested that the characteristic C is analysable into the characteristics C_1, C_2, C_3. (a) We can reject this at once if we can mention an instance of something that *has* C and *lacks* either C_1 or C_2 orC_3. And we can reject it at once if we can produce an instance of something that *has* C_1, C_2, and C_3 and yet *lacks* C.

(b) Suppose we are left with one or more suggested analyses of C, which pass this test of being so far as we know exactly co-extensive with C. We shall next proceed as follows. Granted that I know of nothing which has C and lacks any of the characteristics C_1, C_2 and C_3, and that I know of nothing which has C_1, C_2 and C_3 and lacks C, can I conceive that there *might* be such a thing? If so, I can reject the proposed analysis of C into C_1, C_2 and C_3. For a characteristic and its analysis would be *necessarily* co-extensive. The equivalence of their extensions would not be just a contingent fact like the fact that chewing the cud and having cloven hoofs are exactly co-extensive. (c) Suppose we are left with at least one suggested analysis of C which passes this test and can be seen to be *necessarily* co-extensive with C. There might be several such. E.g., the property of being circular is necessarily co-extensive with each of an enormous number of other complicated properties. For there are innumerable properties which we can prove *must* belong to all circles and *cannot* belong to anything but circles. So we are finally faced with this question. If we know of *only one* set of characteristics which is necessarily co-extensive with the characteristic C, how can we tell whether this set is or is not an analysis of C? And, if we know of *several* such sets of characteristics, how can we tell which, if any, is the *analysis* of C, and which are necessarily and reciprocally connected with C but are not the analysis of C? Suppose, e.g., that it seemed evident that anything which was good would necessarily be a fitting object of desire, and that anything which was a fitting object of desire would necessarily be good. How could one tell whether being a fitting object of desire is the *analysis* of being good, or whether it is just a complex characteristic which is necessarily and reciprocally connected with the quality of goodness, but is not the analysis of that quality?

It seems to me that at this stage further argument becomes impossible. All that an objector can say is "I feel that your proposed analysis of goodness misses out something that I have in mind when I use the word *good*". Or "I can't believe that, when I use the word *good*, I am thinking of anything so complicated as I should be if your proposed analysis of goodness were correct". Now suppose that another person does not feel that the suggested analysis misses out anything that *he* has in mind when he uses the word *good*.

And suppose that he thinks that what he has in mind when he uses the word may easily be as complex as it would be if the suggested analysis were correct. We are assuming that the parties have somehow persuaded themselves that they are thinking of the same characteristic whenever either of them uses the word *good* in similar contexts. What further argument is possible between them?

The real situation, however, is not quite like this. I think it is true to say that all fairly simple analyses of goodness in purely non-ethical terms seem to *most* people to omit something. (Cf., e.g., "to be good" means "to be generally desired as an end".) And all analyses of goodness in purely non-ethical terms which avoid this defect seem to *most* people to be too complex to be correct analyses of what they have in mind. (Cf., e.g. "to be good" means "to be something which most men would approve of themselves or others for desiring.") It is only certain definitions which are partly in ethical and partly non-ethical terms that might seem to many people to avoid both defects. (Cf., e.g., "to be good" means "to be a fitting object of desire".) Now how much weight is to be attached to a fairly *general* feeling that suggested analyses of goodness in purely non-ethical terms either miss out something which we have in mind or are too complex to be correct analyses of what we have in mind? I think we commonly make the following assumptions without ever stating them clearly. We assume that, if I have thought of a certain characteristic *C* often enough to have associated a name with it, then, whenever a proposed analysis is felt by me to be either inadequate or unduly complex, it is pretty certainly incorrect. I think it would be admitted that a proposed analysis might *in fact* be incorrect even though I did *not* feel it to be either inadequate or unduly complex. But, it would be said, if I *do* feel it to have either of those defects then it probably *is* defective. And, if most people who have frequently thought of a certain characteristic agree in feeling that a proposed analysis of it is inadequate or unduly complex, it becomes practically certain that the proposed analysis is defective.

Now as regards this general principle there are two things to be said. (a) I am not much impressed with the importance of a widespread feeling that a proposed analysis is unduly complex. We are assuming that a person can think of a characteristic without *ipso facto* knowing its analysis, if it has one. Now it seems difficult to suppose that he can estimate the *degree* of internal complexity of a characteristic, when he does not know whether it is simple or complex, and does not know its analysis if it has one. (b) More weight should, I think, be attached to a widespread feeling that a proposed analysis is inadequate. This has to be accounted for somehow, and the most obvious explanation is that the analysis really does omit some factor in the characteristic, or that it analyses not *this* characteristic but some other which is allied to it. Unfortunately this is just the place at which the non-attributive theory

becomes highly relevant. It may be that the explanation is simply that the name of the original characteristic has acquired a certain interjectional, evocative or imperative force which is lacking in the phrase that expresses the analysis. We feel the lack of this, and we conclude that the analysis is inadequate.

(4) The fourth point in my statement of Moore's theory was that if the characteristic denoted by "good" is simple, it cannot be a relational property. It must be either a quality or a pure relation. This is quite obvious. But, in order to show that goodness *is* either a quality or a pure relation, it would of course be necessary to add the premiss that the characteristic denoted by "good" is simple. I have tried to show that this has not been proved, and that there is no conceivable way of proving it. The utmost that has been shown is that all analyses in completely non-ethical terms, which have so far been suggested, seem to most people to be either inadequate or unduly complex. For reasons which I have given, I do not think that this proves conclusively even that none of these proposed analyses is correct. And, even if they were all incorrect, it would still be possible that there might be a correct analysis in completely non-ethical terms, which no one happens to have suggested. Again, it would still be possible that there might be a correct analysis, partly in non-ethical terms and partly in other ethical terms, such as "right" or "fitting". There is not even a presumption against this, since certain proposed analyses of this kind do not seem to most people to be obviously inadequate or obviously too complex. It seems to me then that no good reason has been produced for holding that the characteristic denoted by "good", in the primary sense, is not a relational property.

(5) The fifth point was that "good", in the primary sense, is not the name of a relation; and therefore must be the name of a quality. I think it is obvious that "good" is not the name of a relation. If it denotes a characteristic at all, the characteristic which it denotes is either a quality or a relational property. So, if it could be shown that it denotes a simple characteristic, we could admit at once that it denotes a simple quality. The only remark that I wish to make at this point is the following. It does seem to me conceivable that the relation denoted by "better than" might be more fundamental than the characteristic denoted by "good". It might be that the former is simple and unanalysable, and the latter is complex and definable in terms of the former. The suggestion would be that "good" always is an abbreviation for "good of its kind"; and that "good of its kind" means "better than the average member of its kind". This would, of course, make "good" the name of a relational property of a peculiar sort, in which the relation is "better than". If it could be shown that "good", in the primary sense, does not denote a relational property at all, this suggestion could be at once refuted. But I suspect that some peope, who think they have proved this, have not considered the possibility that "good"

might denote a relational property in which the relation is "better than". And perhaps they would not be so sure that it might not denote a relational property of this peculiar kind, even though they were convinced that it could not denote a relational property in which *any other* relation occurred.

1.313. Are "good" and "bad" non-natural characteristics?

(6) The last point in Moore's theory is that "good", in the primary sense, is the name of a non-natural characteristic. This is probably the most important part of the theory. But two questions at once arise. (i) What exactly is meant by the distinction between a "natural" and a "non-natural" characteristic? (ii) What connexion, if any, is there between the doctrine that "good", in the primary sense, denotes a characteristic which is simple and unanalysable, and the doctrine that it denotes a characteristic which is non-natural?

1.314. The distinction between "natural" and "non-natural" characteristics

The division of characteristics into "natural" and "non-natural" was first introduced by Moore many years ago in the *Principia Ethica*. There has been much discussion since then as to whether goodness and certain other alleged characteristics are "natural" or "non-natural". But it seems to me that the distinction itself has never been clearly stated either by Moore or by anyone else. It will be best to begin with an account of Moore's own successive statements on this question.

(1) The first account is to be found in *Principia Ethica* (pp. 40 – 41). It seems to me to be most unsatisfactory. This is admitted by Moore himself in the latest pronouncement which he has made on the subject, viz. in the terminal essay which he contributed to the volume entitled *The Philosophy of G.E. Moore*. He there says that this part of *Principia Ethica* is quite wrong.

The essential points in this passage of *Principia Ethica* may be put as follows. He first describes what he calls a "natural object". This is said to be anything capable of existing in time, e.g. a stone, a mind, an explosion, an experience, etc. Thus it is practically equivalent to a particular existent, whether a continuant or an occurrent. Next it is said that natural objects can have two kinds of characteristics, viz. *natural* ones and *non-natural* ones. He gives the following two distinguishing marks of a *natural* characteristic. (1) Any natural characteristic could be conceived as existing in time all by itself. (2) Every natural object is a *whole*, whose *parts* are its natural characteristics. He defines a "non-natural" characteristic by opposition to a natural one. It is a characteristic which *cannot* be conceived as existing in time all by itself, but can be conceived as existing only as the property of some natural object. And it is not a *part* of any natural object which it characterises.

Now it seems to me plain that there are and can be no characteristics answering to Moore's description here of *natural* characteristics. Take, e.g.,

as an example of a natural object a penny. Surely its brownness and its roundness *cannot* be conceived as existing in time all by themselves. And surely a penny is *not* a whole, of which its brownness and its roundness are *parts*. On the contrary it is a *substance*; of which its brownness and its roundness are *attributes*. Yet Moore certainly regards the brownness and the roundness of a penny as *natural* characteristics.

(2) The second attempt wich Moore made to explain the distinction is in the essay entitled *The Conception of Intrinsic Value* (*Phil. Studies*, pp. 253 – 275). This is a very difficult paper. In reply to criticisms Moore reverted to the subject in the terminal essay of *The Philosophy of G.E. Moore*. He remarks there that he used the phrase "intrinsic property" in a very unfortunate way in this essay. He used it in such a way that there would be no inconsistency between the following three statements, viz. "*P is* intrinsic", "*P is* a property", and "*P is not* an intrinsic property". For, he says, the doctrine which he intended to assert in the essay was that goodness *is* intrinsic and *is* a property, but is *not* an intrinsic property of good things. In view of this it will be best to ignore the earlier essay and to confine our attention to the amended statements of his doctrine which Moore made in the terminal essay of *The Philosophy of G.E. Moore*.

(3) The amended statement, taken together with the parts of the earlier essay which he does not wish to amend, comes to the following. (i) Those properties of a thing, and only those, which *depend solely on its intrinsic nature*, are now to be called "intrinsic properties". (ii) To say that a property P of a thing T depends solely on the intrinsic nature of T is to assert the following two propositions. (a) That it would be impossible for T to have P in one determine form at one time or in one set of circumstances, and to have P in a different determinate form or not to have P at all at another time or in another set of circumstances. (Thus, e.g., the property of looking brown or of looking round would *not* be an intrinsic property of a penny, since it would look red if illuminated by red light and would look elliptical if viewed very obliquely.) (b) That it would be impossible for T to have P in a certain determinate form and for another thing T', exactly like T in all other respects, to have P in a different determinate form or not to have P at all. (E.g., it might be plausibly held that no experience *exactly like* an experience of mine in all other respects could be an experience of any other person's. If that is so, the property of being an experience of *mine* would be an *intrinsic* property of any experience of mine.)

Before passing to the next point in the statement of Moore's position we must make the following two explanatory comments on conditions (a) and (b) above. (α) The word "impossible", which occurs in both conditions, is ambiguous. We must distinguish, in the first place, between being *only relatively* impossible and being *absolutely* impossible. It is only relatively im-

possible, i.e. inconsistent with the actual laws of nature, for an unsupported body in the neighbourhood of the earth to remain where it is and not to fall to the ground. It is *absolutely* impossible that 2×2 should equal 5, or that a body should at the same time be a cube and a sphere. The sense of "impossible" which is involved in the definition of "depending on the intrinsic nature of T'" is not merely relative impossibility. It is absolute impossibility. (β) The phrase "exactly like", which occurs in the second condition is to be interpreted as follows. Two things T and T' are not to be counted as exactly alike unless all the following conditions are fulfilled. Every quality possessed by either must be possessed in precisely the same determinate form by both. If either has parts, both must consist of the same number of parts. To each part of one there must correspond a precisely similar part of the other. And the mutual relations of the parts of one must be the same as the mutual relations of the corresponding parts of the other and must relate them in the same order in both cases. (One can see why these conditions are necessary if one thinks what would be involved, e.g., in two roulette-boards each with sectors of various colours, being exactly alike.) (iii) This being understood, we can pass to the third proposition in Moore's amended statement. I think that it may be put as follows, but I may be mistaken. The *natural* characteristics of a thing fall into two classes. (a) Its *extrinsic* properties, i.e. those which do *not* depend solely on its intrinsic nature. (b) A certain *sub-class of its intrinsic properties*. Moore makes various attempts to state the peculiarities of this sub-class of the intrinsic properties of a thing. The peculiarity which he finally concentrates upon is the following.

Each of the *natural* intrinsic properties of a thing, which is logically independent of the rest, *contributes something* special towards describing the intrinsic nature of the thing. No description of its intrinsic nature could be *complete* if it omitted any one of its *natural* intrinsic properties, unless the presence of the property omitted were logically entailed by the presence of one which was already explicitly included in the description. (The point of these qualifications is this. The sub-class might include, e.g., two such properties as being red and being coloured. Now anything that was red would necessarily be coloured. Therefore if a description of a thing explicitly included the fact that it was red, nothing further would be contributed towards describing it by adding that it was coloured. Similarly a description which explicitly included the fact that it was red would not be made incomplete by not explicitly including the fact that it was coloured.) Subject to this qualification, we may now sum this up as follows. A *natural* property of a thing is either (a) an *extrinsic* property of it, e.g. a coin's property of looking round from here now, or (b) an *intrinsic* property of it which contributes something special towards describing its intrinsic nature and therefore cannot be omitted from any *complete* description of its intrinsic nature, e.g. in the

case of an experience, the property of being an experience of *fear*, of being an experience of *mine*, and so on. (iv) We come now to the last point, viz. the account of a *non-natural* property of a thing. It is plain that this would be a property which (a) *is* intrinsic, and (b) does *not* belong to that sub-class of a thing's intrinsic properties which are *natural*. It would therefore be an *intrinsic* property of a thing which does *not* contribute anything towards describing its intrinsic nature. A description of the thing which omitted to mention this property of it might be complete, although this property was not logically entailed by any of the properties explicitly included in that description.

