IN THE SHADOW OF DESCARTES

SYNTHESE LIBRARY

STUDIES IN EPISTEMOLOGY,

LOGIC, METHODOLOGY, AND PHILOSOPHY OF SCIENCE

VOLUME 272

GEORG HENRIK VON WRIGHT
Academy of Finland, Helsinki,
and University of Helsinki, Finland

IN THE SHADOW OF DESCARTES

Essays in the Philosophy of Mind

Springer Science+Business Media, B.V.

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-90-481-5011-3 ISBN 978-94-015-9034-1 (eBook)
DOI 10.1007/978-94-015-9034-1

Printed on acid-free paper

Originally published by Kluwer Academic Publishers in 1998
Softcover reprint of the hardcover 1st edition 1998

"In philosophy it is always good to put a *question* instead of an answer to a question.
For an answer to the philosophical question may easily be unfair; disposing of it by means of another question is not."

Wittgenstein

To *Lilli* and *Fred*

Friends and companions
in the Cartesian shadow

CONTENTS

PRELIMINARY

My awakening to philosophy took place when I was an adolescent. The first book I read was *Psykologi* by the much esteemed Swedish philosopher and essayist Hans Larsson. It had a section on the mind-body problem which put my thoughts in motion. Soon after I read Wilhelm Jerusalem's *Einleitung in die Philosophie* and was especially fascinated by the account it gave of the empirio-criticist form of identity theory advocated by Mach and Avenarius. I thought out for myself a "monistic philosophy" inspired by the sources mentioned. I cannot remember my "arguments" – only that they seemed to me, at the time, "absolutely convincing".

When in 1934 I started university studies in philosophy under the guidance of Eino Kaila in Helsinki, psychology was still considered to be part of "theoretical philosophy". This meant that I also got a basic education in psychology, including a rudimentary acquaintance with experimental work. I think this was a good preparation for research into the philosophy of mind or of psychology.

Kaila was himself an eminent representative of the two disciplines which were combined with his university chair, and he had a good sense of the philosophical relevance of perceptual psychology and also of the neuroscience of his day. His contributions to the philosophy of psychology seem to me superior to much of the "sense-datum philosophy" which had flourished in England since the turn of the century. As a philosopher Kaila professed a monism (identity theory, parallel theory) which he again and again up to his death in 1958 tried to articulate in writing – without, however, ever being able to give to it a form which would have fully satisfied him.[1]

My own itinerary in philosophy initially took a different direction. Under the influence of Kaila I became interested in logic and the logic-inspired philosophy of the Vienna Circle. My first work was on induction and probability. It was succeeded by work in modal logic. The discovery and study of the modalities now known as deontic contributed to a gradual shift of my interest from the philosophy of logic, first to the philosophy of norms and values, and then to the philosophy of human action. My thinking centred round concepts like cause and reason (of an action), intentionality, explanation of action, and freedom and determinism. By this route I eventually came to the philosophy of mind and my early fascination with the mind-body problem and psycho-physical parallelism was reawakened. The way this happened is reflected in the fact that my approach to the mind-body problem has been, so to speak, from the "output"

G. H. von Wright, In the Shadow of Descartes. Essays in the Philosophy of Mind, ix–xii.

aspect of an agent initiating changes in the physical order of things and not from the "input" aspect of a subject receiving impressions from the outer world through his senses. My first dip into these waters was in the Tanner Lectures I gave in Helsinki in 1984, published under the title "Of Human Freedom". At a symposium in Åbo two years later, Norman Malcolm commented on them in a paper "Mind and Action". My reply to him was called "Reflections on Psycho-Physical Parallelism".[2] It was after these events that I embarked on what I considered a new opening on my philosophic journey.

At first I thought I could link up with the ongoing discussion in the area. Of course I was not unaware of what had been going on "in my absence". Beside Kaila and *Gestalt*-psychology, the sense-datum-philosophy of Russell, Broad, and Moore had been part of my early education. An uninterrupted, at times very intense, occupation with the thought of the "later" Wittgenstein for nearly half a century can be said to have continued and supplemented this education. If in what I have written in later years there are echoes of outside sources they stem mainly from Wittgenstein. I was not ignorant of the new versions of materialism and identity theory which became topics of lively discussion in the late 1950s and still continue strong, nor of the revived Cartesian dualism and the debates to which it has given rise, nor finally of the impact on philosophy made by recent brain research and artificial intelligence study. It was into these post-Wittgensteinian developments that I was hoping to integrate my own thinking. The first thing to do was to read and learn. This I did. I learnt something and my horizons broadened. But when I started writing and had to take issue with what I had read I had a strong and sometimes even frustrating impression that I had to go my own lonely way, and if I could fall back on something earlier in my own philosophical experience, it was nearly always to problems and viewpoints with which I had become familiar in early years when I was still Kaila's student. So I abandoned plans of contributing to and taking part in an ongoing debate and decided to write, to begin with, only for myself in order to clear my thoughts on questions which agitated my mind.

In the years from 1986 on I wrote extensively but did not publish anything on some traditional topics in the philosophy of psychology centring round the notions of perception and sensation and of quality and thing. Copies were circulated to a small number of friends who could be expected to read the material with sympathy for the writer's efforts. For the comments which I succeeded in eliciting I am most grateful.

Years later I returned to these writings, made changes and corrections and purged them of long passages which appeared to me either erroneous or unconvincing. What stood the test is published here. The material is divided into sections corresponding to the order in which they were composed.

Partly overlapping in time with those writings on problems in the philosophy of psychology were successive efforts to deal with the classic mind-body

problem. In the course of years the results of some of these efforts were published – from the Tanner Lectures of 1984 to a paper in the *Journal of Theoretical Biology* ten years later. They are reprinted here with hitherto unpublished material dealing with the same problem or aspects of it.

The writings collected in this volume do not form a unified whole. The same ground is returned to time and time again. Sometimes the successive efforts signalize progress towards greater clarity. More often, perhaps, they reflect a slight change of angle from which the problem is approached. Some things may strike the reader as inconclusive or even as mildly contradictory. It was not always possible for me to make up my mind definitely on alternative positions. I have not wanted to conceal or smooth out the agonies which thinking about the fundamental questions of philosophy always caused me.

July 1997

GEORG HENRIK VON WRIGHT
Academy of Finland

ACKNOWLEDGEMENTS

I owe Professor Tom Reuter special thanks, not only for stimulating discussions of the philosophy of the body-mind relation, but also and not least for letting me profit from his expert knowledge of animal and human neurophysiology. He drew my attention to relevant literature in the area and helped me avoid mistakes which otherwise I would have committed when speculating about matters in a field of science of which I have no experimentalist experience whatsoever and only a rudiment of bookish knowledge.

I wish to thank Dr Mark Shackleton for checking and improving the language of those parts of my English manuscript which had not been published before.

The published essays are:

"Of Human Freedom", *The Tanner Lectures on Human Values* Vol. VI. Edited by Sterling M. McMurrin. Salt Lake City: University of Utah Press, 1985.

"An Essay on Door-Knocking," *Rechtstheorie* **19** (1988).

"On Mind and Matter," *Journal of Theoretical Biology* **171** (1994).

The three papers are here reproduced with minor changes by the kind permission of the sources where they originally appeared.

Finally, I thank Risto Vilkko and Auli Kaipainen for invaluable assistance with the technicalities of producing the manuscript for the press.

NOTES

[1] Cf. my paper "Eino Kaila's Monism" in Ilkka Niiniluoto, Matti Sintonen, and G.H. von Wright (eds.), *Eino Kaila and Logical Empiricism*. Helsinki, Societas Philosophica Fennica, 1992.
[2] Malcolm's comments and my reply are printed in Lars Hertzberg and Juhani Pietarinen (eds.), *Perspectives on Human Conduct*. Leiden, E.J. Brill, 1988.

OF HUMAN FREEDOM

"Hundert irreleitende Bilder kommen hier zusammen, und das macht die Schwierigkeit der philosophischen Situationen aus. Wohin wir treten, wankt wieder der Boden. Die 'grossen', schwierigen Probleme der Philosophie sind es nicht etwa dadurch, dass hier ein unerhört subtiler und geheimnisvoller Sachverhalt ist, den wir erforschen sollen, sondern dadurch, dass an dieser Stelle eine Grosse Zahl irreführender Ausdrucksformen sich kreuzen." Wittgenstein[1]

FIRST LECTURE

1.

It is often said that the problems of philosophy are perennial. They have been discussed throughout the ages, but never solved. This is sometimes interpreted as a sign that in philosophy there is no progress or even that the pursuit of philosophers is fruitless, all in vain.

It is not *quite* true that philosophical problems are perennial. At least their place in the discussion – whether central or peripheral – is shifting. Such shifts often reflect profound changes in the intellectual culture of an era. An example is the problem of the existence of the material or outer world. Another is the problem of "the freedom of the will." The first can hardly be said even to have existed in ancient and medieval European thought. Greek philosophy was not much absorbed in discussion of the second. Both problems got their characteristic modern twist under the impression of the mechanistic world-view which emerged from the revolutions in astronomy and physics in the late Renaissance and Baroque periods. They can be said to have crystallized in the philosophical system of Descartes.

It *is* true, I think, that philosophical problems are not "solved." It sounds absurd to say that G. E. Moore (eventually) "proved" that there exists a world external to my mind – even if one cannot find any fault in Moore's argument. At most Moore succeeded in cutting the discussion short for a time, but one can be sure that it will be revived. One can *not* be sure, however, that it will always be thought important. It may even come to be considered no "problem" at all (any longer).