It is plain, then, that the notion of a non-natural property of a thing involves two factors, one positive and the other negative. The positive factor is that it depends solely on the intrinsic nature of the thing. The negative factor is that it contributes nothing towards describing the intrinsic nature of the thing. It is the combination of these two factors which makes the notion of a non-natural property so paradoxical, and makes one wonder whether there could be such properties.

Let us now take a concrete example and see what all this comes to. Take a malicious emotion occurring at a certain moment in A and referring to B. This would be a complex state of mind consisting of a thought in A's mind of B as suffering pain or disappointment and a feeling of pleasure in A's mind at that thought. Moore would pretty certainly regard this as an instance of something which is *intrinsically bad* in the moral sense. He would also hold that its intrinsic moral badness is a *non-natural* property of it. What exactly does this come to?

On the positive side it would consist in asserting the following two propositions. (1) That it is absolutely impossible that that particular state of mind should under any conceivable difference of circumstances have been morally *good* or morally *indifferent* or have had any *other kind or degree* of moral badness than that which it actually has. (2) That it is absolutely impossible that any other state of mind, which exactly resembled this one in all its other intrinsic properties, should be morally *good* or morally *indifferent* or have any *other kind or degree* of moral badness than that which this one has.

On the negative side it comes to this. To say of this state of mind that it occurs in A, that it is feeling of pleasure, that it is a thought of B as suffering pain and disappointment, and so on, all contribute something special towards describing the state of mind, in a quite familiar sense of the word "describe". One could imagine such statements being piled up to form a *complete* description of it. But to add that it is *intrinsically bad* contributes nothing towards describing it, in this sense of "describe". In this sense of "describe", it could be completely described to a person who had no idea of moral good or evil. And, if he should afterwards acquire these ideas and should then discover that this state of mind is intrinsically bad, this would add nothing to the description which he already had of it.

I will make two comments on this. (1) Moore has nowhere defined the sense of "describe" in which to say that a state of mind is *pleasant* and to say that it is a *thought of another's misfortune* does contribute to "describe" it, whilst to say that it is *morally bad* does not. He just hopes that we shall all be able to recognise it by examples. Unless something more definite can be said about it than this, his attempt to distinguish non-natural properties from natural ones remains very unsatisfactory. (2) It seems to me that a person who held the non-attributive view of ethical words and sentences could account quite plausibly for the facts which have led Moore to distinguish certain properties as "non-natural". Suppose, as this theory alleges, that the word "bad" is not the name of a characteristic at all, but that its function is merely to express or to evoke a certain kind of emotion. Then to call a state of mind "morally bad" would not contribute towards describing it, in the obvious sense of mentioning its properties. Yet, owing to the likeness in grammatical form between, e.g., the two sentences "That experience is *pleasant*" and "That experience is *morally bad*", it might seem paradoxical that the former does and the latter does not contribute towards describing the experience. So a person who had never thought of the non-attributive analysis of ethical sentences might try to account for the difference by supposing (as Moore does) that "morally bad" *is* the name of a characteristic, but that the characteristic is of a *very queer kind*.

1.315. Is there any connexion between "simplicity" and "non-naturalness"? We can now deal with the following question. Is there any special connexion between the doctrine that "good" stands for a characteristic which is *simple and unanalysable*, and the doctrine that it stands for a characteristic which is *non-natural*? Moore certainly holds both these views, but is there any logical connexion between them? I think it is fairly plain that there is not.

(1) It is quite clear that a property might be both natural and unanalysable. The property denoted by the word "yellow" as used in the sentence "That thing looks yellow" is certainly *not analysable*. The word "yellow", when used in that sense, certainly cannot be defined. It can only be exemplified. Now this property is certainly *natural*. To say of a thing that it looks yellow obviously contributes towards describing it. The same is true of many psychological properties. The quality which distinguishes an emotion of *fear* from other experiences is certainly *unanalysable*. It could not be explained to a person who had never felt fear. It is also certainly *natural*. To say of an experience that it is toned with fear certainly contributes towards describing it. So there is no logical impossibility in the words "good" and "bad" standing for characteristics that are at once simple and natural. Therefore to prove that they stand for characteristics which are *simple* would not *ipso facto* prove that they stand for characteristics that are *non-natural*.

(2) If there are any non-natural characteristics there must of course be some *simple* ones. For it is obvious that if there are complex non-natural characteristics, each of them must ultimately be analysable into a conjunction of simple characteristics. And it is also obvious that a *non-natural* characteristic could not be analysed into a conjunction of simple characteristics all of which were *natural*. But if there are any *non-natural* characteristics, there is no reason why *some* of them should not be complex. There would in fact be three possible kinds of complex characteristics, viz. (i) purely natural, (ii) purely non-natural and (iii) mixed. Suppose, e.g., that "morally good" were the name of a characteristic which is both simple and non-natural. Then "morally beneficial" would be the name of a complex characteristic of the mixed kind, for it would mean "tending to produce or to maintain morally good experiences". And the notion of "producing" or "maintaining" is the notion of a *natural* characteristic.

(3) Although there is no *logical* connexion between being a simple characteristic and being a non-natural characteristic, there is the following *de facto* connexion in the particular case of moral goodness and moral badness. There is in fact no *simple* natural characteristic which it would be at all plausible to identify with the characteristic described by "morally good" or "morally bad". Those who have held that these words stand for natural characteristics have nearly always suggested some *complex* natural characteristic. An example would be the theory that to be "morally good" means to be an object of a feeling of approval in all or most impartial spectators. Therefore, if one could show that "morally good" is the name of a *simple* characteristic, one would *in fact* have refuted most of the existing theories which assert that it is the name of a *natural* characteristic.

1.316. Epistemological account of the distinction between "natural" and "non-natural" characteristics

The distinction between "natural" and "non-natural" properties was certainly intended by Moore to be ontological and not epistemological, i.e. a distinction in the objective natures of the two kinds of property, and not a distinction in the ways in which we get to know them. But his attempts to explain the distinction in ontological terms have not been very successful. I think therefore that it may be worth while to try to approach the question from the epistemological point of view.

Suppose we were to say that a characteristic is epistemologically natural if (a) we become aware of it by perceiving particulars which sensibly present it to us or by introspecting experiences which introspectively present it to us; or (b) it is wholly definable in terms of characteristics which fall under (a) together with the notions of cause and/or substance. I think that this would cover every characteristic which Moore or anyone else would want to describe

as "natural". It would, e.g., cover *yellowness*, both in the sense in which it occurs in the sentence "That looks yellow to me from this view" and in the sense in which it occurs in the sentence "Gold is yellow". Yellowness in the first sense comes under (a). In the second sense it comes under (b). For to say that gold is yellow is to say that it is so constituted as to cause sensations of yellowness in all normal persons when they view it in white light. Again, it would cover all psychological characteristics, such as pleasantness, the fear-quality, the anger-quality and so on. For we become aware of these by introspecting experiences which present themselves to our introspection as pleasant, as fearful, and so on. And other psychological characteristics, such as "timorous", are defined in terms of the former and the notion of cause: for to be timorous is to have a tendency to have experiences of fear from very slight causes.

To say that a characteristic is *epistemologically non-natural* would be to say that it is (a) *not* presented to us sensibly by any particular which we perceive nor introspectively by any experience which we introspect, and (b) that it is *not* definable in terms of characteristics presented to us in either of those ways, together with the notions of cause and or substance.

We can now raise the following question. Supposing that the phrase "morally good" or "morally bad" is the name of a characteristic, is that characteristic epistemologically natural or epistemologically non-natural? Since we have defined these phrases, we know what the question involves.

(1) It seems quite obvious that moral goodness or badness is not sensibly presented to us by anything that we perceive with our senses, as yellowness is when we look at the sun or as coldness is when we touch a lump of ice. In the first place, it is pretty clear that "goodness" and "badness", in the moral sense, cannot significantly be ascribed to the sort of things which we perceive with our senses, i.e. to bodies and to physical events. And it is quite obvious that we do not literally see or fear or taste or smell or feel such objects as "good" or as "bad". At the most we may perceive with our senses certain combinations of epistemologically natural properties (e.g. certain combinations of colour or of sound) which are good-*making*.

(2) It seems almost equally obvious that no *simple* psychological characteristic which is presented to us introspectively when we introspect our experiences can be identified with moral goodness or badness. When we introspect our various experiences they present themselves to us as pleasant or unpleasant, as experiences of fear or hate or desire or aversion, and so on. In this way we become aware of various simple psychological characteristics. Now it is true that goodness and badness, in the moral sense, can belong to experiences. Indeed many people would hold that, in the primary sense, they can belong to nothing but experiences. Yet I think that a moment's reflexion will convince one that by calling an experience morally "good" or "bad" one

does not *mean* that it is pleasant or that it is unpleasant, or that it has any one of the various simple psychological qualities which are presented to us by introspection.

If anyone is tempted to identify goodness, as applied to experiences, with any one of these psychological qualities, I think he would do so owing to the following confusion. What he would really believe is that there is one and only one good-*making* quality of an experience, e.g. pleasantness, and that there is one and only one bad-*making* quality of an experience, e.g. unpleasantness. He then fails to notice the distinction between *goodness* or *badness* itself and what he regards as the one and only good-*making* or bad-*making* quality. And so he thinks he believes, e.g., that "good" and "pleasant" are just two names for a single characteristic, as, e.g., "rich" and "wealthy" are. Since *pleasantness* certainly is an epistemologically natural characteristic, he will go on to say that "goodness" is the name of an epistemologically natural characteristic. But I do not think that the belief that one means the same by "good" and by "pleasant", when applied to an experience, would survive for a moment after the distinction between goodness itself and a good-making characteristic had been pointed out to one. And I think that the same would be true, *mutatis mutandis*, of any other simple psychological characteristic which one might be tempted to identify with the characteristic denoted by "good", as applied to experiences.

(3) It is not so obvious *prima facie* that "goodness" might not be the name of some fairly complex characteristic, involving nothing but simple psychological characteristics and perhaps also the notions of cause and/or substance. E.g., it is not *prima facie* obvious that to call a malicious experience "bad" might not be equivalent to saying that such experiences call forth a certain kind of anti-emotion in any normal human being who contemplates them when he is in a normal emotional state. Now, if goodness were a characteristic of this kind, it would be an epistemologically natural one according to our definition.

I think that the upshot of this discussion may be summarised as follows. Suppose that a person utters with conviction such a sentence as "That experience is morally bad" or "That experience is morally good". (We might take as examples of two such experiences (i) a feeling of pleasure at the thought of another's pain or misfortune, and (ii) a desire to help another person believed to be in trouble.) Then the following hypothetical proposition seems to me fairly certain. *If* in uttering such a sentence he is ascribing a characteristic to the experience in question and is not merely evincing a certain kind of emotion towards it, and *if* that characteristic is simple, *then* it is pretty certainly epistemologically non-natural. But it is by no means certain that the antecedents of this conditional proposition are fulfilled. For, in the first place, it is quite possible, as upholders of the non-attributive analysis assert,

that he is merely evincing a certain kind of emotion towards the experience in question, and is not ascribing *any* characteristic whatever to it. And, secondly, it is quite possible that he is ascribing to it a characteristic which is *complex* and analysable. In that case two possibilities are open: (i) That the complex characteristic in question is composed of simple characteristics which are *all* epistemologically natural, e.g. of psychological ones, together with the concepts of cause and/or substance. (ii) That it is composed of simple characteristics, one at least of which is epistemologically non-natural. An example of this alternative is provided by the theory that to call an experience "morally good" is to say that it is a *fitting* object of a feeling of approval, where "fitting" is supposed to stand for a simple epistemologically *non-natural* characteristic.

Now there are two considerations, one direct and the other indirect, which might be brought against the possibility of there being any characteristics answering to my definition of "epistemologically non-natural".

(1) Many people claim to find that it is self-evident that there is one and only one way in which one could conceivably become aware of any simple characteristic. You must be presented either in sense-perception or in introspection with an instance which manifests that characteristic to you either sensibly or introspectively as the case may be. An instance of the former is the way in which yellowness is presented to one when one looks at the sun at midday; and an instance of the latter is the way in which the fear-quality is presented to one when one is consciously feeling afraid of a fierce dog. If this principle be accepted, it follows at once that there can be no *simple* epistemologically non-natural characteristics. And it follows at the next move that there can be no complex characteristics containing any simple epistemologically non-natural characteristic as an element. So anyone who finds that he cannot doubt this principle (which might be called "Hume's" principle") must reject the two following views about the words "good" and "bad". He must deny (i) that either of them is the name of a *simple* epistemologically non-natural characteristic, and (ii) that either of them is the name of a *complex* characteristic which contains, as one of its simple elements, an epistemologically non-natural characteristic.

(2) A second general epistemological principle which many people claim to find self-evident is the following. There can be no necessary *synthetic* truths. Any proposition which is necessarily true must be analytic, i.e. the predicate-term must be part of the meaning or analysis of the subject-term, as in the proposition "All negros are black". Now suppose that "morally good" and "morally bad", as applied to experiences, were names of non-natural characteristics. Consider any sentence which appears to assert a universal connexion between some good-making or bad-making natural characteristic and moral goodness or badness. An example would be "Any experience of feeling

pleasure at the thought of another's pain is morally bad". If "morally bad" stands for a non-natural characteristic this sentence cannot possibly state an *analytic* proposition. For the property of being a feeling of pleasure at the thought of another's pain is epistemologically *naturalistic*. It contains nothing but psychological terms, which we got to know about by introspection, together with the notion of causation. So it cannot contain, as part of its *meaning* or *analysis*, the notion of moral badness, if that be epistemologically *non-naturalistic*. It contains nothing but psychological terms, which we got to know about by introspection, together with the notion of causation. So it cannot contain, as part of its *meaning* or *analysis*, the notion of moral badness, if that be epistemologically *non-naturalistic*. Therefore anyone who accepted this epistemological principle would be forced to draw the following hypothetical conclusion. *If* the words "good" and "bad" in such sentences as these stand for a non-natural characteristic, *then* the propositions which such sentences state must be *contingent empirical* generalisations, like "All cloven-footed animals are ruminants". Now it does not seem at all plausible to hold that such sentences state mere contingent generalisations. Suppose that a person who accepts the principle that all necessary truths must be analytic is not prepared to swallow that conclusion. Then he will have to hold either (i) that such sentences do not state propositions at all because the words "good" and "bad" do not really stand for any characteristic whatever; or (ii) that the words "good" and "bad", as used in such sentences, stand for characteristics which are epistemologically *natural*. On the latter alternative it is at least *possible* that such sentences might state propositions which are analytic, though it is not of course *necessary* that they should do so.