An important aspect of change in philosophy concerns the way its problems are formulated. The problem of freedom is a good example. For a long time it was customary to think that human actions as overt manifestations of behaviour are caused by something called volitions or acts of the will. Human freedom,

G. H. von Wright, In the Shadow of Descartes. Essays in the Philosophy of Mind, 1–44.

First published in *The Tanner Lectures on Human Values*, Volume VI. Edited by Sterling M. McMurrin. Salt Lake City: University of Utah Press, 1985, pp. 107–170. The two lectures under the title "Of Human Freedom" were given in Helsinki on 16 and 17 May 1984.

it was then often said, just consists in this: that an agent's actions are determined by his will and not by external forces over which he has not control or power. This was a way of reconciling freedom with determinism. It was thought important as long as science nourished and sanctioned a deterministic world-view. But a difficulty was lurking in the background.

Granted that action is free when in conformity with our will, what then of the will itself? Are we free to will what we will? Or is the will determined by something else? If the will is not free, action determined by the will can be free at most in some relative sense, it seems.

Questions such as these constitute what I propose to call the "classical" problem of the Freedom of the Will. I think it is right to say that this particular problem is now gradually receding into obsolence.

There is no such thing as "mere" willing. Willing has an object, is *of* something. And the same holds for intending, wanting and wishing. Only seldom do we explain an action by saying that we willed or wanted just it. Giving this answer is much like brushing the question of why we did it aside – like saying "it is none of your business to inquire into the motives for my action." The reason why I did something might be that I coveted or wanted something else to which I thought the action conducive. This other thing was then the object of my will. Willing *it* was the *reason* for my action, that which made me do what I did.

The "classical" way of posing the problem of freedom can be said to obscure the factors which are normally said to determine our actions, viz., the *reasons* we have for performing them.

After these remarks I shall say nothing more here about the traditional Freedom-of-the-Will problem.

There is a second way of posing the problem of freedom which also deserves the epithet "classical," chiefly because it too is related to traditional ideas about determinism and science. It is as follows:

Most human actions have what may be termed a *physical* (bodily, somatic) *aspect* consisting in muscular activity or tension and movements of various limbs and, through this, usually also effecting some changes in the physical environment. This bodily aspect of an action is an event, or sequence of events, in nature, *i.e.*, in space and time. Such events presumably have causes in the neural system, in what one calls innervations of the muscles. The innervations may in turn be caused by antecedent somatic changes, perhaps due to stimuli from outside the body. If all natural events are caused by antecedent natural events, going back maybe in an infinite chain to "the dawn of creation," are not then the bodily aspects of our actions predetermined in a way which is irreconcilable with the purported freedom of the agent in relation to what he does? This was the question which worried Kant, in particular. As a child of his times Kant did not doubt the universal validity of the Law of Causation for

the phenomenal world of events in space and time. But man as agent, he thought, is also a citizen in the noumenal world of "things in themselves" and, as such, free and resposible for his actions. However, if the bodily life of man is governed by "iron laws" of causal necessitation, how can it happen that his limbs, on the whole, move in a way which corresponds to the agent's free actions? The question is obscure. The way to answer it is, I think, to try to formulate it clearly – and then see that there is no question at all to be answered. I shall call this the Problem of Congruence, adopting a term suggested by Professor Frederick Stoutland,[2] and I shall address myself to it in the second lecture.

2.

An aspect of what it is to be free is that one is able to, can do, various things. It is therefore natural to approach the problem of human freedom from considerations about ability and its opposite, inability. My starting point will, in fact, be the latter.

Suppose a man is asked whether he can do a certain thing and answers No, he cannot do it. What could be his grounds for this answer? There are several possibilities:

I cannot drive a motorbike – I never *learnt* to do it. I do not *know how* to do it. I cannot solve this or that problem – it is *too difficult* for me; I doubt whether I could ever acquire the needed skill. I cannot buy myself a new car – I have not got the financial *means*. I cannot park here – one is not *allowed* (supposed) to do so. I cannot let you in – I am not entitled to, have *no right* to do so. I cannot eat intestines – I feel so strong an *aversion* to them. I cannot see this play in Helsinki – there is no *opportunity*. I cannot come tomorrow – I have *no time*. I cannot answer the telephone – my broken leg *prevents* me from getting out of bed.

If I cannot do a certain thing because I have not learnt or do not know how to do it, my inability usually pertains to an action of a certain kind or type which I cannot perform. I shall call such action *generic* and contrast it with the *individual* action I perform or omit on a given occasion. When on the other hand I cannot do a certain thing because I am prevented or have not got the means needed for doing it, my inability pertains to the individual performance of an action of a kind I am able to do. In such cases I both can and cannot do the thing in question. I cannot do it *now*, but could have done it, had it not been for this or that, since it is an action of a *kind* I can do. It makes no sense to say that I am prevented from doing something now if it is a question of something which I do not know how to do. Similarly, it is nonsense to say that I could do something if I only knew how to do it. But to say that I would do it

is not nonsense. Generally speaking: *inability* to perform an individual action presupposes ability to perform the corresponding generic action.

Does ability to perform an individual action, too, require ability to perform the action generally? One must be cautious with the answer. Sometimes one succeeds in doing something, *e.g.*, hitting a target, which one would not claim to be able to do in general. One was lucky. Or, the circumstances made the task easy. The case was exceptional. Normally, however, what I can do on the individual occasion is an action *of a kind* which I can do.

It seems, therefore, that of the two "cans" the generic is primary. One could even reserve the term "ability" for it. One could then contrast "the can of ability" with "the can of succesful performance." This is, for some purposes, useful terminology.

What *is* it to be able to perform an action? The way to tackle the question is to ask: When do we *say*, in colloquial language, that a person *can* perform an action of a certain kind or type, for example jump across a certain ditch without wetting his feet in the water? We say this, if normally or on most occasions when he undertakes to do the action he succeeds in performing it. Instead of "undertakes to do" we could say "chooses to do" or "sets himself to do"; occasionally also, depending upon the nature of the action, "tries to do."

But could one not sometimes say truly of a person that he can do an action of a certain kind even though he never did it? Yes – provided the action is sufficiently like another generic action for which his ability is already established. Perhaps our man never jumped this very ditch, or any ditch at all, but was good at athletics. Then, offhand, he may be pronounced able to perform this special trick too.

What about actions which are such that an agent *always* does them? Normally, if I can do an action of a certain type I do it on *some* occasions which afford an opportunity for doing it, and do not do it on others. Some actions, however, may be such that I do them whenever I have an opportunity. Then there usually is a reason why I always do them – for example that doing them gives me enormous pleasure, or that I am under an obligation to do them. Perhaps the action is one for the doing of which there is not often an opportunity – like going to see a play which is performed at long intervals in the place where I live. If, however, for no particular reason I always, whenever there is an opportunity, do something which I have learnt to do, do it quasi "automatically," "mechanically," one may begin to wonder whether this is still "free action". One would perhaps say that doing it has become an obsession with me, or call it an illness (for example kleptomania). Actions which I have learnt how to do but from which I cannot abstain are more like "reflexes" than "actions" of mine. (Generically they remain, of course, types of *action*.) They are *reactions*, one could also say, to the *stimuli* provided by the opportunities for doing them.

The contrary of performing an action is to omit (performing) it. Actions which one is not able to perform one also cannot omit. One is compelled or forced to leave them undone because of one's inability – but this does not mean that one omits them. That is: I shall use the term "omit" here in such a way that ability to omit logically presupposes ability to do.

Can one also be unable to omit an action? Surely. This is but another way of saying that one *must* (is *compelled* to) perform it. (Except when it means that one is also unable to perform it – but this would be an awkward use of "unable to omit.") Different cases may here be distinguished:

I cannot omit an individual action which I am, as we say, physically compelled to do. What is this? Somebody grabs my arm and makes it go through certain motions, perhaps thereby emitting a signal. I try to resist but I cannot; I am too weak. Was my arm going through those movements the performance of an action by me? I think we must answer "No." The action was by the person who moved my arm, not by me. This type of physical compulsion is better termed "violence". One cannot, strictly speaking, be physically compelled to perform an action or physically prevented from omitting it, which means the same. But one can be physically prevented from performing an action – for example by somebody who grabs my arm and keeps it steady when I am about to move it. Then one is physically compelled to omit its performance.

Physical prevention must be understood to mean prevention from performing an individual action which the agent would have performed on the occasion in question had he not been prevented. Perhaps he sets himself to act and recognizes the obstacle only in the course of his attempted performance. Or the obstacle occurs in the course of his attempt. Or it was there before the action was attempted and the agent knew of it and, therefore, omitted the action which otherwise he would have performed. If, however, the agent had *not* attempted the action, regardless of whether or not there was an obstacle to its performance, we do not say that he was prevented, or that his freedom was, on that occasion, restricted.

A genuine case of inability to omit (compulsion to do) is when one acts under the influence, as we say, of an irresistible desire or temptation or under a fearsome threat. "I cannot stand this smell, I must turn away." "I could not refuse handing him my wallet at gunpoint." Someone retorts that I could have let myself be shot, or, speaking of the smell, could have controlled myself. Could I really? To agree that I could not (have omitted the action) seems like saying that what I did was not "really" an action of mine, but more like a "reflex" or behaviour under physical compulsion. But if my behaviour was not just a scream or a jerk or a turning away from something but was a thing which I knew how to do or the significance of which I had learnt, then what I did was surely also an *action* of mine.

A further case of inability to omit is when one has to or must do something in order to attain a set end or ought to do something, because it is one's acknowledged duty. Although one often, without distorting things, says of such actions that one cannot omit them, it is also clear that normally one would not speak of compulsion in connection with them. A set end is something freely chosen, and an acknowledged duty is something one freely assents to. Both exist as the result of an agent's *self-determination*. This also holds good when what is ackowledged as duty conforms to the customs and traditions or is prescribed by the legal order of a society.

3.

If by ability we understand the "generic can" then one can say that the *range of freedom* of an agent is greater or smaller depending upon the number of kinds of actions he can do. This is why education: learning to do things, acquiring the appropriate know-how, is a factor which enhances human freedom. To keep people in ignorance, to deprive them of opportunities of acquiring skills and improving them through training, is thwarting freedom.