Speaking for myself, I do not find either of these epistemological principles self-evident. And if I did, I should be inclined to feel doubts about them when I came to consider some of their implications. Therefore I could not myself draw the conclusions which I have stated above. But anyone who does accept these principles is entitled, and indeed committed, to draw the conclusions which I have shown to be entailed by these premisses.

1.4. Naturalistic theories

The only reasonable alternative to the view that goodness is a non-natural characteristic is that it is a complex natural characteristic. Such theories may be called *naturalistic*. They may take many different forms according to the different natural characteristics which are supposed to be involved in the analysis of goodness. But, from a philosophical point of view, the following divisions are the most important.

(1) Goodness may be indentified either with a complex *intrinsic* natural characteristic or with a complex *extrinsic* natural characteristic. On the first

alternative, to say that a thing is good would be to say that it has a certain complex natural quality or a certain kind of internal structure. I suppose that people who have identified goodness of character with a certain kind of harmony or balance in one's dispositions, actions and experiences hold a naturalistic theory of the first kind. On the second alternative, to say that a thing is good would be to say that it stands in a certain kind of natural relation to some other thing or person or class of things or persons. An example of this kind of theory would be Hume's view to call a thing good means that all or most men who contemplated it would feel an emotion of approval towards it.

I think it is important to see that a naturalistic theory of good need not make goodness an extrinsic characteristic of good things.

(2) Naturalistic theories which make goodness an extrinsic characteristic of good things may be divided into psychological and non-psychological, according as to whether the relational property which is identified with goodness does or does not involve a psychological relation or term. Hume's theory is a psychological form of naturalism, since it defines goodness in terms of a certain emotional relation. But suppose a good action were defined as one which tends to keep society stable or to make it more complex. Then goodness would be an extrinsic natural characteristic, but it would not be a psychological characteristic.

(3) Theories which identify goodness with some extrinsic psychological characteristic are sometimes called "subjective" theories of goodness. The word "subjective" is so terribly ambiguous that it is probably best to avoid using it altogether. But, as other people use it, we must point out its ambiguities. In the widest sense it means the same as "psychological". In its narrowest sense a judgement is called "subjective" if it is a judgment by a person that he is now having a certain experience. The statement "I am feeling cold now" would express a judgement of this kind. In a slightly wider sense a judgement by a person that he generally has a certain kind of experience in a certain kind of situation would be called "subjective". The statement "I dislike the smell of apples" would express a judgement of this kind. If a person makes a judgement about the experiences or dispositions of a class of people, which may or may not include himself, his judgement is not subjective in any but the widest sense. E.g. the statement "Most children like chocolate" expresses a judgement which is subjective only in the sense that it is psychological.

We might call judgements like "I am feeling cold", and "I dislike the smell of apples" *intra-subjective*. And we might call judgements like "Most children like chocolate" *trans-subjective*. A person is not likely to be mistaken in his intra-subjective judgements. And if he is mistaken, it is almost impossible for anyone else to show that he is. But trans-subjective judgements can be supported or refuted by providing favourable or unfavourable statistical evidence. We see then that psychological naturalistic theories of goodness

may be divided into intra-subjective and trans-subjective. Hume's theory is trans-subjective. But suppose it were held that, when *I* call a thing good, I mean simply that *I* feel a certain kind of emotion towards it, and that when *you* call a thing good, you mean simply that *you* feel a certain kind of emotion towards it. That would be an example of an intra-subjective psychological theory about the nature of goodness.

1.41. Extrinsic psychological naturalism

Now it seems to me that the only naturalistic theories of goodness which are plausible enough to be worth considering are those which make goodness to be an extrinsic psychological characteristic of things that are good. Obviously there is a kind of emotion which most people feel from time to time which may be called "moral approval" or "moral disapproval". And it is doubtful whether a person would be able to make or understand a moral judgement if he had never felt this kind of emotion. It is therefore plausible to suggest that, if goodness can be analysed into purely natural characteristics, a reference to this kind of emotion will be an essential factor in its analysis.

We will now consider the intra-subjective and the trans-subjective forms of psychological naturalism in turn.

The commonest and most plausible arguments against the intra-subjective form are the following.

(i) According to this theory, when I call anything "good" I am asserting that I am feeling a certain kind of emotion towards it. When I call anything "bad" I am asserting that I am feeling towards it a certain opposite kind of emotion. Now introspection seems to show that I may be quite convinced that something is extremely good or extremely bad without at the time feeling any strong emotion towards it. (ii) Suppose that I call certain experience "good" and that you call the same, or a precisely similar, experience "bad". Then, if this theory is true, there is no conflict of opinion between us. There is no inconsistency between my feeling an emotion of a certain kind towards a certain object, and *your* feeling an emotion of the opposite kind at the same time towards the same or a precisely similar object. Yet it does seem that our opinions conflict, as they would if I were to call a certain thing "black" and you were to call it "white". (iii) Not only do ethical opinions appear to conflict. People who disagree in their ethical judgements often try to persuade each other by arguments. Now how could you possibly persuade me by argument that I am not feeling a certain kind of emotion towards a certain object, but am feeling the opposite kind of emotion towards it? It seems almost incredible that I could make this sort of mistake. And, if I had made it, what possible line of argument could you use to persuade me that I had?[1]

1. In the enlarged version of the section on naturalistic theories which for reasons explained in my Preface I have discarded, Broad considers also those theories according to which ethical sentences do not express propositions, true or false, but are "interjectional" or "evocative" or "imperative". (Cf. Section 1:311 of the present Chapter.) He points out, however, that the objections stated in the above paragraph apply, *mutatis mutandis*, to theories of this type, too.

It must be noticed that none of these objections are relevant to the trans-subjective form of psychological naturalism. According to this theory when I call *X* "good" I am asserting that all or most men, or all or most members of a certain class of men, feel a certain kind of emotion when they contemplate the occurrence of *X*. Obviously I can make such a statement without feeling this emotion myself; just as I can believe that most people like butter, though I personally dislike it. And obviously there can be real conflict of opinion about ethical questions, and people can try to persuade each other by argument.

The *prima facie* objections to the trans-subjective form of psychological naturalism are the following. (i) Suppose that, when an ancient Athenian called *X* "good", he meant that most contemporary Athenians would contemplate the occurrence of *X* with approval. Suppose that, when a modern Englishman calls *X* "bad", he means that most contemporary Englishmen would contemplate the occurrence of *X* with disapproval. Then their opinions would not conflict. They may both be true. And yet we should be inclined to think that the opinions which they are expressing do conflict, and that one at least of them must be false. This difficulty could be avoided only if each person who calls *X* "good" means that the majority of the human race throughout all the ages would contemplate the occurrence of *X* with approval. But, if this be the meaning, it seems doubtful whether any judgement of the form "*X* is good" or "*X* is bad" is true. And it is difficult to see why anyone should have thought that he had any ground for believing any such sweeping historical generalisations. (ii) If the trans-subjective form of psychological naturalism were true, there would be one absolutely conclusive method of proving or disproving any ethical statement. Suppose that, when I call *X* "good", I mean that most contemporary Englishmen contemplate the occurrence of *X*, or of events exactly like *X*, with approval. Then the truth or falsity of my statement can be settled conclusively by taking a census, and finding what proportion of contemporary Englishmen do in fact habitually feel approval when they contemplate the occurrence of *X* or *X*-like acts or experiences. If my opponents can show me that less than 50% do so, I must admit at once that I was mistaken in thinking that *X* is good. If I can show my opponents that more than 50% do so, they must admit at once that they were mistaken, and that I was correct in my opinion that *X* is good. Now *I* should certainly think that the results of such a census, so far from being conclusive, would be almost completely irrelevant to the truth or falsity of *my* opinion that *X* is good or that *X* is bad. And I fancy that each man would think that the results of such a census would be almost completely irrelevant to the truth or falsity of *his* opinion that *X* is good or that *X* is bad.

There remains one objection which is common to both forms of psychological naturalism. I will defer it until I have stated the intra-subjective theory

in a form which, I think, will remove the first objection to it. The first objection was that I may be quite convinced that *X* is very good or very bad without at the time feeling any strong emotion of approval or disapproval towards *X*. In dealing with this objection we must notice the following points. (a) We must distinguish between genuine first-hand moral judgements and second-hand or conventional ones. In the latter we use sentences as parrots might do, and we are not expressing actual judgements by them at all. (b) We must distinguish between "good" in its various secondary senses, such as "benefic", and in its primary sense. We are concerned here only with first-hand judgements, in which "good" and "bad" are used in the primary sense. (c) The intra-subjective theory need not hold that, when I judge *X* to be good, I am stating that, at this very moment, I am feeling an emotion of approval towards *X*. A much more plausible view would be the following. I am expressing the belief that, if I contemplate the occurrence of *X* when I am in a state of normal bodily health and free from any specially perturbing emotion, such as lust, jealousy, anger, fear etc., my cognition of *X* will always be toned with approval or with disapproval, as the case may be. Now I may have reason to believe this proposition about my own emotional dispositions even when I am not seriously contemplating the occurrence of *X* at all. And I may have reason to believe it on occasions when I am contemplating the occurrence of *X* in a state of illness or lust or jealousy or anger, which suppresses or reverses the emotion of approval or disapproval that I should normally feel. To take a parallel case. When I am bilious the thought of any very rich food is disgusting. But I know even then that, in my normal state of health, I enjoy *pâté de foie gras*. I think that the intra-subjective theory, when stated in this form, is immune to the first objection. So I shall henceforth assume that it is to take this form.

I will now mention an objection which might be made to both the intra-subjective and the trans-subjective forms of psychological naturalism. The ordinary person thinks that, when he feels approval towards *X*, an essential factor in causing him to feel this emotion towards this object is his belief that *X* is good. Similarly, when he feels disapproval towards *Y*, an essential factor in causing him to feel this emotion towards this object is his belief that *Y* is bad. Now on the intra-subjective theory his belief that *X* is good is his belief that he will normally feel approval on contemplating the occurrence of *X*. Thus on the intra-subjective theory the ordinary person's view comes to this: "An essential factor in causing me to contemplate the occurrence of *X* with approval is my belief that I shall normally contemplate the occurrence of *X* with approval". Now it seems quite clear that this is not what one means. And it is easy to show that it could not possibly be true. My belief that I *shall* normally contemplate *X* with approval cannot begin to exist until I *already have* felt approval towards *X*, or *X*-like things, in the past. And it cannot

have been a factor in causing those emotions of approval towards X which occurred before it had begun to exist.

Let us now consider how the trans-subjective theory would deal with this common sense belief. Suppose that my belief that X is good is my belief that most contemporary Englishmen would contemplate the occurrence of X with approval. Then the common sense belief would come to this: "An essential factor in causing *me* to contemplate the occurrence of X with approval is my belief that most contemporary Englismen would contemplate such events with approval". Now it is quite likely that *some* of my emotions of approval are caused in this way. Most of us like to share the emotions of our contemporaries; and so my belief that most of my contemporaries feel approval of X may call forth a similar emotion towards X from me. But it is quite incredible that all or most emotions of approval in all or most people should be caused by the belief that other people feel similar emotions.

1.5. The descriptive theory

I will now give what seems to me to be, on the whole, the most satisfactory account of what I mean when I make a first-hand judgement of the form "This is good" and "That is bad", when "good" or "bad" are used in the primary sense. (i) It seems to me that I never use the words "good" or "bad" in a certain sense in which I often use the words "hot" or "cold". I am acquainted from time to time with particulars which sensibly manifest to me a peculiar determinable characteristic, viz. sensible temperature. Some of them manifest this determinable in a certain determinate form, and I give the name "hotness" to this. Others manifest this determinable in an opposed determinate form, and I give the name "coldness" to this. Thus, the words "hot" and "cold", when used by me in their primary non-dispositional sense, are used as proper names of two determinate forms of a certain determinable characteristic which I know by acquaintance. I also use them in a secondary and dispositional sense, which is definable in terms of the primary sense. If I come into a room, and say "It feels cold", I am using "cold" in the primary sense. If I say "It *is* cold", I am using it in the secondary sense. I mean at least to express the opinion that it will feel cold to any normal person who enters it. Now, as far as I can see, I never use the word "good" as a proper name for a peculiar kind of characteristic which I know by acquaintance. Therefore I never use it in the *non*-dispositional sense in which I often use the word "hot". It follows that I also never use it in the *dispositional* sense in which I often use the word "hot". For the dispositional sense presupposes the non-dispositional sense, and is definable in terms of it. (ii) On the other hand, I am introspectively acquainted from time to time with cognitions which are emotionally toned with approval or emotionally toned with disapproval.

Thus, I can and do use the words "toned with approval" or "toned with disapproval" as proper names for a pair of opposed psychological characteristics with which I am acquainted. In terms of these characteristics I can define and understand the phrases "contemplated by me with approval under normal conditions" or "contemplated by most Englishmen with approval under normal conditions". (iii) I find that the things that I normally contemplate with approval are extremely various, and that I contemplate different members of this class with different degrees of approval. I find that the things that I normally contemplate with disapproval are also extremely various, and that I contemplate different members of this class with different degrees of disapproval. (iv) I suppose or assume or take for granted that there must be a certain characteristic or set of characteristics common and peculiar to the members of the first class, which is an essential cause-factor in causing me to contemplate them and only them with approval. I suppose that this must be present in various degrees in various members of the class to explain the various degrees of approval with which I contemplate them. I make a precisely similar assumption, *mutatis mutandis*, about the class of things which I normally contemplate with disapproval. (v) Now approval and disapproval are opposed forms of a certain determinable emotional quality. I therefore suppose that the characteristic which I have assumed as the cause of my approval of the things which I approve is opposed in a similar way to the characteristic which I have assumed as the cause of my disapproval of the things which I disapprove. (vi) It follows that the judgement which I express by the sentence "*X* is good" is really a descriptive judgement. It would be expressed more accurately by the following sentence "There is one and only one characteristic or set of characteristics whose presence in any object that I contemplate is necessary to make me contemplate it with approval, and *X* has that characteristic". A precisely similar analysis would be given, *mutatis mutandis*, of the judgement which I should express by the sentence "*Y* is bad".