Freedom in this sense could also be called *potentional* freedom. To be free (able) to do or omit an individual action can, by contrast, be called *actual* freedom. It follows from what has already been said that actual *non*freedom is a restriction upon an existing potential freedom: the agent *cannot*, on the individual occasion, do something which, in the generic sense, he *can* do. The agent's actual freedom, therefore, is greater or lesser depending upon the number of restrictions which there are on his (existing) potential freedom.

Such restrictions can be *external* or *internal*. Restrictions of either kind, moreover, are either *preventive* or *compulsive*. The members of the second pair are interdefinable. To be compelled to act is to be prevented from omitting an action – and to be compelled to omit (forbear, abstain) is to be prevented from doing (acting).

External restrictions on freedom I shall divide into *physical* and *normative* (or *deontic*). I have already argued that whereas one can be by physical obstacle prevented from doing various things – as, for example, a chained prisoner from escaping – and thus compelled to forbearance, one cannot rightly be said to be physically compelled to do anything, and therefore one cannot be physically prevented from forbearing anything either. This is a noteworthy asymmetry inherent in the concept of free action.

External normative restrictions on an agent's freedom are those prohibitions of a legal or moral character which are instituted in the social order, or orders, to which the agent belongs. Let it be observed in passing that the term "prohibition" is normally applied to *actions* which it is forbidden to perform. Prohibitions apply symmetrically to *omissions* too, however, in which case they

are more commonly called "obligations." (Prohibition to do = obligation to omit doing; obligation to do = prohibition to omit.)

Internal restrictions on freedom can be divided into *psychological* and *normative* (*deontic*). By the first I understand "mental forces" such as desire and temptation, fear or aversion which, as the saying goes, either "irresistibly" compel us to do actions or constitute "insurmountable" hindrances to our embarking upon them. Psychological compulsion (for example acting under a threat) can sometimes come to resemble physical compulsion in that it is questionable whether the compulsory behaviour should be classified as an "action". If we come to think that it cannot be thus regarded we do not impute responsibility for it to the agent. That is: we do not regard him as "free" or as an "agent" in relation to this particular behaviour. But not every case of which it is correct to say "he could not abstain" or "he could not bring himself to act" is of this character. Most cases are not, and of those which are it would be better to say that the notions of omitting and acting are no longer applicable to them.

If psychological compulsion and prevention relates to a *generic* action, then it annihilates ability and does not count as a restriction on existing potentialities of the agent. In the case of compulsion this means that the agent always, whenever there is an opportunity, does the action. He never omits it. This kind of compulsion which annihilates ability (to omit) is like an illness or an obsession which seizes an agent after he has once learnt to do a certain thing. Prevention which annihilates ability (to do) is more common. It is usually spoken of as "inhibition". The agent simply cannot bring himself to do a certain kind of action. Maybe he once upon a time was able to perform it, but later aquired an "insurmountable aversion." Then he not only never performs the action any more; he also no longer omits performing it. He *cannot do* it, and therefore he *cannot omit* it either.

Internal normative restrictions on an agent's freedom are the prohibitions which the agent acknowledges as his duty to observe. They can also be called self-imposed restrictions. But it should be noted that many such duties are societal norms which the agent has *internalized*, *i.e.*, adopted as ultimate reasons for his actions and abstentions. This means that he observes the prohibitions, *because* he thinks he ought to and not, for example, because he is anxious to avoid getting into trouble with the norm-authorities. It may be suggested that *all* self-imposed duties (prohibitions, obligations) are, in fact, internalized norms of external origin. Duties which the agent has, so to speak, invented for himself are not "real" duties but *decisions* or *resolutions* of his to adopt a rule for his personal conduct. Some such rules would be like *habits* (for example, always to go for a walk before dinner).

4.

Perhaps no man is absolutely free in the sense that he is never compelled to do or to abstain from doing anything which, in the generic sense, he can do. But let us stop for a moment to consider what such a free man, if he existed, would be like.

He would, first of all, never meet with any physical obstacle which prevents him from doing something which he can, *i.e.*, has learnt or knows how to do, should he choose to do it. That such is the case might be a matter of luck with this man – but it could also be due to either an instinctive or a reasoned avoidance of the obstacles on his part.

Second, he would be so constituted that no temptation is ever "irresistible," nor any aversion or inhibition so strong that he cannot overcome it.

Third, he would never feel compelled to act under the pressure of norms. This means two things. One is that he would never observe a prohibition prescribed by some authority because he feared the consequences of refusing to obey. The second is that he would never consider it his unconditional duty to obey any rule, either self-imposed or given.

Strength to overcome aversions and resist temptations may be regarded as praiseworthy features of a man's character and also as a mark of "freedom." But what shall we think of a man whose actions are never strictly bound by norms? He is not perhaps praiseworthy. But is he even free?

In trying to answer this question we should note that refusal to let oneself be compelled to follow rules does not preclude one's actions from being in accordance with the legal and moral and other norms of society. The agent may never be in a position where he has a reason to trespass – or if he comes to be in such a position he may have an even stronger overriding reason for acting in conformity with the norm. But he would never feel "bound" by the norm, either in the sense that he feels compelled to bow to the norm-authority's will, or in the sense that he makes obedience to the norm his self-imposed duty.

Norm-authorities have sometimes thought that the "true freedom" of their subjects consists in action conforming to the norms. It has also been thought that only action in conformity with self-imposed duty is "truly free."

Ideas like these need not be sheer nonsense or hypocrisy. One can try to support them by rational arguments. Such arguments would have to be conducted in axiological rather than in deontological (normative) terms. A norm provides a person to whom it is addressed with a reason for acting in a certain way. Reasons, however, can be rated as better or worse. One could make the goodness of the reasons a measure of the degree of freedom of the action. If one wants to argue that true freedom consists in norm-bound action, one would have to argue that the reasons provided by norms of a certain kind, be they the laws of the state or the laws of our moral consciousness, are *the best reasons*

on which a man can act. The pros and cons of such arguments, however, will not be examined here.

5.

It is often thought that the sign that an action was performed freely is that it could have been omitted – and, reciprocally, that an omission was free if the agent could have performed the omitted action. Whenever I say truly "I could have acted otherwise" what in fact I did I did freely.

No doubt this idea touches the core of human freedom. We have no reason to doubt its *truth*. But we have, I think, great difficulties understanding precisely what it *means*.

In the justly celebrated chapter on free will in his book *Ethics*, Moore suggested that "I could have done otherwise" means that I should have done otherwise had I chosen to do otherwise.[3] Thereby he drove a wedge between freedom of action and freedom of choice. If my choice, too, was free I could presumably have chosen otherwise. When faced with the question of what *that* means, one thing Moore suggested was that "I could have chosen otherwise" means that I should have chosen otherwise had I chosen to choose otherwise.[4] Thereby the problem of freedom was only pushed one step back. In order to escape from an infinite regress Moore resorted to an epistemic move: I did not know for certain beforehand which choice I was going to make, and in this sense of "not knowing beforehand" it was *possible* that I should choose differently, that I *might* have chosen differently.

Moore, however, was not sure whether this wedge between freedom of action and of choice was necessary for solving his problem. He "confessed" that he could not feel certain that the truth of the statement that we could have done what we did not do was, in many cases, "*all* we usually mean and understand by the assertion that we have *Free Will*."[5] Let us therefore lay aside the problem and concentrate on the phrase "could have acted otherwise."

To say that I could have acted otherwise (omitted the action which I performed) is to affirm that my action was *contingent*. But in what sense was my act "contingent"?

No one would say that an action which I perform is logically necessary. So every action is, *ipso facto*, logically contingent. This is *a* sense of "could have acted differently," but hardly a very interesting one.

The statement that no action is logically necessary is not, however, as clear and uncontroversial as it may seem at first sight. Given an action of a kind or type which I *can* perform, and given an opportunity for performing it, I shall, of logical necessity, either do or omit it right then. To count omission as a mode of action makes good sense. So why not also count the disjunction "do or omit" as a mode of action? This would then be a "tautologous action" which an

agent will necessary "perform," provided that he has the required ability and that the occasion provides an opportunity for exercising it. Given these prerequisities, he could *not* "act otherwise." Such actions are not "free." But they are actions of a very special kind, and it would be quite feasible to refuse to call them "actions" at all.

I have decided to do something. There is no doubt about my ability to do the thing in question. I do not reverse my decision. Nothing preventive intervenes. The opportunity is there. Is it not then, relative to these assumptions, logically necessary that I perform the action? If one is prepared to ascribe every conceivable failure to perform either to some preventive interference or to a reversal of decision ("change of mind"), the answer is "Yes." But the (logical) necessity of the action is then *relative* to assumptions which are themselves (logically) contingent. *Simpliciter* the action is a logical contingency. This is trivial. We feel instinctively that the meaning of "could have acted differently" is more interesting than this. But in what way?

Consider an action of a kind or type which I have learnt or otherwise know how to do. Then, normally, when I set myself (choose, undertake) to do it I succeed. However, I normally do not perform the action whenever there is an opportunity, but only sometimes. This is proof that the performance of the action is contingent – just as the fact that it is sometimes raining and sometimes not raining is proof that the fact that it is raining is contingent.

Are these facts about ability sufficient grounds for saying that an agent who on some occasion performed a certain action might also have omitted it, "could have done differently"?

One would wish to answer "No" to the question. That the action I performed was free must mean that I could *then*, on the very occasion for its performance, have omitted it. How can I know this? The fact that on some other occasion I omit the same (generic type of) action is no proof. So what does it mean that I could *then* have omitted it, acted otherwise?

The comparison with rainfall is useful here. The fact that it is ("happens to be") raining here and now is contingent by virtue of the fact that it is sometimes raining and sometimes not raining here. But this is fully compatible with the possibility that *whenever* it is raining this is due to some causes which make rainfall a (physical) necessity under natural law. Similarly, might not the fact that I sometimes do, sometimes omit, an action which I can do be compatible with the possibility that *on those occasions* when I do it I could not have omitted it – and on those occasions when I omitted it I could not have done it? If that actually were the case, would actions then be free? One is tempted to say "No."