I am going to call this the *descriptive theory* of the meaning of ethical judgements. The first point to notice about it is that, in itself, it is neither naturalistic nor non-naturalistic. It says that we can think of goodness and badness only descriptively, and it says that the descriptions by which we think of them are in terms of a natural psychological characteristic. But it leaves open the question whether there is a characteristic answering to the description or not. And it leaves open the question whether, if there is such a characteristic, it is natural or non-natural. It is quite possible that a characteristic which is *non-natural* could be thought of only as the characteristic which answers to a certain description in which all the terms are *natural*.

In order to show that the descriptive theory can deal with the objections which were brought against psychological naturalism it is necessary to take

into account the following psychological facts. (i) In the main, people of the same race and class and period approve the same things and disapprove the same things, though there are considerable divergences between any two such people in detail. (ii) As Gallie points out ("Oxford Moralists", *Philosophy*, No. 27), each of us wants others to approve what he approves and to disapprove what he disapproves. (iii) Each of us wants to approve what is approved by those whom he likes and respects, and to disapprove what they disapprove. (iv) Each of us is capable of reflecting on his own approvals and disapprovals severally, and contemplating each with reflexive approval or disapproval. Now each of us wants to be able to approve reflexively of his own first-order approval and disapprovals. (v) Lastly, each of us contemplates his own beliefs, desires, and emotions collectively as forming a system. He contemplates himself with approval, in so far as they seem consistent; and with disapproval, in so far as he finds them inconsistent. We will now consider the theory in the light of the various objections

(1) It is quite plain that, if the descriptive theory is true, I can judge a thing to be good without at the time actually contemplating it with approval. Suppose, e.g., that other people whose approvals and disapprovals generally agree with mine tell me that some experience which I have never had is good. I shall have reason to expect that I should approve of it if I had it. And therefore I shall have reason to believe that it has the quality whose presence in objects is necessary to make me contemplate them with approval if I contemplate them at all. And that is what I mean by calling a thing "good", on the descriptive theory. Suppose that I eventually have a certain experience, which people have told me is good, and that I find myself feeling neither approval nor disapproval for it. I could still admit that it may be good. It may have the characteristic whose presence in an object is *necessary* to make me feel approval for it. But some other necessary conditions may not have been fulfilled in this particular case or on this particular occasion, and so my feeling of approval has been inhibited. Suppose finally that, when I eventually have this experience, I find myself feeling disapproval for it. Then, on the present theory, I could not call it good and I should probably call it bad. But before doing so, I should have to make sure that my emotion really was disapproval and not some other emotion which is rather like it or is often associated with it such as aesthetic disgust or a superstitious fear of punishment. Now it seems to me that in actual fact we judge in all these alternative cases just as we might be expected to judge if the descriptive theory were true.

(2) In what sense can ethical judgments conflict if the descriptive theory is true? Suppose I judge that *X* is good, and you judge that *X* is bad. I mean that *X* has the charactertistic whose presence in objects which *I* contemplate is necessary to make *me* contemplate them with approval. You mean that *X* has the characteristic whose presence in objects which *you* contemplate is neces-

sary to make *you* contemplate them with *dis*approval. Now it is true that these two judgments, taken by themselves, do not directly conflict. But, taken together with the fact that in the main you and I agree in our approvals and our disapprovals, they do raise a problem for both of us. Let us denote the one characteristic whose presence in an object is necessary to make A contemplate it with approval by Φ_A. Let us denote the one characteristic whose presence in an object is necessary to make A contemplate it with disapproval by Ψ_A. Then Φ_A and Ψ_A are incompatible. Similarly let us denote the one characteristic whose presence in an object is necessary to make B contemplate it with approval by Φ_B. Let us denote the one characteristic whose presence in an object is necessary to make B contemplate it with disapproval by Ψ_B. Then Φ_B and Ψ_B are incompatible. Now the predominant agreement between the approvals of A and B will have made it highly probable that Φ_A and Φ_B are two names for one and the same characteristic Φ. And the predominant agreement between the disapprovals of A and B will have made it highly likely that Ψ_A and Ψ_B are two names for one and the same characteristic Ψ. But now arises a case where A judges X to be good and B judges X to be bad. Now on descriptive theory, if A is correct in this judgment, X has Φ_A. And if B is correct in his judgment X has Ψ_B. But in that case it is impossible that Φ_A and Φ_B should be identical. For then X would have both Φ_B and Ψ_B, which is impossible. Similarly it is impossible that Ψ_A and Ψ_B should be identical. For then X would have both Φ_A and Ψ_A, which is impossible. Yet the predominant agreement between the ethical judgments of A and B has very strongly suggested that $\Phi_A = \Phi_B$ and that $\Psi_A = \Psi_B$. Thus, although is is *logically possible,* on the descriptive theory, for A's and B's judgments to be both true, yet they can both be true only if we are prepared to reject something which the predominant agreement between A and B has rendered very probable. We shall therefore be inclined in such cases to begin by trying to show that either A or B or both of them is mistaken about the goodness or badness of X.

(3) This brings us to the question "What is implied by the fact that people try to persuade each other by argument to alter their ethical judgments?" (a) I have just shown that, when two people who generally agree in their approvals and disapprovals, make the judgment "X is good" and "X is bad" respectively, there is a real intellectual problem to be solved, even on the descriptive theory. (b) Apart from this, there is the fact that each of us wants others to approve what he approves and to disapprove what he disapproves. And there is the fact that each of us wants to approve what is approved, and to disapprove what is disapproved, by those whom he likes and respects. There are thus ample motives for employing and listening to arguments directed to altering a person's ethical judgments. (c) Arguments with other people about their ethical opinions always take the following forms. (i) We may first try to make sure that both of us are using ethical terms in the same

sense. Perhaps one of us is using "good" in some secondary sense, like "benefic", and the other is using it in the primary sense or in some other secondary sense, such as "contributively good".
(ii) Next we should try to make sure that we are referring to the same, or to precisely similar, things when we both say that we are judging about *X*. And we shall have to make sure that we have exactly similar beliefs about the non-ethical qualities, relationships, and dispositions of *X*. (iii) Then we should consider whether either of us has been mistaking some other emotion, such as aesthetic disgust or superstitious fear, for moral disapproval. And we should raise a similar question about moral approval. In considering this question we should enquire whether either of us is influenced by some special emotional bias, either occasional or dispositional, in connexion with such objects as *X*. (iv) Lastly, we should consider the consistency or inconsistency of our judgments about *X* with our other ethical judgments. Suppose it could be shown that I, who judge *X* to be good, judge other things, which I admit differ in no relevant respect from *X*, to be bad. And suppose it could be shown that my opponent, who judges *X* to be bad, has no such inconsistency in his system of ethical judgments. This would be counted as a point against my judgment and in favour of his. (v) Beyond this no further argument is possible except to count heads. One of us may be able to show that most men agree with him in their judgment about *X*, and that his opponent is alone in his opinion or is in a very small minority.

Now all these arguments could reasonably be used if the descriptive theory of the meaning of ethical judgments were true. And no other arguments could be used even if the most extreme form of objective theory, such as Moore's were true. Suppose that goodness were a non-natural intrinsic characteristic whose presence in an object can be intuited, and suppose that badness were an opposite characteristic of the same kind.
Suppose *A* thinks that he intuits the presence of goodness in *X*, and *B* thinks that he intuits the presence of badness in it. When all the possible sources of disagreement which I have mentioned have been removed there is nothing further to be done except to count heads and to see whether those who agree with *A* or those who agree with *B* are in a majority.

Before leaving this subject there is one point which I want to make quite clear. The question which we have been discussing is: What does a person *mean* when he makes the kind of judgment which would naturally be expressed by uttering the sentence "*X* is good"? To this, it seems to me, the descriptive theory gives a satisfactory answer. Now there is an entirely different question which can be raised at this point: Supposing that your account of what people mean by such judgments were correct, is there any reason to believe that any such judgments are true? All that I need say about this second question is the following. If the descriptive theory is correct, then every judg-

ment of the form "*X* is good" that I have ever made has been false unless there is one and only one characteristic or set of characteristics whose presence in any object that I contemplate is necessary to make me contemplate it with approval. Similarly every judgment of the form "*X* is bad" that I have ever made has been false unless there is one and only one characteristic or set of characteristics whose presence in any object that I contemplate is necessary to make me contemplate it with disapproval.

Chapter 5

METAPHYSICS OF MORALS

1. Determinism, indeterminism, and libertarianism

1.1. Obligability and substitutability

Judgements of obligation about past actions may be divided into two classes, viz. (a) judgments about actions which were actually done, and (2) judgments about conceivable actions which were not done. Each of these divides into two sub-classes, and so we get the following fourfold division. (1.1) You did X, and X was the action that you ought to have done. (1.2) You did X, and X was an action that you ought not to have done. (2.1) You did not do X, and X was the action that you ought to have done. (2.2) You did not do X, and X was an action that you ought not to have done.

The common phrase "You ought to have done so-and-so" and "You ought not to have done so-and-so" are generally equivalent to our (2.1) and (1.2) respectively. For the former is generally used to mean "You did not do so-and-so", but that was the action that you ought to have done". The latter is generally used to mean "You did so-and-so, but that was an action which you ought not to have done". But our judgments (1.1) and (2.2) are not superfluous. We sometimes want to say that a person did what he ought on a certain occasion; and that is expressed by (1.1). And we sometimes want to say that a person omitted to do something which he ought not to have done on a certain occasion. For this is exactly the state of affairs which exist when a person has rejected a course of action which is in other respects strongly attractive to him, but which would have been morally wrong.

Now both judgments of the first class entail that you could in some sense have avoided doing what you in fact did. If the action which you did can be said to be one that you ought to have done, or if it can be said to be one that you ought not to have done, it must be one that you *need not* have done. Both judgments of the second class entail that you could in some sense have done an action which you did not in fact do. If a conceivable action which you did not do can be said to be one which you ought to have done, or if it can be said to be one which you ought not to have done, it must be one that you *could have* done.

We will call an action "obligable" if and only if it is one concerning which it is sensible to say that it "ought to have been done" or that it "ought not to

have been done". We will call an action "substitutable" if and only if either it was done but could have been left undone or was left undone but could have been done. We may then sum up the situation as follows. An action is obligable if and only if it is, in a certain sense, substitutable. Unless all judgments of obligation are false in principle and not merely in detail, there are obligable actions. Therefore, unless all judgments of obligation are false in principle, there are actions which are, in this sense, substitutable.

At this stage two problems arise. (i) Can we discover and state the sense of "substitutable" in which being substitutable is a necessary condition of being obligable? (ii) If we can do this, can it be admitted that any action is substitutable in this sense? The first may be called a question of *analysis*. The second may be called a question about *logical and empirical possibility*.

1.2. Various senses of "substitutable"

There are several senses of "could" in which nearly everyone would admit that some actions which were done could have been left undone and that some conceivable actions which were left undone could have been done. There are thus several senses of "substitutable" in which it would generally be admitted that some actions are substitutable. But it seems doubtful whether an action could be obligable if it were substitutable *only* in these senses. Let us now consider the various senses.

1.21. Voluntary substitutability

Let us begin by considering an action which has actually been performed. In some cases we should say that the agent "could not have helped" doing it. We should certainly say this if we had reason to believe that the very same act would have been done by the agent in these circumstances even though he had willed that it should not take place. It is obvious that there are actions which are "inevitable" in this sense, since there are actions which take place although the agent is trying his hardest to prevent them. Cf., e.g., the case of a conspirator seized with an uncontrollable fit of sneezing.

Next consider a conceivable action which was not in fact done. In some cases we should say that the agent "could not possibly" have done it. We should certainly say this if this conceivable action would not have taken place in these circumstances no matter how strongly the agent had willed it. It is obvious that there are conceivable acts which are "impossible" in this sense, since there are cases where such an act fails to take place although the agent is trying his hardest to bring it about. Cf., e.g., the case of a man who is bound and gagged, and tries vainly to give warning to a friend.

We will call acts of these two kinds "not voluntarily substitutable". It is plain that an act which is not voluntarily substitutable is not obligable. No

one would say that the conspirator ought not to have sneezed, or that the bound and gagged man ought to have warned his friend. At most we may be able to say that they ought or ought not to have done certain things in the past which are relevant to their present situation. Perhaps the conspirator ought to have sprayed his nose with cocaine before hiding behind the presumably dusty arras, and perhaps the victim ought not to have let himself be lured into the house in which he was gagged and bound. But these are previous questions.

We see then that to be voluntarily substitutable is a *necessary* condition for an action to be obligable. But is it a *sufficient* condition? Suppose I performed the action A on a certain occasion. Suppose that I should not have done A then if I had willed with a certain degree of force and persistence not to do it. Since I did A, it is certain that I *did not* will with this degree of force and persistence to avoid doing it. Now suppose that at the time I *could not* have willed with this degree of force and persistence to avoid doing A. Should we be prepared to say that I ought not to have done A?

Now take another case. Suppose that on a certain occasion I failed to do a certain conceivable action B. Suppose that I should have done B if I had willed with a certain degree of force and persistence to do it. Since I did not do B, it is certain that I *did not* will with this degree of force and persistence to do it. Now suppose that at the time I *could not* have willed with this degree of force and persistence to do B. Should we be prepared to say that I ought to have done B? It seems to me almost certain that, under the supposed conditions, we should not be prepared to say either that I ought not to have done A or that I ought to have done B.

Consider, e.g., the case of a man who gradually becomes addicted to some drug like morphine, and eventually becomes a slave to it. At the early stages we should probably hold that he could have willed with enough force and persistence to ensure that the temptation would be resisted. At the latest stages we should probably hold that he could not have done so. Now at every stage, from the earliest to the latest, the hypothetical proposition would be true "If he had willed with a certain degree of force and persistence to avoid taking morphine, he would have avoided taking it". Yet we should say at the earlier stages that he ought to have resisted, whilst, at the final stages, we should be inclined to say that "ought" and "ought not" have ceased to apply.

1.211. Primary and secondary substitutability

An action which was in fact done, but would not have been done if there had been a strong and persistent enough desire in the agent not to do it, will be called "primarily avoidable". Suppose in addition that there could have been in the agent at the time a desire of sufficient strength and persistence to prevent the action being done. Then the action might be called "secondarily

avoidable''. If this latter condition is not fulfilled, we shall say that the action was ''primarily avoidable, but secondarily inevitable''. Similarly, an action which was not in fact done, but would have been done if there had been in the agent a strong and persistent enough desire to do it, will be called ''primarily possible''. Suppose in addition that there could have been in the agent at the time a desire of sufficient strength and persistence to ensure the action being done. Then the action may be called ''secondarily possible''. If this latter condition is not fulfilled, we shall say that the action is ''primarily possible, but secondarily impossible''. An action will be called ''primarily substitutable'' if it is either primarily avoidable or primarily possible. It will be secondarily substitutable if it is either secondarily avoidable or secondarily possible. In order that an action may be obligable it is not enough that it should be primarily substitutable, it must be at least secondarily substitutable.

We are thus led on from the notion of voluntarily substitutable *actions* to that of substitutable *volitions*. Suppose that, on a certain occasion and in a certain situation, a certain agent willed a certain alternative with a certain degree of force and persistence. We may say that the volition was substitutable if the same agent, on the same occasion and in the same circumstances, could instead have willed a different alternative or could have willed the same alternative with a different degree of force and persistence. Now there is one sense of ''could'' in which it might plausibly be suggested that many volitions are substitutable. It seems very likely that there are many occasions on which I *should* have willed otherwise than I did, *if* on previous occasions I had willed otherwise than I did. So it seems likely that many volitions have been voluntarily substitutable.

It is necessary to be careful at this point, or we may be inadvertently granting more than we are really prepared to admit. Obviously it is often true that, if I had willed otherwise than I did on certain earlier occasions, I should never have got into the position in which I afterwards made a certain decision. If, e.g., Julius Caesar had decided earlier in his career not to accept the command in Gaul, he would never have been in the situation in which he decided to cross the Rubicon. This, however, does not make his decision to cross the Rubicon substitutable. For a volition is substitutable only if a different volition could have occurred in the agent in the *same* situation. Again, it is often true that, if I had willed otherwise than I did on certain earlier occasions, my state of knowledge and belief would have been different on certain later occasions from what it in fact was. In that case I should have thought, on these later occasions, of certain alternatives which I did not and could not think of in my actual state of knowledge and belief. Suppose, e.g., that a lawyer has to decide what to do when a friend has met with an accident. If this man had decided years before to study medicine instead of law, it is quite likely that he would now think of, and perhaps choose, an alternative which his lack of

medical knowledge prevents him from contemplating. This, however, does not make the lawyer's volition in the actual situation substitutable. For, although the external part of the total situation might have been the same whether he had previously decided to study medicine or to study law, the internal part of the total situation would have been different if he had decided to study medicine, instead of deciding, as he did, to study law. He would have become an agent with different cognitive powers and dispositions from those which he in fact has. No one would think of saying that the lawyer ought to have done a certain action, which he did not and could not contemplate, merely because he would have contemplated it and would have decided to do it if he had decided years before to become a doctor instead of becoming a lawyer.

Having cleared these irrelevances away, we can now come to the real point. A man's present conative-emotional dispositions, and what we may call his "power of intense and persistent willing", are in part dependent on his earlier volitions. If a person has repeatedly chosen the easier of the alternatives open to him, it becomes increasingly difficult for him to choose and to persist in pursuing the harder of two alternatives. If he has formed a habit of turning his attention away from certain kinds of fact, it will become increasingly difficult for him to attend fairly to alternatives which involve facts of these kinds. This is one aspect of the case. Another, and equally important, aspect is the following: If a man reflects on his own past decisions, he may see that he has a tendency to ignore or to dwell upon certain kinds of fact, and that this has led him to make unfair or unwise decisions on many occasions. He may decide that, in future, he will make a special effort to give due, and not more than due, weight to those considerations which he has a tendency to ignore or to dwell upon. And this decision may make a difference to his future decisions. On the other hand, he may see that certain alternatives have a specially strong attraction for him, and he may find that, if he pays more than a fleeting attention to them, he will be rushed into choosing them, and will afterwards regret it. He may decide that, in future, he will think as little as possible about such alternatives. And this decision may make a profound difference to his future decisions.

We can now state the position in general terms. Suppose that, if the agent had willed differently on earlier occasions, his conative-emotional dispositions and his knowledge of his own nature would have been so modified that he would now have willed differently in the actual external situation and in his actual state of knowledge and belief about the alternatives open to him. Then we can say that his actual volition in the present situation was "voluntarily avoidable", and that a volition of a different kind or of a different degree of force and persistence was "voluntarily possible". An action which took place was secondarily avoidable if the following two conditions are fulfilled. (i) That this action would not have been done if the agent had willed with a cer-

tain degree of force and persistence to avoid it. (ii) That, if he had willed differently in the past, his conative-emotional dispositions and his knowledge of his own nature would have been such, at the time when he did the action, that he would have willed to avoid it with enough force and persistence to prevent him doing it. In a precisely similar way we could define the statement that a certain conceivable action, which was not done, was secondarily possible. And we can thus define the statement that an action is secondarily substitutable.

Can we say that an action is obligable if it is secondarily substitutable, in the sense just defined, though it is not obligable if it is only primarily substituable? It seems to me that the same difficulty which we noticed before reappears here. Suppose that the agent could not have willed otherwise than he did in the remoter past. It is surely irrelevant to say that, *if* he had done so, his conative dispositions *would* have been different at a later stage from what they in fact were then, and that he *would* have willed otherwise than he then did. One might, of course, try to deal with this situation by referring back to still earlier volitions. One might talk of actions which are not only primarily, or only secondarily, but are tertiarily substitutable. But it is quite clear that this is useless. If neither primary nor secondary substitutability, in the sense defined, suffice to make an action obligable, no higher order of substitutability, in this sense, will suffice. The further moves are of exactly the same nature as the second move. And so, if the second move does not get us out of the difficulty, none of the further moves will do so.

1.22. Categorical substitutability

The kind of substitutability which we have so far considered may be called "conditional subsitutability". For at every stage we have defined "could" to mean "would have been, if certain conditions had been fulfilled which were not". Now I have concluded that merely conditional substitutability, of however high an order, is not a sufficient condition for obligability. If an action is to be obligable, it must be *categorically* substitutable. We must be able to say of an action, which was done, that it could have been avoided, in some sense of "could" which is not definable in terms of "would have, if". And we must be able to say of a conceivable action, which was not done, that it could have been done, in some sense of "could" which is not definable in terms of "would have, if". Unless there are some actions of which such things can truly be said, there are no actions which are obligable. We must therefore consider whether any clear meaning can be attached to the phrase "categorically substitutable", i.e. whether "could" has any clear meaning except "would have, if". And, if we can find such a meaning, we must enquire whether any actions are categorically substitutable.

1.221. Various senses of "obligable"

Before tackling these questions I must point out that the words "ought" and "ought not" are used in several different senses. In some of these senses obligability does not entail categorical substitutability.

(i) There is a sense of "ought" in which we apply it even to inanimate objects. It would be quite proper to say "A car ought to be able to get from London to Cambridge in less than three hours", or "A fountain-pen ought not to be constantly making blots". We mean by this simply that a car which did take more than three hours would be a poor specimen of car, or would be in a bad state of repair. And similar remarks apply to the statement about the fountain-pen. We are comparing the behaviour of a certain car or fountain-pen with the average standard of achievement of cars of fountain-pens. We are not suggesting that *this* car or *this* pen, in its present state of repair, unconditionally could go faster or avoid making blots. Sometimes when we make such judgments we are comparing an individual's achievements, not with those of the *average* member, but with those of an *ideally perfect* member, of a certain class to which it belongs. We will call "ought", in this sense, "the comparative ought". And we can then distinguish "the average-comparative ought" and "the ideal-comparative ought".

(ii) Plainly "ought" and "ought not" can be, and often are, used in this sense of human actions. But, in the case of human actions, there is a further development. Since a human being has the power of cognition, in general, and of reflexive cognition, in particular, he can have an idea of an average or an ideal man. He can compare his own achievements with those of the average, or the ideal, man, as conceived by him. And he will have a more or less strong and persistent desire to approximate to the ideal and not to fall below the average. Now it is part of the notion of an ideal man that he is a being who would have a high ideal of human nature and would desire strongly and persistently to approximate to his ideal. Obviously it is no part of the notion of an ideal horse or an ideal car that it is a being which would have a high ideal of horses or cars and a strong and persistent desire to live up to this. When we say that a man ought not to cheat at cards we often mean to assert two things. (a) That the average decent man does not do this, and that anyone who does falls in this respect below the average. And (b) that a man who does this either has a very low ideal of human nature or a very weak and unstable desire to approximate to the ideal which he has. So that, in this further respect, he falls below the average.

Now neither of these judgments implies that a particular person, who cheated on a particular occasion, categorically could have avoided cheating then; or that he categorically could have had a higher ideal of human nature; or that he categorically could have willed more strongly and persistently to live up to the ideal which he had. For an action to be obligable, in this sense, it

is plainly enough that it should be secondarily substitutable, in the sense already defined.

1.2211. The categorical ought. Some philosophers of great eminence, e.g. Spinoza, have held that the sense of "ought" which I have just discussed is the only sense of it. Plainly it is a very important sense, and it is one in which "ought" and "ought not" can be applied only to the actions of intelligent beings with powers of reflexive cognition, emotion, and conation. I think that a clear-headed determinist should hold either that this is the only sense; or that, if there is another sense, in which obligability entails *categorical* substitutability, it has no application.

Most people, however, would say that, although we often do use "ought" and "ought not" in this sense, we quite often use them in another sense, and that in this other sense they entail categorical substitutability. I am inclined to think that this is true. When I judge that I ought not to have done something which I in fact did, I do not as a rule seem to be judging merely that a person with higher ideals, or with a stronger and more persistent desire to live up to his ideals, would not have done what I did. Even when this is part of what I mean, there seems to be something more implied in my judgment, viz. that I *could* have had higher ideals or *could* have willed more strongly and persistently to live up to my ideals, where "could" does not mean just "would have, if". Let us call this sense of "ought" the "categorical ought". It seems to me then that we must distinguish between an action being obligable in the comparative sense and being obligable in the categorical sense; and that, if any action were categorically obligable, it would have to be categorically substitutable.

1.222. Analysis of categorical substitutability

We can now proceed to discuss the notion of categorical substitutability. It seems to me to involve a negative and a positive condition. I think that the negative condition can be clearly formulated, and that there is no insuperable difficulty in admitting that it may sometimes be fulfilled. The ultimate difficulty is to give any intelligible account of the positive condition. I will now explain and illustrate these statements.

Suppose that, on a certain occasion, I willed a certain alternative with a certain degree of force and persistence, and that, in consequence of this volition, I did a certain voluntary action which I would not have done unless I had willed this alternative with this degree of intensity and persistence. To say that I categorically could have avoided doing this action implies at least that the following negative condition is fulfilled. It implies that the process of my willing this alternative with this degree of force and persistence was not completely determined by the occurrent, the dispositional, the nomic, and the

background conditions which existed immediately before and during this process of willing. In order to see exactly what this means it will be best to contrast it with a case in which we believe that a process is completely determined by such conditions.

Suppose that two billiard-balls are moving on a table, that they collide at a certain moment, and that they go on moving in modified directions with modified velocities in consequence of the impact. Let us take as universal premisses the general laws of motion and of elastic impact. We will call these "nomic premisses". Let us take as singular premisses the following propositions. (i) That each ball was moving in such and such a direction with such and such a velocity at the moment of impact. We will call these "occurrent premisses". (ii) That the masses and coefficients of elasticity of the balls were such and such. We will call these "dispositional premisses". (iii) That the table was smooth and level before, at, and after the moment of impact. We will call this a "background premiss". Lastly, let us take the proposition that the balls are moving directly after the impact in such and such directions with such and such velocities. Then this last proposition is a *logical consequence* of the conjunction of the nomic, the occurrent, the dispositional, and the background premisses. That is to say, the combination of these premisses with the denial of the last proposition would be *logically inconsistent*. It is so in exactly the sense in which the combination of the premisses of a valid syllogism with the denial of its conclusion would be so.

1.2221. The negative condition. We can now work towards a definition of the statement that a certain event e was completely determined in respect of a certain characteristic. When we have defined this statement it will be easy to define the statement that a certain event was not completely determined in respect of a certain characteristic. I will begin with a concrete example, and will then generalise the result into a definition.

Suppose that a certain flash happened at a certain place and date. This will be a manifestation of a certain determinable characteristic, viz. colour, in a certain perfectly determinate form. It may, e.g., be a red flash of a certain perfectly determinate shade, intensity, and saturation. We may call shade, intensity, and saturation the three "dimensions" of colour, and we shall therefore symbolise the determinable characteristic colour by a three-suffix symbol C_{123}. When we want to symbolise a certain perfectly determinate value of this we shall use the symbol C_{123}^{abc}. This means that the shade has the determinate value a, that the intensity has the determinate value b, and that the saturation has the determinate value c. Each *index* indicates the determinate value which the dimension indicated by the corresponding *suffix* has in the given instance.

Now the statement that this flash was completely determined in respect of

colour has the following meaning. It means that there is a set of true nomic, dispositional, occurrent, and background propositions which together entail the proposition that a manifestation of colour, of the precise shade, intensity, and saturation which this flash manifested, happened at the place and time at which this flash happened. To say that this flash was *not* completely determined in respect of colour means that there is *no* set of true nomic, dispositional, occurrent, and background propositions which together entail the proposition that a manifestation of colour, of the precise shade, intensity, and saturation which this flash manifested, happened at the place and time at which this flash happened.

There are two remarks to be made at this point. (i) It seems to me that the second statement is perfectly *intelligible,* even if no such statement be ever true. (ii) It is a purely *ontological* statement, and not in any way a statement about the limitations of our knowledge. Either there is such a set of true propositions, or there is not. There may be such a set, even if no one knows that there is; and there may be no such set, even if everyone believes that there is.

We can now give a general definition. The statement that a certain event e was completely determined in respect of a certain determinable characteristic C_{123} is equivalent to the conjunction of the following two proposition. (i) The event e was a manifestation of C_{123} in a certain perfectly determinate form C_{123}^{abc} at a certain place and date. (ii) There is a set of true nomic, dispositional, occurrent, and background propositions which together entail that a manifestation of C_{123} in the form C_{123}^{abc} happened at the place and date at which e happened. The statement that e was *not* completely determined in respect of C_{123} is equivalent to the conjoint assertion of (i) and denial of (ii).