Assume that I perform the action *for some reason*. Perhaps I was fulfilling a promise. The fact that I had given the promise was the reason for my action. Or perhaps I was complying with an order or request. The fact that I had been

ordered or requested to do something might then have been the reason why I did it. (Let us assume that the reasons why I acted actually are stated. This *need* not be so, since, for example, the "real" reason why I fulfil a promise need not be that I have promised, but may be something else (cf. below, p. 17).)

That an agent acted for a certain reason normally means that something was, for this agent, a reason for doing something *and* that he set himself (chose, proceeded, maybe upon deliberation) to do this thing *for that reason*. To say this is to intimate that he could, in fact, have acted otherwise. He could have neglected the reason and omitted the action. Or he could have performed the action for some other reason which he *also* happened to have. Or, finally, he could have performed or omitted the action but done this for no reason at all and not for any reason which he had. Normally, it is, as one says, "up to the agent" to act or not on given reasons. Action for reasons is *self-determined*.

But if he actually did not neglect a certain reason but acted on it, how could he *then* have acted otherwise? If the "then" is so understood that it, so to speak, "includes" the fact that he acted (for that reason), then he could, of course, not have omitted the action. One and the same occasion does not afford "logical space" both for performing and omitting one and the same action. "What is is necessary, *when it is*," as Aristotle said. Nothing *can be* otherwise from what it is. But it could, perhaps, *have been* different (from what it is). And this is precisely what we claim to be the case with most actions. (By insisting upon the "then" in the phrase "could have acted differently *then*" one can produce a kind of philosophical "cramp" or "frenzy" which blinds one to the distinction between "could have been" and "can be.")

But do we not sometimes say that a reason was *compelling* and that therefore I could not have acted otherwise. I, as we say, "had no choice." My freedom was restricted, the "freedom to the contrary" annihilated.

I give away a secret under torture. My reason for doing this can be that otherwise I could not have rid myself of a most horrible pain. In thus describing the reason it is presupposed that I suffered from the pain, wanted to get rid of it, and thought (or knew) that in order to achieve this I must confess the secret. The *pain* as such is no reason for my action. Its rôle is rather that of a *cause*. It "compels" or "forces" me to act for the reason mentioned. Was my action then free, *i.e.*, was it "up to me" to act or not to act in the way I did? The question can only be answered by considering a wider context than just this one occasion. If, on some other occasion or maybe several other occasions, I could withstand (in all appearances) an equal or even greater pain, then we would (probably) think of my action as free. One would say that I *can* withstand a pain of this intensity – "can" meaning now that I have the required ability (the generic "can do" mentioned above). This being so, it was still "up to me" to act on the "compelling" reason; my confession was a product of my

self-determination. I could have acted otherwise. But if I am notoriously bad at standing pain, the case may be judged differently. Not necessarily, however. Other persons are known to have withstood even greater pain; some to have let themselves to be tortured to death. Am I sufficiently unlike them to warrant the judgement that I could *not* have acted differently? The answer would depend upon further facts about me (and about those more heroic people). Maybe a sufficient number of such facts are known or can be ascertained so as to enable us to answer the question one way or another. But it may also be that a factual basis for a wellgrounded answer cannot be established. Then we simply cannot tell (decide) whether my action was free, whether I could have acted differently, whether it was "up to me" to perform or omit the action.

I got frightened by a bull and screamed. If I screamed in order to call for help or in order to frighten away the bull, I acted for a reason. I could then also have suppressed the scream and done something else instead. But a scream of fright can be "automatic," "mechanical," "uncontrollable," "a reflex." Then my reaction, screaming, is not an action. And there surely are such primordial reactions of fright – and also of delight.

Sometimes an agent performs an action *for no particular reason*. We agree it was an action; it was not done by mistake. Let us also assume that the action is of a kind which the agent does not always do, whenever there is an opportunity, "mechanically," like a reflex. So, in a sense his performance was contingent; he might not have done it just then, on that occasion. But does saying that he could, on that occasion, have acted differently now mean anything over and above that we do not know why he did the thing in question then (nor does he), but we know (and so does he) that on some occasions he does it, on others not? It does not make much sense to say that he was free or that he was not free to act differently *on that very occasion*. And this is so just because his action had no reason, was "fortuitous." If, however, what he did was something annoying or obnoxious we might ask him to control or watch himself better in future – and thereby we should give him a *reason* for *not* doing the thing in question "for no particular reason" on other occasions.

Cases of fortuitous actions are perhaps not very frequent. But assume that they become very frequent with an agent with regard to one or several types of action. He quite often does certain things without deliberating beforehand and without being able to connect them with any reason when challenged to reflect on them in retrospect. He cannot account for these actions of his. Can he be held responsible for them? Was he free to do or omit them? Shall we perhaps after all classify them with reflexes rather than with actions? Such questions may be interesting to consider – sometimes because they challenge questions of sanity and mental illness – but one should resist a temptation to force a clearcut answer to them.

To sum up: The phrase "could have acted otherwise," *i.e.*, "could have omitted what was done or done what was omitted" has not one but several (related) meanings. In the weakest sense the phrase is true of anything which can truly be called an action (or omission) and means simply that the performance and omission of actions are logical contingencies. In a stronger sense the phrase is true of the performance and omission of any (normal) action which the agent is able (has learnt to, knows how) to perform or omit. Then it means that there are occasions when the agent performs the action and other occasions when he omits it. In a still stronger sense the phrase is true when an agent *for some reason* performs (omits) an individual action of a type which he is (generically) *able* to perform but also to omit. Then the action (omission) springs from the self-determination of the agent. Of a good many such actions, however, the phrase "could *not* have acted differently," is *also* true – meaning that the reason which prompted the action was, as we say, compelling. Then the freedom of the agent was restricted. In marginal cases the restriction is so severe that we judge it impossible for the agent to have acted otherwise. This happens when we, usually on the basis of experience of analogous occasions, would deny that the agent has the *ability* to omit that which on this individual occasion he did. In cases, finally, when an action takes place apparently for *no* reason we sometimes look for (physical) causes and hesitate to call the behaviour (full-fledged) "action." Our attitude will then depend on the frequency and character of such fortuitous behaviour – and on how we evaluate it morally. It is doubtful whether we should call such actions "free" *when they occur.*

6.

Normally, we said, it is "up to the agent" *whether* he will act for such and such reasons which are there for him to act upon, or not.

But is it also "up to the agent" to have the reasons which he happens to have? If "up to the agent" means that the agent could choose, on a given occasion, which reasons to *have* for his action, the answer is "No." Such a choice simply makes no sense. But if the phrase means that he, normally, can choose which reasons to *act upon* (among those he has), the answer is "Yes."

The reasons for acting which an agent has, on a given occasion, are often "given" to him independently of his own (previous) action. An order could be an example – but also something "internal" such as a sudden wish to take some physical exercise or listen to music.

A man wants and shuns, likes or desires, hates or fears certain things and he knows, or thinks he knows, ways of securing for himself what he wants and avoiding what he shuns. By virtue of this he has (gets) reasons for and against certain actions of his. He has, moreover, been brought up to know what is

expected of him in various situations and he has been placed, or has placed himself, in positions connected with duties and rights in relation to his fellow human beings. His involvement in the social fabric constantly provides him with reasons for and against certain actions.

The existence of reasons for a man to act in certain ways are facts about him. They are not *his* makings in the same sense as his *actions for such and such reasons* can be said to be his makings, *i.e.*, result from his self-determination. But the majority of reasons an agent has for his actions are there as the result or consequence of human action, including the actions of the agent under consideration himself. Things have been done to him; he has for example been given a certain education or training or, on the contrary, been excluded from education or training. His tastes for various things have been cultivated, partly by others, partly by himself. He has by birth a certain place in the social order, and this place has been changed in the course of his life, partly dependent on his doings, partly independent of them. To the extent that the reasons a man has for his actions depend on his own actions in the past one may say that it has been "up to him" to have them or not.

In these facts about the reasons is reflected the way in which the range of a man's actual freedom, *i.e.*, of things he will do if he chooses to do them, will wax and wane as a result of what happens to him or how he "builds" his own life. It is also possible to say that the *more* reasons an agent has for and against actions which he can do, the *greater* his freedom of action (choice). But greater freedom may also imply greater difficulties and uncertainty in taking decisions – and in this way freedom of choice may inhibit action.

7.

The word *reason* in English refers to the rational faculties of man. A reason for action is something which, *prima facie*, it is rational or reasonable to act upon. The two adjectives, incidentally, are not used as synonyms in ordinary language. "Reasonable" carries a stronger value-load than "rational." Of some actions which took place for a reason one would say that they were rational but not (very) reasonable.

A reason for action can also be called a *ground*. In German, a reason is called *Grund*, or sometimes *Vernunftsgrund*, which intimates a relation to the faculty of reason. In Swedish there is, in addition to the word *grund* also a word *skäl*. To both one can prefix *förnufts-* ("of the reason"). Adding the prefix in German or Swedish serves the purpose of distinguishing ground as reason from ground as cause. But the reason–cause distinction is not a clear one – neither in language nor at the level of concepts.

What then is a reason for action? One could answer that a reason is anything to which the action is an adequate response. But what does this mean?

A reason can be given to an agent in the form of a challenge the meaning or purpose of which is that the agent should react to it in a certain way. The response is expected, maybe even required or obligatory. For example: I do something. Why? The answer is that I had promised to do this thing. The person to whom I gave the promise expects this action from me; it is my duty (obligation) to him to perform it. Or, I stop my car in front of the red traffic light. Why? One is forbidden to drive against it.

It should be noted that the fact that a challenge makes its appearance "in the world" (a command being shouted out, the red light appearing in front of my car) is not, by itself, a reason for any action. It becomes a reason in virtue of the fact that the agent to whom it is addressed is aware of and understands (the "meaning" of) the challence, *i.e.*, knows how to react to it adequately. Whether he then reacts or not is another question.

The presentation of the challenge has, so to say, to be sieved through the medium of the understanding in order to become a reason for the agent.

A reason is often also presented in the form of something an agent covets or wants (to be, to do, to get, to have or to promote) in combination with an opinion of his that a certain action is conducive to or otherwise useful for the attainment of his goal or end of action. The action which takes place for that reason could be something very simple and direct like opening a window to get fresh air, or it could be something complex and remote like registering for a course in order to promote one's education.

Ends of action are often considered means to some remoter ends. Having the latter in view is then a reason for pursuing the former. The ultimate ends are things a man cherishes as good in themselves. They are his "ultimate goods" or "ultimate values," things which, as we say, give "meaning" to his life. Which they are and how a man chooses to pursue them will vary from man to man. They are not necessarily things we all agree are noble or praiseworthy.

It may be suggested that the ideally rational agent is one whose reasons for action are always anchored in ultimate ends. Perhaps no man can live up to this ideal. How many of us can tell which our ultimate ends (goods, values) in life are? But the farther towards something ultimate we can push our answers to the question *why* we undertake to do what we do, the better do our reasons for action deserve to be called rational.

If by the "apparent good" of an agent we mean all that he values as good in itself, then we could say that, ideally, a man's reasons for action should be those things which make his actions rational from the point of view of his apparent good. If, furthermore, one distinguishes between a man's apparent and his real good, one can go a step further and say that a (truly) reasonable man is one whose actions are based on care for his real good.

One may also wish to say of such an ideally reasonable man that he has attained the highest degree of freedom. But I shall not pursue here this moralistic thinking about reasons, rationality, freedom, and the good.

Sometimes we say that the reason a man has for some action of his is really *no reason* why he should do it. This can mean several things. It can mean, for example, that his opinion (belief) about the conduciveness of a certain action to a certain end is erroneous (false, superstitious). By making him "know better" the means-end connections we can influence his freedom and therewith also his actions. But it can also mean that what *for him* is a reason for an action would not be a reason for us; for example because we censure or disapprove of something he aspires after and wish to change his valuations – not his opinion about the means but his pursuit of ends.

8.

One distinguishes between *reasons* and *motives* (for an action). Ordinary language does not uphold this distinction very clearly. Reasons are often spoken of as motives, and vice versa. One must not be pedantic about the use of the words. But some conceptual observations may be called for.

Motives have not the same link with the rational faculties of man that reasons have. Motives can be irrational. And irrational motives can prompt a man to act perfectly rationally for reasons. I shall try to explain.

An important class of motives are constituted by "passions" such as jealousy, hatred, greed. They tend to "move" people to action; under their influence people do various things. That a man, for example, hates another man will usually manifest itself in various ends of action which he then pursues. He may want to inflict harm on the object of his hatred. Having such objectives is not so much a "consequence" of his passion as something "constitutive" of it; his objectives are the criteria on the basis of which we attribute the passions in question to him. If now a man with such objectives thinks that a certain action will be conducive to their attainment – say, harm the person whom he hates – then the fact that he has this objective and opinion will constitute a *reason* for him to do the action in question. It is of such reasons that we sometimes say that they are "no reasons" on the ground that we disapprove of the objective and of the feeling which it manifests. "You hate him and doing this to him would harm him, I agree; but that is no reason why you should do it. I realize that you hate him considering what he has done to you; but try to understand him and you will feel compassion for him and pity him rather than hate him."

The "good" passions are motives for action, too. Supreme among them is love. The lover will do a number of things for the reason that he considers them promotive of the happiness and well-being of the beloved. His actions are

motivated by love, but one would not normally call his love a "reason" for what he does.

There are other ways, too, in which one can mark a distinction between motives and reasons. Having a reason involves *understanding* something: for example the meaning of a practice (promising, answering questions) or a causal relation between means and ends. Motives may be "blind" like sometimes love and hatred, or have an animal character like hunger or thirst.

We need not here uphold a sharp separation between motives and reasons, however. By the *motivation(al) background* of an action I shall understand the complex web of factors (motives, reasons) to which we refer when we explain why something was done or omitted, or of which we say that they led to or prompted the action or made the agent act or moved him to action.

9.

In a good many cases of simple actions the agent has just one reason for doing or omitting it. But in other cases the motivation background of an action is *complex*. The complexity can be either one of *number* or one of *strength* of the reasons.

The fact which I call the *complexity of the motivation* is well known to psychologists and psychoanalysts. As far as I can see, this fact has not been much noted in recent philosophical discussion of action and action-explanation. This is a limitation which we must overcome.

There can exist many reasons why an agent should act as he does. For example: An agent does something which he has promised to do. But he also expects a reward or a service in return from the promisee. Would he have fulfilled his promise had he not had that expectation?

Sometimes there are reasons *for* but also *against* a certain action. (A reason *against doing* something is a reason *for omitting* it.) For example: The thing the agent had promised to do and for which he is expecting a service in return is perhaps something shady, disreputable or, maybe, even criminal. In this situation the agent has to "form a balance": he has to "weigh" the "sum total" of the reasons for and against the action. How he then acts shows which one of the (sums of) reasons was heavier (stronger).

Also among the reasons, if there are several, which are all for (or all against) an action some may be stronger than others. And the strength of a particular reason may be influenced by the presence in the motivation background of other reasons for or against the action. For example: considering the disreputable character of the act and the agent's awareness of this, the fact that he had promised was a rather weak reason why he should ("after all") do it. But the expectation of reward may have constituted, for him, such a strong reason for the action that, because of this, he did it. Maybe he did not attach

any weight at all to the fact that his action was a fulfilment of a promise *as a reason for his action.* ("I know full well that promises of such acts need not be kept.") But the fact that his action was the fulfilment of a promise *and* disreputable may be highly relevant to his expectation of a reward. ("If I promise to do this shady trick for his benefit, I am sure he will reward me.")

When reasons are balanced against each other and one found heavier than another, contrary reason, the first is said to be *overriding* in relation to the second. An overriding reason is not necessarily a reason of the kind we call *compelling*, nor vice versa. A reason can be called compelling also in the absence of any contrary reason. Often at least, in calling reasons compelling one excludes them from deliberation. They leave no choice open to the agent.

When, in deliberation or in retrospect, reasons are rated for strength they are often called good or bad, better or worse. But rating reasons for goodness can also be a moral evaluation of them. And a morally commendable reason for an action is often called "strong." But the strength which on moral or other grounds we attribute to reasons must be distinguished from their (actual) strength in moving agents to actions and abstensions.

10.

When the motivation background is complex one can usually not point to any *one* reason when trying to explain why the action was performed or omitted. A full description of the background may be needed for the sake of understanding what took place. This description will also contain estimates if the relative strength (weight) of the reasons known to have been present. Some of the reasons for the action will be thought to have contributed more, others less to its actual performance. Some may have been completely "inefficient," others again so strong that they alone, in the absense of all the others, would have conquered, overridden, the restraining influence of possible reasons against the action. Then we say that the action was *over-determined.*

The existence of reasons for an action is an ambiguous concept. When an action is judged from "outside," *i.e.*, by someone other than the agent himself, it is often said that there were (good) reasons why the agent should not have performed it. But the agent did not consider them. He was not aware of their presence or did not understand their significance. We sometimes blame an agent for such ignorance. "He ought to have known what this meant" (for example the hooting of a horn).

Reasons of this kind, I shall say, were *not present for the agent* (did not "exist for him") at the time of his action. They may, in various ways, be relevant to the *evaluation* (blaming or praising) of the action. But they are not relevant to its *explanation* since they do not belong to the motivation background of the action. And the same is true of those reasons which were present

for the agent, which belonged to the motivation background, but which he chose to ignore. We often blame an agent for not having taken them into account.

Consider the following example. I am invited to a party. I decline, giving as a reason that I have another engagement. My reaction (declining the invitation) is a perfectly adequate response in view of this fact. It is a valid *excuse*. But is it the reason *why* I declined? The party would have bored me. I am shy – I hate to be in the presence of so many people. I might have met X at the party; I dislike him intensely; I am, in fact, afraid of meeting him.

All the things mentioned are reasons for declinig the invitation. But I did not mention any of them when I was challenged to explain why I declined. Perhaps I did not think about them very much, since I had a valid excuse. Maybe it did not even occur to me that I might meet X at the party. If this is *really* so, *i.e.*, that it did not occur to me, then the fact that I would have feared meeting him was not one of the reasons present for me. But is it quite certain that the possibility did not "occur" to me? Surely I knew that X is a great friend of the family to whom I was invited, that he often visits them. Since I knew this, I *must*, "subconsciously," have known, too, that I was likely to meet him there. Who is to tell?

We shall presently have to say more about such cases. Here we only note the following two things. First, that it is not always clear and easy to tell which reasons for or against a certain action shall count as belonging to the agent's motivation background. And second that reasons which undoubtedly belong to this background – for example that I am a shy person and do not like big parties – do not necessarily "contribute" to my actual conduct. It is, in other words, important to distinguish between reasons existing for the agent and reasons influencing his action – between *existing* reasons and *efficacious* reasons. An existing but not efficacious reason can serve as an *excuse* for doing something. But it is not part of the explanation. Only of efficacious reasons do we say that the agent acted *for those reasons* or *because* of them.

11.

To explain an individual action is to answer the question why this action was performed.

In its general form the formulation covers several *types* of action explanations. The only type which will be discussed here is explanations in the terms of reasons. Such explanations I shall also call *understanding explanations*.