The next point to notice is that an event might be partly determined and partly undetermined in respect of a certain characteristic. As before, I will begin with a concrete example. Our flash might be completely determined in respect of shade and saturation, but not in respect of intensity. This would be equivalent to the conjunction of the following two statements.
(i) That there is a set of true propositions, of the kind already mentioned, which together entail that a flash, of precisely the shade and saturation which this flash had, happened at the place and date at which this flash happened.
(ii) There is no such set of true propositions which together entail that a flash, of precisely the intensity which this flash had, happened at the time and place at which this flash happened. We thus get the notion of "orders of indetermination" in respect of a given characteristic. If an event is undetermined in respect of one and only one dimension of a certain determinable characteristic, we say that it has "indetermination of the first order" in respect of this characteristic. If it is undetermined in respect of two and only two dimensions of a certain determinable characteristic, we say that it has "indetermination of the second order" in respect of this characteristic. And so on.

It is obvious that there is another possibility to be considered, which I will call "range of indetermination in respect of a given dimension of a given characteristic". Suppose that our flash is undetermined in respect of the intensity of its colour. There may be a set of true propositions, of the kind mentioned, which together entail that a flash, whose intensity falls within certain limits, happened at the time and place at which this flash happened. This range of indetermination may be wide or narrow. Complete determination in respect of a dimension of a given characteristic is the limiting case where the range of indetermination shuts up to zero about the actual value of this dimension for this event. Thus the "extent of indetermination" of an event with respect to a given characteristic depends in general upon two factors, viz. (i) its order of indetermination with respect to the dimensions of this characteristic, and (ii) its range of indetermination with respect to those dimensions for which it is not completely determined.

We can now define the statement that a certain event *e* was completely determined. It means that *e* has zero range of indetermination for every dimension of every determinable characteristic of which it is a manifestation. The statement that a certain event *e* was *not* completely determined can now be defined. It means that *e* had a finite range of indetermination for at least one dimension of at least one of the characteristics of which it was a manifestation.

And now at last we can define "determinism" and "indeterminism". Determinism is the doctrine that *every* event is completely determined, in the sense just defined. Indeterminism is the doctrine that some, and it may be all, events are not completely determined, in the sense defined. Both doctrines are, *prima facie*, intelligible, when defined as I have defined them.

There is one other point to be noticed. An event might be completely determined, and yet it might have a "causal ancestor" which was not completely determined. If Y is the total cause of Z, and X is the total cause of Y, I call both Y and X "causal ancestors" of Z. Similarly, if W were the total cause of X, I should call Y, X, and W "causal ancestors" of Z. And so on. If at any stage in such a series there is a term, e.g. W, which contains a cause-factor that is not completely determined, the series will stop there, just as the series of human ancestors stops with Adam. Such a term may be called the "causal progenitor" of such a series. If determinism be true, every event has causal ancestors, and therefore there are no causal progenitors. If indeterminism be true, there are causal progenitors in the history of the world.

We can now state the negative condition which must be fulfilled if an action is to be categorically substitutable. Suppose that, at a certain time, an agent deliberated between two alternatives, A and B, and that he actually did A and not B. Suppose that the folowing conditions are fulfilled. (i) The doing of A by this agent at this moment was completely determined. (ii) The total cause

of *A* being done contained as cause-factors a desire of a certain strength and persistence for *A* and a desire of a certain strength and persistence for *B*. (iii) These two desires were not completely determined in respect of strength and persistence. (iv) The range of indetermination was wide enough to include in it, as possible values, so strong and persistent a desire for *B* or so weak and fleeting a desire for *A* as would have determined the doing of *B* instead of the doing of A. Conditions (iii) and (iv) are the negative conditions which must be fulfilled if *B* is to be categorically substitutable for A. They amount to the following statement. It is consistent with (a) the laws of nature, including those of psychology, (b) the facts about the agent's dispositions and the dispositions of any other agent in the world at the moment of acting, (c) the facts about what was happening within and without the agent at that moment, and (d) the facts about the general background conditions at that moment, that the strength and persistence of the desires mentioned in (ii) should have any value that falls within the range mentioned in (iv).

Before we go further there is one point to be mentioned. Strictly speaking, what I have just stated are the negative conditions for *primary* categorical substitutability. For I have supposed the incomplete determination to occur at the *first* stage backwards, viz. in one of the cause-factors in the total cause of the action *A*. It would be quite easy to define, in a similar way, the negative conditions for secondary, or tertiary, or any other order of categorical substitutability. All that is needed is that, at *some* stage in the causal ancestry of *A*, there shall be a total cause which contains as factors desires of the agent answering to the condition which I have stated. That is to say, all that is necessary is that *A* shall have a causal ancestor which is a causal progenitor, containing as a factor an incompletely determined desire of the agent's.

We come now to the final question. Supposing that this negative condition were fulfilled, would this *suffice* to make an action categorically obligable? It seems to me plain that it would not. Unless some further and positive condition were fulfilled, all that one could say would be the following. "The desire to do *A* happened to be present in me with such strength and persistence, as compared with the desire to do *B*, that I did *A* and avoided *B*. The desire to do *B* might have happened to be present in me with such strength and persistence, as compared with the desire to do *A*, that I should have done *B* and avoided *A*." Now, if this is all, the fact that I did *A* and not *B* is, in the strictest sense, an accident, lucky or unlucky as the case may be. It may be welcomed or it may be deplored, but neither I nor anything else in the universe can properly be praised or blamed for it. It begins to look as if the categorical ought may be inapplicable, though for different reasons, both on the hypothesis that voluntary actions have causal progenitors and on the hypothesis that none of their causal ancestors are causal progenitors.

1.2222. The positive condition. Let us now try to discover the positive conditions of categorical obligability. I think that we should naturally tend to answer the sort of objection which I have just raised in the following way. We should say "I deliberately identified myself with my desire to do *A*, or I deliberately threw my weight on the side of that desire. I might instead have made no particular effort in one direction or the other; or I might have identified myself with, and thrown my weight on the side of, my desire to do *B*. So my desire to do *A* did not just happen to be present with the requisite strength and persistence, as compared with my desire to do *B*. It had this degree of strength and persistence because, and only because, I *reinforced* it by a deliberate effort, which I need not have made at all and which I could have made in favour of my desire to do *B*". Another way of expressing the same thing would be this "I forced myself to do *A*; but I need not have done so, and, if I had not done so, I should have done *B*". Or again "I might have forced myself to do *B*; but I did not, and so I did *A*".

It is quite plain that these phrases express a genuine positive experience with which we are all perfectly familiar. They are all, of course, metaphorical. It will be noticed that they all attempt to describe the generic fact by metaphors drawn from specific instances of it, e.g. deliberately pressing down one scale of a balance, deliberately joining one side in a tug-of-war, deliberately thrusting a body in a certain direction against obstacles, and so on. In this respect they may be compared with attempts to describe the generic facts about time and change by metaphors drawn from specific instances, such as flowing streams, moving spots of light, and so on. The only use of such metaphors is to direct attention to the sort of fact which one wants one's hearers to contemplate. They give no help towards analysing or comprehending this fact. A metaphor helps us to understand a fact only when it brings out an analogy with a fact of a different kind, which we already understand. When a generic fact can be described only by metaphors drawn from specific instances of itself it is a sign that the fact is unique and peculiar, like the fact of temporal succession and the change of events from futurity, through presentness, to pastness.

Granted that there is this unique and peculiar factor of deliberate effort or reinforcement, how far does the recognition of it helps us in our present problem? So far as I can see, it merely takes the problem one step further back. My doing of *A* is completely determined by a total cause which contains as factors my desire to do *A* and my desire to do *B*, each of which has a certain determinate strength and persistence. The preponderance of my desire to do *A* over my desire to do *B*, in respect of strength and persistence, is completely determined by a total cause which contains as a factor my putting forth a certain amount of effort to reinforce my desire for *A*. This effort-factor is not completely determined. It is logically consistent with all the nomic, disposi-

tional, occurrent, and background facts that no effort should have been made, or that it should have been directed towards reinforcing the desire for *B* instead of the desire for *A*, or that it should have been put forth more or less strongly than it actually was in favour of the desire for *A*. Surely then we can say no more than that it just happened to occur with a certain degree of intensity in favour of the desire for *A*.

I think that the safest course at this stage for those who maintain that some actions are categorically obligable would be the following. They should admit quite frankly what I have just stated, and should then say "However paradoxical it may seem, we do regard ourselves and other people as morally responsible for accidents of this unique kind, and we do not regard them as morally responsible, in the categorical sense, for anything but such accidents and those consequences of them which would have been different if the accidents had happened differently. Only such accidents, and their causal descendants in the way of volition and action, are categorically obligable". If anyone should take up this position, I should not know how to refute him, though I should be strongly inclined to think him mistaken.

This is not, however, the position which persons who hold that some actions are categorically obligable generally do take at this point. I do not find that they ever state quite clearly what they think they believe, and I suspect that this is because, if it were clearly stated, it would be seen to be impossible. I shall therefore try to state clearly what I think such people want to believe, and shall try to show that it is impossible. I suspect that they would quarrel with my statement that, on their view, the fact that one puts forth such and such an effort in support of a certain desire is, in the strictest sense, an accident. They would like to say that the putting forth of a certain amount of effort in a certain direction at a certain time *is* completely determined, but is determined in a unique and peculiar way. It is literally determined *by the agent or self*, considered as a substance or continuant, and not by a total cause which contains as factors *events in* and *dispositions of* the agent. If this could be maintained, our puttings-forth of effort would be completely determined, but their causes would neither be events nor contain events as cause-factors. Certain series of events would then originate from causal progenitors which are continuants and not events. Since the first event in such a series would be completely determined, it would not be an accident. And, since the total cause of such an event would not be an event and would not contain an event as a cause-factor, the two alternatives "completely determined" and "partially undetermined" would both be inapplicable to it. For these alternatives apply only to events.

I am fairly sure that this is the kind of proposition which people who profess to believe in free will want to believe. I have, of course, stated it with a regrettable crudity, of which they would be incapable. Now it seems to me

clear that such a view is impossible. The putting-forth of an effort of a certain intensity, in a certain direction, at a certain moment, for a certain duration, is quite clearly an event or process, however unique and peculiar it may be in other respects. It is therefore subject to any conditions which self-evidently apply to every event, as such. Now it is surely quite evident that, if the beginning of a certain process at a certain time is determined at all, its total cause *must* contain as an essential factor another event or process which *enters into* the moment from which the determined event or process *issues*. I see no *prima facie* objection to there being events that are not completely determined. But, in so far as an event *is* determined, an essential factor in its total cause must be other *events*. How could an event possibly be determined to happen at a certain date if its total cause contained no factor to which the notion of date has any application? And how can the notion of date have any application to anything that is not an event?

Of course I am well aware that we constantly use phrases, describing causal transactions, in which a continuant is named as the cause and no event in that continuant is mentioned. Thus we say "The stone broke the window", "The cat killed the mouse", and so on. But it is quite evident that all such phrases are elliptical. The first, e.g., expresses what would be more fully expressed by the sentence "The coming in contact of the moving stone with the window at a certain moment caused a process of disintegration to begin in the window at that moment". Thus the fact that we use and understand such phrases casts no doubt on the general principle which I have just enunciated.

Let us call the kind of causation which I have just described and rejected "non-occurrent causation of events". We will call the ordinary kind of causation, which I had in mind when I defined "determinism" and "indeterminism", "occurrent causation".

Now I think we can plausibly suggest what may have made some people think they believe that puttings-forth of effort are events which are determined by non-occurrent causation. It is quite usual to say that a man's putting-forth of effort in a certain direction on a certain occasion was determined by "reason" or "principle" or "conscience" or "the moral law". Now these impressive names and phrases certainly do not denote events or even substances. If they denote anything, they stand for propositions or systems of propositions, or for those peculiar universals or systems of universals which Plato called "ideas". If it were literally true that puttings-forth of effort are determined by such entities, we should have causation of events in time by timeless causes. But, of course, statements like "Smith's putting-forth of effort in a certain direction on a certain occasion was determined by the moral law" cannot be taken literally. The moral law, as such, has no causal efficacy. What is meant is that Smith's *belief* that a certain alternative would be in accordance with the moral law, and his *desire* to do what is right,

were cause-factors in the total cause which determined his putting-forth of effort on the side of that alternative. Now this belief was an event, which happened when he began to reflect on the alternatives and to consider them in the light of the moral principles which he accepts and regards as relevant. And this desire was an event, which happened when his conative-emotional moral dispositions were stirred by the process of reflecting on the alternatives. Thus the use of phrases about action being "determined by the moral law" may have made some people think they believe that some events are determined by non-occurrent causation. But our analysis of the meaning of such phrases shows that the facts which they express give no logical support to this belief.

1.3. Libertarianism

We are now in a position to define what I will call "libertarianism". This doctrine may be summed up in two propositions. (i) Some (and it may be all) voluntary actions have a causal ancestor which contains as a cause-factor the putting-forth of an effort which is not completely determined in direction and intensity by occurrent causation. (ii) In such cases the direction and the intensity of the effort are completely determined by non-occurrent causation, in which the self or agent, taken as a substance or continuant, is the non-occurrent total cause. Thus, libertarianism, as defined by me, entails indeterminism, as defined by me; but the converse does not hold.

If I am right, libertarianism is self-evidently impossible, whilst indeterminism is *prima facie* possible. Hence, if categorical obligability entails libertarianism, it is certain that no action can be categorically obligable. But if categorical obligability entails only indeterminism, it is *prima facie* possible that some actions are categorically obligable. Unfortunately, it seems almost certain that categorical obligability entails more than indeterminism, and it seems very likely that it entails libertarianism. It is therefore highly probable that the notion of categorical obligability is a delusive notion, which neither has nor can have any application.

2. Arguments for and against determinism

We can now tackle the second part of our problem, viz. whether there is any good reason for accepting or for rejecting determinism. Since determinism and indeterminism are contradictory opposites, any argument for either is *pro tanto* an argument against the other, and any argument against either is *pro tanto* an argument for the other.

Possible arguments on the subject may be subdivided first into *ethical* and *non-ethical*, and the non-ethical ones can be subdivided into *empirical* and *a*

priori. We will now consider them in that order.