Another type of explanation is *medical*. An explanation of this type attributes an action, or a failure to act, to a diagnosed illness or deficiency – due perhaps to something "somatic" and thus to a "cause" rather than to a "reason". Still

another kind of explanation is *sociological*. It is concerned with abilities, or the lack of abilities, rather than individual actions. It explains, for example, why an agent can or cannot do certain things because of economic status, education, or social position.

Action explanations of the types here called "medical" and "sociological" are in a certain sense *scientific* explanations. They usually have a background in some *theory* about man or about society. Their purpose is often *to cure* an agent of some illness or to remove some hindrance to his development. Reason-giving explanations, by contrast, are not typically what we would call "scientific." The purpose they serve is usually *evaluative*. Does the agent deserve blame or praise for what he did? The answer may crucially depend upon the reason which he had. Hence we must *understand* the action before we can *judge* the agent.

12.

In giving an "understanding" action explanation it is presupposed that the action has been correctly identified as an action of a certain type and that the agent actually had the reasons mentioned in the explanation. The action and the agent's reasons are, so to speak, the facts of the case. The presupposition that they have been established, however, is not trivial.

What the behaviour of the agent was, or what it caused to be, may be identified as a result of a good many generic actions which, however, cannot be imputed to the agent as *his* actions. The agent's arm moved in a way constituting a signal. Did he signal? Perhaps he had not the faintest idea that he was doing such a thing. Then the action cannot be imputed to him. But if he knew the significance of the movements as a signal we can impute the action to him even if he did not "mean" (intend) to signal but meant something else, say to reach out for an object. If he did not mean to signal, he had no reason for signalling, and his action cannot be explained (understood) as that of giving a signal. We may blame him for his action ("you should have realized —"), but in order to explain it we must look for another way of identifying it. We must try to identify it as an action for the doing of which the agent had some reason(s).

Our identification of an action for the purpose of explaining (understanding) it is thus guided by what we think of as possible reasons for it. The reasons for signalling are different from those for reaching out for some object. We know, roughly, which they *are*. *Had* the agent reasons for an action of either type? He may have had for one, or for both, or for neither. If he had reasons for both, were the reasons for both efficacious? Reasons which are not efficacious do not "contribute" to the explanation.

So our problem is: how do we identify efficacious reasons?

To this question I shall give an answer which at first may be thought shocking. The efficacious reasons are those in the light of which we explain the action. I maintain, in other words, that one cannot separate the question of the *efficaciousness* of the reasons from the act of *understanding* the action as having been performed for those reasons. This means that the truth of the action explanation has no basis in facts *other* than the understanding itself of the action in the context of its reasons.

The obvious objection to this is that it seems to open the gates for boundless subjectivism in action explanation. Must we not be able to discriminate between understanding and *mis*understanding, when explaining an action, or at least between a *better* understanding and a less good one? What then are the criteria for making these distinctions if not some facts about the action and the reasons on which our understanding of their connection may be based?

13.

Understanding something requires a subject, somebody who understands. When there is a wide consensus about how something should be understood one also talks of understanding in an impersonal, derivative, sense: "It is (commonly) understood that –."

When I say that to explain an action is to connect it in the understanding with the reasons for its performance, *whose* understanding am I then thinking of? There are two possibilities to be considered:

Understanding can be by the agent himself or by one or several outside observers of him and his action. In the first case we speak of the agent's self-understanding; in the second I shall talk about "outside understanding" or "understanding from outside." One could also call them first-person and third-person understanding, respectively.

It is clear that self-understanding is, somehow, *basic* to action explanation. Normally, an agent knows what, on a certain occasion, he did, *i.e.*, under which description(s) his action is intentional. He also knows which reasons there were for him to act. In normal cases, moreover, he knows for which reasons he acted. If we, outsiders, wish to know why the agent did what he did, the obvious way to get to know this is by *asking* him.

Of most actions, no explanation is ever required. Should the agent stop to reflect why he did a certain thing he would know the answer, and should he be asked he would give it without hesitation. Nobody would have a reason to doubt it. There would be complete agreement, consensus, about the case. It is in such agreement that the "truth" of an action explanation, if an explanation be required, consists.

Many cases, perhaps even a majority of cases, when an explanation for some reason or other is required, are not cases where there is consensus – at least

not initially. An outsider *wonders* why the agent did what he did. (He may also wonder which action to impute to the agent, how to identify the action. But this difficulty we now assume to be solved). He may know something about the agent's reasons for the action but he can also see reasons against doing an action of this kind and wonders why the agent did not *omit* it. He asks the agent and the agent's answer does not satisfy him. The case looks "suspect." There must have been other reasons why he did it and which he conceals from us, we think. Or we say that he did it, not for the reason he gave, but for another reason which we know he had.

Consider our previous example of the promise (above, p. 17). The agent had given a promise. This was a reason for doing what he did. But what he did was something shady, maybe criminal, something one ought not to do. This he presumably understood was a reason against doing it. However, by doing the thing he greatly obliged the promisee and could expect a service in return. This he obviously knew too and that gave him another ("selfish") reason for doing what he did. He says, however, that he did it because he had promised. Did he not realize that what he did was something bad? Yes, but "a promise is a promise." We are left wondering.

How should a case like this be decided?

Perhaps the situation is quite clear. The agent is openly lying. He knows full well why he did what he did and that this was not for the reason he gave us. Then his self-knowledge need not conflict at all with the outsider's suggested explanation of his case. There is in fact consensus, although it is "tacit."

The situation need not be like this, however. The agent may, as we say, be "lying to himself," too, about his reasons (motives). He fulfilled his promise and did the shady thing because of a selfish calculation, but he does not "acknowledge" this (even) to himself. Or he honestly misunderstands his own action – thinking, for example, that the sole reason why he fulfilled his promise was that he had promised and not that he expected to be rewarded.

(The border between cases of "lying to others" and "lying [also] to oneself" may not be sharply distinguishable.)

On what grounds could an outsider defend his claim to understand the agent (his motives) better than the agent himself? The outsider would, for example, refer to his knowledge, presumably based on past experience, of the agent's character. Perhaps he says: "He, the agent, *is* that kind of person who gives and fulfils promises only when this is clearly to his own advantage. The moral obligation to fulfil promises does not mean anything to him. We know this." The outsider thus views the conduct of the agent in this particular case in the broader setting of the picture we have of his character. The explanation of the action offered by the outsider is more consistent or in tune with the rest of our knowledge of the agent.

The outsider's view gets further support if it turns out to be a safe basis for predictions. "You will see: when in future he promises something he will disappoint the promisee, unless he also has a selfish motive for fulfilling the promise. He is not to be relied upon." The prophecy may fail in some cases, but if it holds in many cases this supports the explanation which the outsider offered of the particular case in which he disputed the agent's own explanation of his action.

14.

In case of disagreement it may of course happen that the agent convinces the outsider that the latter has misunderstood him. The outsider is then, so to speak, "converted" to the view of the agent. This case may be quite common but not of much interest either from a philosophical or from a psychological point of view

Of more interest is the case in which the outsider stands by his view and tries to convert the agent to a new self-understanding. The outsider says perhaps that the agent's lips profess that he did the action for the reason *X*, but in his heart he knows that he did it for the reason *Y*. Maybe we can convert him and make him "confess" the truth.

There is an idea that the agent must be the supreme judge, the highest authority in the matter. He and he alone can see the truth directly. The outsider's evidence for *his* explanation can only be external and indirect. Agreement with the agent's self-knowledge therefore seems the ultimate test of truth in the matter.

What kind of argumentation would the outsider resort to if he tried to convert the agent? Mere *persuasion* would not be fair. If it succeeded, *i.e.*, led to consensus, it would be a result of "brainwashing." What is a brain-washed agent's self-knowledge worth as a testimony? Even if we do not dismiss it as completely worthless, we would hardly accord to *it* "highest authority." The highest authority is now in the hands of the outsider (the "brainwasher").

The *rational* arguments which the outsider could use would be, roughly, the same grounds and evidence on which he based his initial disagreement with the agent's professed explanation. He would, for example, try to make the agent see his present action in the setting of a larger fragment of his life-history. He would point to incidents in the agent's past which are "public knowledge" and which the agent would not deny. He would also hold up for him the image of his character which others have formed and ask the agent to ponder the facts which led to the formation of this image and to compare it with his self-image. He may warn him of his future actions, ask him to watch himself better.

Obviously, the border between rational argumentation and "brain-washing" is not always sharp. This being so, why should we think that the "internal

evidence" which the agent professes to have after a "conversion" has a privileged position in relation to truth (correctness of understanding)? Perhaps there is no good reason for thinking this at all.

Assume that a "conversion" takes place. The agent says perhaps: "I now admit that I did not do it because I had promised but because I counted upon a service in return." Or: "The reason why I did not go to the party was that I surmised that *X* was going to be there; the appointment I had could easily have been cancelled or changed; giving it as a reason why I declined the invitation was pretence only." And assume that we do not challenge the sincerity of these new explanatory declarations by the agent, but accept them.

The question of philosophic importance is now: How shall we *correctly describe* the imagined situation? Shall we say that *now* the agent sees the truth about himself? It, the truth, was always there to be seen although hidden from the agent's sight by the veils of his self-deception. When the veils are removed *he* sees clearly what the outside observer had already sighted, although the latter could not be sure of the veracity of his impression until he had it confirmed by the agent himself? *Or* shall we say that the agent now sees his former action in a *new* light, that his selfconsciousness has changed, and that he has acquired a new understanding of his past? Shall we, in other words, say that a connection (between an action and its reasons) which was already there has been discovered, or shall we say that a new (different) connection has been made?

It should be noted how permeated by metaphor the talk of truth is here. The truth was there to be "seen" ("in his heart"), but it was "veiled." When the "conversion" had taken place it was "revealed" to the agent, who, as it were, then "recognized" his "true self."

We are in the neighbourhood of what may be called the *epistemology of psychoanalysis*. A psychoanalyst would perhaps speak of a subconscious understanding by the agent's super-ego of the connection between the action and the reasons. The existence of this connection would then be brought to the surface of the consciousness of the ego which had repressed it. But this is a metaphor too.

It is tempting to resort to such metaphors as those we mentioned. They almost force themselves upon us. They are good *metaphors* and when used as such may be perfectly innocuous. The danger is that their use gives birth to *conceptual* mythology and mystification. One builds a "theory" of the workings of the subconscious, a "dynamic psychology." Here the task of the philosopher sets in. It is a task of "demystification." And this means a task of trying to describe the actual situation in terms which do not mislead. This is difficult.

In order to see how misleading talk of truth can be here let us ask the following question: What is supposed to have been veiled, the agent not to have seen? And let the answer be: the connection between the action and the reason

which made him perform it. But this connection had not yet been established. (Unless, of course, he lied "openly.") Because "establishing" the connection means understanding the action as having been performed for that reason. So under the veil there was in fact nothing to be seen! The object of vision was created in the very moment when the veil was lifted! What is *now* established, *viz.*, the connection in the understanding, simply was not there *then*.

The assumption is that the agent did not lie about his reasons when first asked to explain his action. If he did not lie he was sincere. But how can he have been sincere since later he admitted that the reason was something different? Unless we wish to say that he was brain-washed we must, I think, insist that he cannot have been *quite* sincere. He was, so to speak, half sincere, half lying. How shall this state then be described?

Consider again the example of the promise. If we attribute its fulfilment to a selfish expectation by the agent, the agent must somehow have had this expectation at the time of the action. Otherwise we could not say truly that there existed this reason *for him* for fulfilling the promise. He must have known, for example from previous dealings with the promisee, that he was doing something for which a service in return could be expected. Perhaps he did not think of this at the moment of his action. Maybe he felt "ill at ease" in face of the shameful thing; the thought of a service just "flashed" before his mind but was turned aside by the voice of conscience which said "you promised and cannot deceive your friend." This, for example, would be a description of what it is to be half-sincere when one has to explain one's action. The description shows *in which sense* the connection between the action and the selfish reason for doing it *was already* there from the beginning, albeit in an "embryonic" form, and *not only* from the moment of conversion.

It will be helpful here to warn against a temptaion to *insist* upon the existence of an explanation of any action which has a complex motivation background. The complexity may not consist only in the fact that there are *many* reasons, or reasons *for* and *against*, or reasons of various *strength*. "Complexity" can also mean that the background is *opaque*. And here opaque does not signify merely that we cannot *see* through the web of motives but that the motives *are*, in fact, confused. The opaqueness is, so to speak, "ontic" and not only "epistemic." When we then explain the action in the setting of its reasons (motives) we actually *create* an order where before there was none.

I shall therefore say that what happens in a "conversion" of the kind which we are considering is that the agent connects in his understanding *in a new way* some action of his with the motivational background for its performance. He explains his action differently – not because new facts about its reasons have come to light but because facts already there are connected (arranged, articulated) in a new way. If this new understanding is called better, more correct or more true, than the previous one this is because it matches the

broader frame of facts about the agent's past history in which the outsider had from the beginning been reviewing his present action.

In view of what has been said, what happens to the idea of the agent as supreme authority in understanding his own case (action)?[6] I think we must say that it withers away. The conversion is not a revelation of *truth*, but a reaching of a *consensus*.

The idea of the agent's authority has, of course, a rational foundation. This, however, is easily misinterpreted. The agent is likely to know more facts about the case than the outsider – particularly about existing reasons for his action. Therefore the outsider who distrusts the explanation of the agent will have to elicit information from him. The keys to a new understanding of the action are thus, in the main, in the hands of the agent himself and have to be obtained from him. But as for the new understanding itself, the agent is not necessarily better equipped than the outsider. The outsider may be superior. To neither of the two belongs exclusively the right to pass a final judgement.

15.

Assume, however, that no conversion takes place but that the outside observer stands by *his* explanation of the agent's action. Does this mean that the case remains undecided?

It is good to remember here that "decided" means that consensus is reached. It does not mean that the agent upon scrutinizing himself testifies to the truth in the matter.

But what is required in order that we may talk of consensus having been reached? Is it necessary to have the agent's endorsement of the outsider's explanation? Once we have demolished the idea of the agent's privileged position with regard to (access to) truth, the question is worth considering. It is clear that in normal cases the agent's agreement is desirable, even essential. If we come to think that his professed self-understanding can be ignored, we must have special reasons. *One* possible reason is that we are convinced that he is lying – and thus really agreeing with us. But this possibility we shall here ignore as being of minor interest. A more interesting case is when we judge the agent's character so morally corrupt or perverse that he is unable to give a coherent and honest account of the motives and aims of his actions. We simply disqualify him as a judge in his own case. Only the opinions of outsiders now count for obtaining consensus about how his actions are to be explained. And all outsiders may, in fact, agree – with the possible exception of some whose judgement we think, on independent grounds, cannot be trusted or can be ignored. Then the case is "decided."

That cases like this occur cannot be denied. But there is something tragic about them. That somebody else should have supreme authority in cases which

concern my "inner life" may be thought humiliating. May not such an authority misuse his position for "brain-washing" – perhaps with a view to furthering uniformity in people's thoughts and actions? And may not this lead to the gravest injustice in treating a person? Of these dangers we have good reason to be aware.

How much easier would not things be if we could believe in an absolute truth in these matters, a truth which exists independently of what anybody thinks about the reasons for our actions? It is characteristic that those who misuse their authority when they disqualify the testimonies of the agents often do this in the name of a "higher" truth, perhaps sanctioned by "science," which the recalcitrant agent is been forced to accept. And it is also characteristic that those who resist often seek comfort in the belief that there is an "inner" truth to which they *alone* have access and which they know. The insight that there is no such truth, neither "inner" nor "outer," is the weapon with which we must try to fight both the self-righteousness of excessive subjectivity and the pretensions of false objectivity in matters of understanding human action.

16.

To explain an action is a facet of understanding the agent as a *person*. The same holds for the imputation of actions to him, and for the attribution to him of reasons for actions.

One can distinguish layers of facts about an agent attributed to him in the understanding of him as a person. Facts of an inferior layer are often unquestionably taken for granted in efforts to establish facts about him on a superior level. Thus, for example, we may without question regard it as a fact that he did a certain action and also that he had such and such reasons, but be hesitant about the explanation. Did he do it for this reason or for that one? This may lead us to re-examine the already accepted facts of the inferior level. Perhaps we had mistakenly imputed to him the action, *i.e.*, his behaviour was not intentional under the description we had first given to it.

In attributing reasons for action to an agent we normally also attribute to him various abilities, beliefs, desires and inclinations, the understanding of institutions and practices of the community, and other things which characterize him as a person. Some of these features may date far back in his life history. They constitute a kind of background or "program" which has to be assumed if certain things he did or which happened to him shall count as reasons for subsequent action (for example, that he understands a certain language). These other things, then, speaking metaphorically, are "inputs" playing on the "keyboard" of his programmed personality. His action is the "output."

SECOND LECTURE

1.

Not all actions are performed for reasons. Actions can be unintentional, done by mistake, or "for no particular reason." Some such actions shade into reflex. If we wish to explain them we have to look for *causes* in stimulations of the agent from inside or outside his body. From the point of view of their explanation, these actions are movements, or the inhibition of movements, of the limbs and organs of the human body.

Actions which are performed for reasons also have a "bodily aspect." As its *primary* form I shall regard overt ("visible") movements of the body or some parts of it. These movements may effect further changes outside the body. Some such effected changes are normally used for *identifying the action, i.e.*, for telling *what* the agent did – for example opened a door. They are what I have called elswhere the *results* of the action.[7] Further changes effected by the results of actions I shall call (causal) *consequences* of those actions.[8]

In some simple cases the overt bodily movements themselves are regarded as results of an action – for example the action of raising one's arm. But more often the bodily movements are only (causal) prerequisities of (the results of) an action. These overt prerequisities have in their turn a covert background in the tension and relaxation of muscles. Muscular activity again has a causal background in processes in the nervous system. In the last resort, causes for these processes may be sought in stimulations of the nervous system from outside the agent's body. In this way the causal prerequisities for (the results of) our actions may be traced back to things which took place "in the world" outside our bodies and independently of us (our actions).

Not every human action results in a change in the world. Preventive or suppressive action, if successful, results in a not-change. Such action has nevertheless a physical (somatic) aspect, the characteristic form of which is muscular tension. For example, I press my hand against a door, thus preventing it from opening when someone else is trying to push it open.

There is a noteworthy asymmetry between *performance* and *omission* of action in relation to bodily manifestations. In the normal cases, omissions do not require any (physical) effort. They lack a somatic aspect. Omission of actions for which there existed no reason, for or against, would hardly ever be even noticed or require an explanation. The typical quest for explanation of an omission has the form: Why did an agent not do this or that for the doing of which he had a reason and opportunity (and which he can do)? And sometimes the answer is that he was prevented by an outer or inner *physical* factor (force).

That every action (other than omission) should have a somatic aspect is, I think, a conceptual or intrinsic feature of action. One can imagine "action at a distance" – for example that people could make things move or fall to pieces

just by looking at them or by pronouncing some words in a low voice. Looking too is "somatic," and so is subvocal speech. But what about the possibility of causing changes to take place by "mere" thinking or willing? What would this mean, if not some exertion of bodily effort such as frowning, clenching one's fist, closing one's eyes, compressing one's lips, *etc.*? One can imagine that such changes in the soma would effect changes outside the body even though in fact they do not do so. But a concept of action which is completely detached from somatic change would no longer be *our* concept of action.

I am not denying that there are *mental acts* and that some of them, such as imagining or thinking, are subject to the will. But the results of such action – if we call it by that name – are not changes and not-changes "in the world." Pure mental activity, as we know it, is therefore conceptually different from what here, in comformity with common usage, I call human action.