2.1. Ethical arguments

The ethical argument for indeterminism may be put as follows. It is certain that some of our actions are categorically obligable. It is certain that we have done some things which we categorically ought not to have done, and that we have left undone some things that we categorically ought to have done. But any action that is categorically obligable must be categorically substitutable. Therefore there have been some categorically substitutable actions. Now any action which was categorically substitutable must have had a causal ancestor which contained a cause-factor not completely determined by occurrent causation. Therefore there have been events not completely determined by occurrent causation. Therefore indeterminism is true.

I think that this is much the strongest argument for indeterminism. But I do not think that it is conclusive, for the following reason. Although categorical obligability certainly does entail indeterminism, it looks as if this were insufficient. It looks as if it also entailed the non-occurrent causation of certain events which are not completely determined by occurrent causation. Now this is impossible. Therefore, if categorical obligability does entail this, no action can have been categorically obligable. And, if no action has been categorically obligable, this argument for indeterminism breaks down at the first move. Against this the following objection might be made. We certainly have the notion of categorical obligability, whether it applies to any action or not. If no action that has ever been done, or has ever been contemplated and left undone, has been categorically obligable, how did we get the notion of categorical obligability? To this objection we might I think, make the following answer. We might deny that we have a non-descriptive idea of categorical obligability. We might deal with the alleged notion of categorical obligability in a somewhat similar way to that in which Hume tried to deal with the alleged notions of perfect straightness, perfect flatness, and so on. We certainly have the positive notion of *conditional* obligability, correlated with the notion of *conditional* substitutability. There is no difficulty in accounting for the origin of this notion, for there is no reason to doubt that some actions are conditionally substitutable and therefore conditionally obligable. We also have the notion of orders of substitutability, and, correlated with this, the notion of orders of conditional obligability. This may be compared with the notion of degrees of straightness. To such a series of orders of conditional obligability there is no intrinsic highest term. If we can think of an action having conditional obligability of the n-th order, we can equally think of an action as having conditional obligability of the (n + 1)-th order. Now we may think of this series as having an upper limit, which is not a member of it. And

so our idea of categorical obligability may be a merely descriptive idea of the following kind. We may think of categorical obligability simply as the upper limit of the series of ascending orders of conditional obligability. If in fact the series has no upper limit, there will be no term answering to our description of categorical obligability. Nevertheless, in one quite common sense of "idea", we shall have an idea of categorical obligability; just as a person can have an idea of the ratio whose square is equal to the ratio of 2 to 1, though there cannot be such a ratio. We follow the series in thought for a certain distance; see that it could be followed further; get tired of thinking; and postulate an upper limit. And the idea of categorical obligability is the product of thinking so far and refusing to think further. Taking these facts and possibilities into consideration I am not prepared to accept the moral argument for indeterminism as conclusive.

2.2. Non-ethical arguments

2.21. Empirical

2.211. Argument from immediate conviction

Some people tell us that, at the moment when they make a voluntary decision in favour of alternative *A*, they are convinced that they could have decided instead in favour of alternative *B*. This has been used as an argument for indeterminism. There are several remarks to be made about this. (i) Many people profess to find it self-evident that every event must be completely determined by occurrent causation. It is plain that the co-existence of these two convictions diminishes the importance of both of them. The situation is rendered still more unsatisfactory by the fact that they may co-exist in the same person. The very same man, when he reflects on the notions of "event" and "causation", may find it self-evident that *every* event must be completely determined by occurrent causation; and, when he is actually making a decision, may find it obvious that *this* event is not completely determined by occurrent causation. He must be mistaken in one of these convictions, and each of them may be equally strong when taken in isolation from the other. (ii) It seems to me very likely that the alleged certainty, at the moment of decision, that one could have decided otherwise, may be the product of confusion of ideas and incomplete knowledge of fact. In the first place, the agent may be confusing conditional substitutability of a high order with categorical substitutability. Secondly, he may be confusing the fact that determination of his volition by motives is a perfectly unique kind of occurrent causation, with the fancy that his volitions are not completely determined by occurrent causation. Lastly, even if he makes neither of these mistakes, his conviction is of very little importance, for the following reason. If his decision were

completely determined by occurrent causation, it is most unlikely that all the occurrent cause-factors in the total cause of his decision would be open to introspection. Hence the fact that the sum-total of the occurrent cause-factors which he can introspect are often plainly inadequate to account causally for his decision is no evidence whatever that his decision is not completely determined by occurrent causation. Ever since Leibniz pointed out this perfectly elementary fact, the present argument for indeterminism has ceased to be respectable. But, like the elderly ladies mentioned by Pope, it still "haunts the places where its honour died".

2.212. Other empirical arguments for determinism or indeterminism

2.2121. Indirect arguments. We come now to empirical arguments which are specially directed to make it highly probable or highly improbable that every human voluntary action and all its causal ancestors are completely determined by occurrent causation. These may be divided into two classes, viz. analogical arguments from non-mental events, and direct arguments. Arguments of the first class may be stated and dismissed without much ceremony. (i) It used to be asserted that there is overwhelming empirical evidence that all physical and physiological events are completely determined by occurrent causation. It was then argued that it is most unlikely, in view of this fact, that human voluntary action should be incompletely determined or should have causal ancestors which contain cause-factors not completely determined by occurrent causation. To this three answers can be made. (a) Complete determinism is a proposition of a kind which could not possibly be established by empirical arguments. All observable and measurable characteristics of things and events can be observed and measured only within certain limits of accuracy. It always remains possible that an event may be undetermined in respect of any observable characteristic within limits which are narrower than the limits of accurate observation and measurements. (b) It has now become doubtful whether all physical events are completely determined even within the limits of accurate measurement. (c) Even if the premiss were certain, the argument is deplorably weak. Volitions are utterly unlike physical or physiological events, and the causation of volition by motives is utterly unlike the causation of physical or physiological events by other events of the same kind. Hence any argument by analogy from the premiss that the latter are completely determined by occurrent causation to the conclusion that the former are also thus determined is of the weakest kind. (ii) If anyone should feel inclined nowadays to use a similar argument from the principle of indeterminacy in atomic physics to the *incomplete* determination of human voluntary actions or their causal ancestors, he would be open to very similar objections. It is still uncertain whether the indeterminacy

which atomic physicists now recognise is ontological or only epistemological. And, even if it were indubitably ontological, any argument by analogy from the incomplete determination of certain atomic events to the incomplete determinism of human voluntary action or their causal ancestors would be extremely weak.

I suspect that both determinists and indeterminists who use these arguments tacitly assume a suppressed premiss about the determination of mental events by events in the brain and nervous system. If this be assumed, the arguments cease to be arguments by analogy, and they take the following form. The determinist argument may be put as follows. If all physical events are completely determined by occurrent causation, and all mental events are completely determined by certain physical events in the brain and nervous system, every mental event will be completely determined, and all its causal ancestors will be completely determined, by occurrent causation. The weak point in the argument is that no reason is given for the premiss that every mental event is completely determined by events in the brain and nervous system. This premiss seems to be no more certain than the conclusion which it is used to support, viz. that every mental event and all its causal ancestors are completely determined by occurrent causation.

We can now consider the modified form of the indeterminist argument. It is doubtful whether the indeterminist could consistently use the premiss that all mental events are completely determined by events in the brain and nervous system. If there is indeterminism *somewhere*, it is difficult to see how one could be sure that there is complete determinism just here. Fortunately this premiss is not needed. The fairest way to state the amended indeterminist argument would be as follows. Either human volitions are completely determined by occurrent causation, or they are not. If they are not, indeterminism is admitted at the first move. If they are, then the total cause of a volition will certainly contain events in the brain and nervous system as cause-factors, even if there be purely mental cause-factors too. Now some physical events are not completely determined by occurrent causation, and those events in the brain and nervous system which are cause-factors in the total cause of a volition may be physical events of this kind. If so, this volition will have a causal ancestor which contains a cause-factor not completely determined by occurrent causation. This is a much better argument than the amended determinist argument; for the premiss here is an extremely plausible hypothetical proposition, instead of an extremely sweeping categorical proposition, about the causation of mental events.

I think, however, that reflexion on this argument and its conclusion reinforces my contention that indeterminism is not a sufficient condition of obligability. What possible ethical significance could there be in the fact that some of the occurrent cause-factors in the total cause of a volition are the in-

completely determined jumps of electrons from one orbit to another?

2.2122. Direct arguments. Finally we come to direct empirical arguments for and against determinism as applied to human voluntary actions. The empirical facts are these. We constantly do make predictions with a considerable degree of confidence about the voluntary actions of ourselves and our friends. We do so with still more confidence about certain voluntary actions of large classes of men. As Russell says "Bradshaw's time-table consists entirely of predictions about the voluntary actions of engine-drivers". Now such predictions have been repeatedly verified; and, if they had not been, human society could never have existed or continued. On the other hand, persons of the most settled habits, whom we think we know through and through, do form time to time behave in the most unexpected ways, as, e.g., when a confirmed elderly bachelor marries, or a "tough" bookmaker undergoes religious conversion. Persons who insist on the former set of facts maintain that it is most unlikely that our predictions would have been verified to anything like the extent to which they have been if indeterminism were true. Those who insist on the latter set of facts maintain that these are inconsistent with determinism as applied to human voluntary actions.

It appears to me that there is nothing in either argument. All the facts about predictability and unpredictability of human actions are compatible with either theory. I will now take the two theories of determinism and indeterminism in turn, and try to show this.

(i) Let us suppose that every voluntary action and all its causal ancestors are completely determined. In order to predict a completely determined action it is necessary to know beforehand the relevant laws, the relevant dispositional properties of the agents, the relevant processes which were going on in the agents, and the relevant external relations of the agents. Now let us compare human beings and their actions, on the one hand, with physical things and their actions, on the other, in this respect. (a) The laws which govern the motions of bodies, i.e. the laws of mechanics, are the same for all bodies and for all motions. They are quite independent of the particular materials of which a body is made and are quite independent of the particular way in which the motion is produced. They can therefore be discovered once and for all by suitable experiments on certain motions of certain selected bodies and then applied to all motions of all bodies. Now the laws of psychology are not in this position. No doubt there are some psychological laws which apply to all mind, e.g. the laws of retentiveness, association, etc. And there are probably more specific psychological laws which apply to all human minds. But it is quite possible that each human mind may be subject to a still more specific psychological law which is peculiar to itself and not deducible from the psychological laws which apply to all human minds. If so, the

psychological laws which are characteristic of a given individual could be discovered only by a special study of the behaviour of that individual. (b) The dispositional properties of physical objects are practically the same for all samples of a given kind of material prepared in the same way. If we determine the mass and the coefficient of elasticity of a single billiard-ball, we can be practically certain that the mass and the elasticity of any other billiard-ball of the same brand made by the same firm will be the same. There is nothing analogous to this in the case of the psychological coefficients of a human individual. Human minds cannot be regarded as so many different but similar samples made from a common material. Each starts with its own characteristic dispositional properties, and the dispositional properties of any human mind can be discovered, if at all, only by studying *that* mind. (c) The dispositional properties of many physical objects are practically constant for long periods and are scarcely affected at all by most of the transactions in which these objects take part. E.g. the mass and elasticity of a billiard-ball will remain unaltered for centuries unless it is treated with extreme violence. The dispositional properties of human minds are continually altering. If billiards were played with balls made of plasticine, the motion of the balls would still be completely determined, but it would be practically impossible to predict them. But the dispositions of human minds change much more than this. For each mind is continually having experiences, each experience leaves its trace, and each trace modifies and is modified by the pre-existing system of dispositions. (d) We do not expect a physical object suddenly to manifest a disposition which has previously been latent throughout the whole of its history. But it is logically possible for this to happen, and, in the case of human minds, this possibility is often realised. (e) We can now pass from dispositional to occurrent cause-factors. No one has any direct acquaintance with any of the internal processes of any mind but his own. He has to infer what has been going on in another mind from conversation, gesture, facial expression, overt action, and so on. This is obviously a very precarious process as compared with direct observation of what is going on in physical objects. (f) Each of us has only a very imperfect knowledge of what is going on in his own mind. It seems certain that the experiences which I can remember or introspect are a very small part of the total processes going on in my mind. And the facts of abnormal psychology and the work of the psycho-analysts seem to show that some of the most important occurrent cause-factors are not open to introspection. From all these facts two results emerge. (α) That repeated failure to predict the actions of a human being would be quite likely to happen even if all his actions and all their causal ancestors were completely determined. Consequently the degree of failure in prediction which we actually find casts hardly any doubt on determinism. (β) On the other hand, determinism, as applied to human actions and experiences, must always remain a mere

scientific postulate. There is not the least chance of proving it experimentally.

(ii) Let us now suppose that no human action is completely determined in respect of any dimension of any characteristic. It might still be the case that the range of indetermination is narrow for every dimension of every characteristic. And it might become narrower and narrower as certain habits were established, certain temptations repeatedly fallen to or repeatedly conquered, and so on. One way of looking at the matter is the following. It might be that at first any effort that fell within the division *XB* of the total range of indetermination *AB* would suffice to ensure that the temptation to take a certain drug would be resisted $\overline{A\ X\ B}$. Let us call *XB* the range of *effective* effort. If I repeatedly fail to make such an effort, and thus establish a habit it may be that only an effort which falls within the narrower range *YB* $\overline{A\ X\ Y\ B}$ would be adequate to ensure refusal of the drug. Let us suppose that, all through, any *one* degree of effort within the total range of indetermination *AB* is as likely to be made as any *one* other. Still there is more chance of such a degree falling within the wider range *XB* than in the narrower range *YB*, since the former covers more possibilities than the latter. Thus, as time goes on, the range of *effective* effort continually shrinks, and occupies a smaller and smaller fraction of the total range of *causally possible* effort. And so it becomes less and less likely that the effort which happens to be made will fall within the range of effort which will be effective in preventing me from taking the drug. Thus a reasonable form of indeterminism is quite compatible with the fact that we can make highly probable conjectures about the conduct of individuals in many cases.

Probable predictions about collective action, such as those of Russell's engine-drivers, can be reconciled still more easily with a reasonable form of indeterminism. In the first place, there are generally several drivers available; and, if one decided not to drive his train at the advertised time, another would almost certainly be ordered to do so. Bradshaw predicts, e.g., only that the Flying Scotsman will leave King's Cross at 10 a.m. It does not predict that it will be driven by so-and-so. Secondly, it is unlikely that the incompletely determined volitions of a number of engine-drivers would all happen at the same time to take the form of refusing to drive a certain train. Now, it is only if this unlikely coincidence were fulfilled that the train would fail to start at the advertised time. Lastly, Bradshaw is not infallible. His predictions about the volitions of engine-drivers break down hopelessly when there is a railway-strike.

The upshot of the discussion is this. If a limited form of indeterminism were true, one could make probable predictions about human conduct; and, if complete determinism were true, one could make no more than probable predictions about it. Thus the actual facts about the partial predictability of human conduct are quite compatible with either theory.

3. Consequences of determinism

Let us now take determinism simply as an hypothesis, and see what consequences, ethical and otherwise, would follow from it. The fundamental ethical consequence which would follow is that the notion of categorical obligability would have to be rejected as delusive. Any other ethical notion which involved this would therefore have to be rejected or modified, but ethical notions which are independent of it might remain unchanged. I will now consider some of the logical consequences of assuming determinism.

(i) It is sometimes said that, if determinism were true, the future would be completely fixed already. If I am faced with alternatives *A* and *B*, it is already determined that I shall choose *A* or it is already determined that I shall choose *B*, though I do not at present know which of the two choices is determined to happen. In fact, if determinism is true, there are no *real* alternatives. When *A* and *B* are said to be both possible, this is relative to my partial ignorance of my own dispositions, of the laws of human psychology, of my own non-introspectable mental processes, etc. Objectively either *A* alone is causally possible and *B* is causally impossible, or *B* alone is causally possible and *A* is causally impossible. It is sometimes concluded from this that a consistent and clear-headed determinist would have no motive for deliberating and no motive for trying to resist temptation or to improve his own character.

Now the first part of this contention is certainly correct. If determinism is true, the future is already fixed and nothing can happen in the future except what is already determined to happen. But there are two points to notice. (i) The future is not determined *independently* of the present and the past. If *A* is determined to happen and *B* not to happen, this is because it is determined that I shall choose *A* and reject *B* and because my choice of *A* is a necessary cause-factor in the total cause of *A*. The fact that my choice is completely determined by previous events does not prove or suggest that it is not an essential factor in determining subsequent events. People are liable to think that, if *X* is the total cause of *Y* and *Y* is the total cause of *Z*, *Y* is somehow superfluous and *X* is the total cause of *Z*. This is obviously a mistake. *X* is indeed a *causal ancestor* of *Z*, but it is not the *cause* of *Z*; and it is a causal ancestor of *Z* only because *Y* intervenes as the total effect of *X* and the total cause of *Z*. The theory that deliberation and decision are causally ineffective, and that the same results would always have followed whether they had taken place or not, and whether they had gone in one direction or the other, may be called *fatalism*. It is in no way entailed by determinism, and it is quite consistent with indeterminism and with libertarianism. (ii) Unless this confusion between determinism and fatalism is made, a belief in determinism does not remove the motives for deliberating or for trying to resist temptation. Suppose I am faced with alternatives *A* and *B*, and that in fact *B* is determined

to happen. I never know that it is *B* and not *A* which is determined to happen until it has happened. But I do know that *A* is certain to happen if I choose it and certain not to happen if I reject it. And I do know in many cases that I am more likely to choose *A* and persist in my choice if I deliberate on the merits and defects of *A* and *B* and put forth an effort in favour of *A* than if I do not deliberate or put forth an effort at all. This knowledge, together with the desire to act prudently, constitutes a motive for deliberating and putting forth effort. And this motive is quite independent of my belief that my choice is completely determined, and is quite independent of the fact (which I do not and cannot know) that it is determined that I shall eventually do *B* and not *A*. (iii) It is sometimes said that, if determinism were true, people would not be morally responsible for their characters and actions. You could still, of course, talk of a bad man or a good man as you can talk of a bad car or a good car; and you could still say that a certain man acted well or badly on a certain occasion, as you could say that a certain car ran well or ran badly on a certain occasion. The meaning would be roughly as follows. To say that a man is bad would mean that his dispositions are such that he inevitably succumbs to temptations under conditions in which most men would inevitably resist them. Now it is said that we mean something more than this when we call a man good or bad, and that this something more is incompatible with complete determinism. Moreover, we do not consider a bad car responsible for its badness or a good car for its occasional bad running; we ascribe responsibility to its makers or to the people who have misused it. It is said that this is because the actions of cars are completely determined; and that, if the actions of men were completely determined, we could not hold men responsible for them.

The first point to notice is that, even if determinism be true, there are absolutely fundamental dissimilarities between human minds and all machines. It is therefore possible that the differences in the sense of "good" and "bad", "responsible" and "irresponsible", as applied to men and to machines, depend on these differences and not on the question of determinism and indeterminism. In the first place, there is the distinction between intentional and unintentional behaviour in the case of men. In machines all behaviour is like the purely reflex actions of human organisms, and we do not regard human beings as responsible for the latter. *A fortiori* there is nothing in machines analogous to determination by motives, ideals, etc., in men. Secondly, whether determinism be true or not, men differ from machines in the fact that they can and do deliberately modify their own characters. We draw a distinction between two kinds of good men, viz. those who were born with a happy balance of dispositions and placed in fortunate surroundings, and those who with great difficulty and against obstacles succeeded in building up and maintaining a good character. We are inclined to say that

both are "good", but that men of the second kind have a special "merit" which does not belong to men of the first kind. They are more responsible for their own goodness than men of the first kind. There is nothing in the least analogous to this in the case of machines.

It seems to me that the only actions for which an agent can be held directly responsible are his intentional actions. He is directly responsible for his character only in so far as he has modified it by actions which he believed would be likely to modify it and which he performed because of or in spite of this belief. He may be indirectly responsible for his character in so far as he has modified it by intentional actions which he did not believe to be likely to modify it, either because he did not consider the question of their effect on his character at all or because he mistakenly believed that they would have no such effects. This explains why machines and animals are *not* regarded as responsible for their actions or characters. It is because their actions are not intentional, and are certainly not intended to modify their own natures; it is not because their actions are completely determined. But does it suffice to explain why men *are* regarded as responsible for some of their actions and for some aspects of their characters? Many people would say that it does not suffice. They would say that, if men's actions are completely determined by occurrent causation, they are not responsible even for their intentional actions or even for that part of their characters which they have intentionally modified.

Now I believe that these people have something true and important in their minds, but I do not think that they have expressed it clearly. If you were to press them as to why they hold that a man would not be responsible even for his intentional actions if all events are determined completely by occurrent causation, I think they would eventually answer as follows. If determinism is true all my intentional acts have a causal ancestor which consists of the following factors, viz. (a) the innate character and dispositions with which I started to exist at the moment of conception, and (b) the external situation, with its agents and processes in which I started to exist. All that has happened in me and to me since then has been the inevitable outcome of this causal ancestor. Now I certainly am not in any sense responsible for being conceived and coming into existence with my initial character and dispositions. And I certainly am not in any sense responsible for the initial external situation, with its agents and processes, in which I started to exist. Now I cannot possibly be responsible for anything which is the inevitable consequence of a causal ancestor in which every factor was existentially and qualitatively independent of me. Therefore I cannot be responsible even for my intentional actions or for those developments of my character which are due to my intentional actions. It will be noted that this argument, if valid at all, would hold equally whether we regard the production of a human soul as due

entirely to natural causes or due to the miraculous action of God. The only difference is that, on the former supposition, no one would be responsible for my actions, whilst, on the latter, God would be responsible for them.

Now it seems to me that the reasoning here is valid; and that all the premisses, with just one exception, are obviously true. The one premiss which can be questioned is that human minds come into existence at the moment of conception. This is not entailed by determinism. It would be quite consistent with determinism that every human mind should have had no beginning and should have existed throughout all past time. This premiss is therefore independent of determinism, and could consistently be dropped whilst determinism was retained. And, if it were dropped, responsibility and determinism could be held together. For the character and dispositions with which I begin any one of my incarnations would be developed out of my previous character and dispositions by my own doings and sufferings. And, if I never began to exist, every one of my incarnations would be preceded by another incarnation or by a phase of discarnate existence. So there would be no stage at which my character and dispositions would have a causal ancestor in which every factor was existentially and qualitatively independent of me. Thus we reach the following interesting conclusion. Moral responsibility can be reconciled with determinism if and only if we assume that human minds never begin to exist, but that each human mind has existed through all past time and has developed through its own doings and sufferings.

I will now make some comments on this argument. (i) It does not prove that human minds have existed through all past time unless we admit both that determinism is true and that moral responsibility is a notion which does apply to some of our actions. Now determinism does not seem to me to be certain, and moral responsibility may be a delusive notion which really applies to nothing. Indeed some people might be inclined to reverse the argument. They might say: Since the beginningless pre-existence of human minds is obviously false, it follows that either determinism is false or moral responsibility is a delusive notion. (ii) I should not be prepared to accept this reversed argument; for it does not seem to me that the beginningless pre-existence of human minds is obviously false. I do not think that Western philosophers have ever considered seriously enough the extreme difficulties involved in the notion of the coming into existence of a mind. A mind seems to be a substance or continuant. Now we know what we mean by the coming into existence of a new complex continuant, e.g. a watch, or a car, or a drop of water. We mean that certain pre-existing continuants, e.g. atoms of oxygen and atoms of hydrogen, which previously stood in other and less intimate relations, began at a certain moment to stand in certain more intimate relations, and continued for some time to do so. The complex continuant, thus formed, had certain qualities of its own, which did not belong to the pre-

existing continuants in their separated state. But it certainly does not look as if a mind were a complex continuant, formed by pre-existing continuants coming into and remaining in certain specially intimate relations. And to talk of the coming to be of a non-complex continuant is to use words which convey no clear meaning. Presumably then there must be *some* continuants which never began to exist; and, if there are any, it seems not unreasonable to suppose that human minds may be such continuants. At any rate the supposition avoids the necessity of assuming a perfectly unintelligible kind of event at each conception of a human being. Of course this particular difficulty would also be avoided by the materialistic assumption that minds are not continuants and that mental processes are in some way by-products of processes in the brain and nervous system. But this assumption has its own difficulties. And it is, I think, clearly incompatible with the validity of the notion of moral responsibility.

It remains to consider whether moral responsibility is compatible with indeterminism. It seems clear to me that, just in so far as an action is determined by a total cause which contains incompletely determined events as factors it is an irresponsible action. So indeterminism by itself can do nothing to help the notion of moral responsibility. No doubt many people who accept indeterminism also accept libertarianism. They hold that what is left undetermined by occurrent causation is determined by the self as a substance or continuant exercising non-occurrent causation. I have already said that this doctrine seems to me to be nonsensical. So, if moral responsibility involved libertarianism, it would have to be rejected as a delusive notion.

I will now sum up the conclusions of this discussion. In order that a person may be morally responsible for an action the following conditions must be fulfilled. (a) The action must be intentional. (b) He is responsible for it only in so far as it is determined by his character and dispositions. (c) He is responsible for it only in so far as his character and dispositions at the time when he did it are the products of his previous character, dispositions, experiences, and actions. In order to reconcile conditions (b) and (c) it seems necessary to assume that the person has existed and has been having experiences and doing actions through all past time, though there may have been periods during which he was completely quiescent and unconscious. Thus moral responsibility seems to entail that human minds have persisted throughout all past time.

Guide to authors/subjects*

* Only the first page upon which a concept is discussed, is cited in this guide.

NIJHOFF INTERNATIONAL PHILOSOPHY SERIES

1. Rotenstreich N: Philosophy, History and Politics – Studies in Contemporary English Philosophy of History. 1976. ISBN 90-247-1743-4.
2. Srzednicki JTJ: Elements of Social and Political Philosophy. 1976. ISBN 90-247-1744-2.
3. Tatarkiewicz W: Analysis of Happiness. 1976. ISBN 90-247-1807-4.
4. Twardowski K: On the Content and Object of Presentations – A Psychological Investigation. Translated and with an Introduction by R Grossman. 1977. ISBN 90-247-1726-7.
5. Tatarkiewicz W: A History of Six Ideas – An Essay in Aesthetics. 1980. ISBN 90-247-2233-0.
6. Noonan HW: Objects and Identity – An Examination of the Relative Identity Thesis and Its Consequences. 1980. ISBN 90-247-2292-6.
7. Crocker L: Positive Liberty – An Essay in Normative Political Philosophy. 1980. ISBN 90-247-2291-8.
8. Brentano F: The Theory of Categories. Translated by RM Chisholm and N Guterman. 1981. ISBN 90-247-2302-7.
9. Marciszewski W (ed): Dictionary of Logic as Applied in the Study of Language – Concepts / Methods / Theories. 1981. ISBN 90-247-2123-7.
10. Ruzsa I: Modal Logic with Descriptions. 1981. ISBN 90-247-2473-2.
11. Hoffman P: The Anatomy of Idealism – Passivity and Activity in Kant, Hegel and Marx. 1982. ISBN 90-247-2708-1.
12. Gram MS: Direct Realism – A Study of Perception. 1983. ISBN 90-247-2870-3.
13. Srzednicki JTJ and Rickey VF (eds): Leśniewski's Systems – Ontology and Mereology. ISBN 90-247-2879-7.
14. Smith JW: Reductionism and Cultural Being – A Philosophical Critique of Sociobiological Reductionism and Physicalist Scientific Unificationism. 1984. ISBN 90-247-2884-3.
15. Zumbach C: The Transcendent Science – Kant's Conception of Biological Methodology. 1984. ISBN 90-247-2904-1.
16. Notturno MA: Objectivity, Rationality and the Third Realm: Justification and the Grounds of Psychologism – A Study of Frege and Popper. 1984. ISBN 90-247-2956-4.
17. Dilman I: Philosophy and Life. 1984. ISBN 90-247-2996-3.
18. Russell JJ: Analysis and Dialectic – Studies in the Logic of Foundation Problems. 1984. ISBN 90-247-2990-4.
19. Currie G and Musgrave A (eds): Popper and the Human Sciences. 1985. ISBN 90-247-2998-X.
20. Broad CD: Ethics. Edited by C Lewy. 1985. ISBN 90-247-3088-0.
21. Seargent DAJ: Plurality and Continuity – An Essay in GF Stout's Theory of Universals. 1985. ISBN 90-247-3185-2.
22. Atwell JE: Ends and Principles in Kant's Moral Thought. 1985. ISBN 90-247-3167-4.