2.

There was a time when one did not know anything about the rôle of the nervous system in relation to muscular activity and overt bodily movement. *Logically*, it is of course contingent that there exists a causal connection between the two at all. Suppose that this connection had not (yet) been discovered, that we still lived in "blissful ignorance" of it. Would this have been relevant to the problem of freedom of action or of the will?

The question is worth asking, and in one sense of "relevant" the answer is: "Yes, probably," because it is certainly not a historical accident that the form in which the problem of free action has tormented philosophers for the last three centuries or so dates from the very time when the fundamental discoveries were made concerning the physiological mechanisms of the body, among them the nervous system. Descartes holds a key position in these developments. It was under the influence of the "new philosophy" of mechanistic determinism, the "scientific revolution" of the seventeenth and eighteenth centuries, that the problem aquired the typical form which it has retained to this very day, *viz.*, whether one can "reconcile" the idea of free action with the idea of a strictly deterministic course of events in nature.

Did the problem then not exist before Descartes? In Ancient philosophy we find discussion of determinism and also of voluntary action, but not much discussion of the two in relation to one another. In the Christian philosophy of the Middle Ages our problem has a definite ancestor, the question how to reconcile the notion of man as a free agent with the existence of an omnipotent and omniscient God.

It is interesting to compare these two variants of our problem, the "theological" and the "scientific" – as they might be called. When the idea of an omnipotent and omniscient God gradually withered away, the rôle which it had

exercised in the intellectual imagination of a culture was taken over by the idea of mechanistic determinism. This latter is now in its turn gradually being eroded under the influence of scientific developments. These developments too are likely to affect the form which the problem of freedom is going to assume and the rôle it is going to play in the philosophy of the future. For the time being one can only speculate about this, and we shall not do so here.

3.

Philosophers may be divided into two main groups depending upon whether they regard freedom (of action) and universal determinism (in nature) as compatible with one another or not. Philosophers of the first group are said to defend a *compatibility thesis*, those of the second group an *incompatibility thesis*.

A supporter of the view that freedom is incompatible with universal determinism is facing a choice between the following two positions: Either he has to deny that the physical aspect of our actions is completely determined by antecedent physical states and natural laws, or he has to deny freedom – label free action some sort of illusion.

Each of the two positions exists in many variants. In our century, indeterminism has sometimes been defended with arguments from microphysics (quantum theory). Physics is no longer wedded to the idea of universal determinism in the way it was in the eighteenth and nineteenth centuries. This is true – but the question whether indeterminism in physics is "ontic" or "epistemic" is still open to debate. If it is the latter, indeterminism in physics reflects limitations in our knowledge and is compatible with determinism in nature.

It is an old idea in philosophy that the freedom of our actions is an "epistemic illusion" due to our ignorance of their causes. This idea is related to one of Moore's suggested interpretations of "could have done otherwise" (cf. above, p. 9f.). Since, at least in many cases, we do not know what our choices (of course of action) are going to be, we say that it is possible that we are going to do a certain thing but also possible that we are going to omit the action. This corresponds to a common and natural use of "possible," roughly equivalent to the phrase "for all we know." A determinist who thinks that our choices (of course of action) are, in effect, determined, would then label the idea that man is "free" to choose his actions an epistemic illusion.

There is a classic defence of the compatibilist position which should be mentioned here. It enjoyed a certain popularity with writers on ethics of the former century. They thought that reasons and motives of actions are (comparable to physical) causes.[9] If every action "flows" from a motive, then actions are just as rigidly determined as events in nature. But then actions spring from

the agent's *self-determination* and not from external causal factors. Determinism must not be confused with fatalism.[10] Human freedom consists exactly in this, that human actions are determined by the agent's (own) reasons.

With the last statement we may agree. It is also true that motives and reasons are often called "causes of actions." There is no objection to this way of speaking as long as one does not let it obscure the conceptual differences between causes of events in nature and reasons for action. A minor objection to this position just described is that it is overly "rationalistic" if it assumes that *all* actions have a motive-explanation and that *no* action is therefore (completely) fortuitous.

This way of "reconciling" freedom and determinism is an interesting reflection of the prestige which the deterministic ideas have enjoyed in our intellectual culture. By calling reasons for actions "causes," one can defend human freedom and at the same time pay lip-service to the deterministic world-view of classical natural science.

This defence of compatibilism leaves another problem unsolved, however. One could call it the problem of *congruence* or *parallelism* (cf. above, p. 3). Granting that reasons are causes, we seem to have two parallel but independent causal chains here. On the one hand we have reasons causing actions, and on the other hand we have innervations and other neural processes causing muscular activity. The two chains converge in the physical aspect of the actions. How shall we understand the "congruence" or seeming "coincidence" that when I do a certain thing for one reason or other, the required physical aspect of my action makes its appearance under the influence of causes, perhaps acting from without my body, and in any case "external to my will"?

4.

I open a lock – my arms and hands go through certain movements. Why do I open the lock? I want to fetch something from the locked cupboard. By *moving* my hand I *achieve* the unlocking of the cupboard. The *movements* of my hands *caused* the lock to open. What *made me* move my hands in a certain way? The fact that I wanted to unlock the cupboard or, perhaps, the fact that I wanted to fetch something from the cupboard. What *made my hands* move in a certain way? Some innervations of the muscles from the brain. What made those innervations take place (just) then? With this question the "problem of congruence" is raised – and the conceptual muddle begins.

I shall next introduce the notion of the *context* of an action.

Consider again the action of opening a lock. It has a beginning: I "embark" on the task, as we say, proceed to action. The action has a certain duration, lasts for some time during which my arms and hands go through certain

movements. And it comes to an end: the lock opens. The thing just mentioned constitute (describe) the context of the action.

Where in relation to this context shall we "locate" the innervations of the muscles? Obviously they do not begin when my arms and hands are already moving. They must be there when I embark on the task. They must belong *in* the context of my action. Perhaps they could be called the "physical aspect" of that somewhat intangible episode which I call "embarking on" an action. What is this?

My desire to fetch something from the cupboard may have already existed before I set myself to open the lock. The same holds for my want to open the lock. The origination of a want may be impossible to locate exactly in time. If the want was there before I embarked on the action (and its existence thus falls partly outside the "context" of the action), then proceeding to action consisted just in this, that some innervations put my arms and hands in motion. Embarking on the action *was* my want "becoming active," and this happened when the innervations put my arms and hands in motion. But are not these two things: proceeding to action and the nervous impulses moving my hands really *the same*, only described in different ways? One description is in obscure "mentalistic" terms ("embarking on the action," "my want becoming active"), the other in, seemingly, clearer physical (neural) terms. I shall return to this question below.

5.

Assume that the only explanation I could offer for the action is that I wanted to open the lock. Just this. Not that I wanted to find out whether I could open it or that I wanted to fetch something from the cupboard. It would be rather strange, just wanting that. It would be like saying "an irresistible desire overcame me." One could ask: Was my action free? There is not much *point*, it seems, in calling the action free if its context is, in the sense described, "self-contained."

Assume, however, that my action has a fuller explanation. I opened the lock because I wanted to fetch a bottle of wine from the cupboard. Why did I want this? Perhaps I was expecting guests for dinner. When the action is placed in this setting it seems artificial to speak of a (separate) "want" to open the lock.

The fuller explanation points beyond the context of the action. It points to the future – to an end being aimed at. It also points to the past – to a pre-existing want conditioned by an expectation. When set in this perspective, one would not hesitate to call my action of opening the lock free. The context of the action is now embedded in a larger context of reasons and motivations.

This larger context is still finite in the sense that the chain of ever-remoter reasons has an end. I expected quests for dinner. This I obviously did for some

reason. The normal reason would be that I invited the people. But why? Perhaps because I had been invited to visit them before. By inviting them back I observe a rule of good manners in our society. And perhaps there are some other reasons too. But I shall probably not be able to advance in my explanation much beyond this point.

Although an explanation in terms of reasons may point far beyond the context of the action in time, the reasons must yet, all of them, be *present* in the context. The agent need not be aware ("thinking") of (all of) them when he proceeds to action. But they must be present in the sense that he subsequently can say, if challenged, that he *had them* then. He did not invent them afterwards, nor had he completely forgotten about them. He would have been able to state them when proceeding to action had he, for whatever reason, reflected on why he was doing what he was doing. But the borderline is often blurred between pre-existing reasons and a subsequent "rationalization" of an action.

When I set myself to act for some reasons, the motivation background present in the context of the action "activates me" – and the physical aspect of this activation is the innervations which make my muscles contract and relax and thus direct the bodily movements which constitute the physical aspect of my action. But how can the motivation background which moves *me*, the agent, to action have this power over the innervations which move my *muscles* if there is not something answering to this background on the physical side, *i.e.*, in the brain or the nervous system of the agent? The answer, presumably, is that the motivation background could *not* have this power unless it had some such "physical counterpart."

Assume that my action was the response to an order or was the answer to a question. I heard some noises which were an *order* to *me* to *do* a certain thing which I can do – and I proceeded to do it. The order was the reason why I acted. But the command had to be *understood* (not only heard) in order to activate me. What is this? In order to understand an order I have to know the language in which it is issued and to hear it when it is issued. I also have to know the meaning of orders as reasons for action. All this must already be "embedded" in my past history, if the order is to move me to action. This again presupposes, as far as we know – and this is a matter of *empirical* (scientific) and *not* conceptual knowledge – that my nervous system has been duly prepared or "programmed" (cf. above p. 27) in the course of my development, *i.e.*, growth and learning process. If, then, I receive an order and react to it, this means, in physical terms, that certain soundwaves affect my hearing nerves, and the "message" is transported to the brain and effects a change in the neural patterns which eventually "releases" the innervations.

But must not *the brain* "understand" the "message" of the soundwaves in order to emit to the muscles the "message" of the innervations? Certainly –

but it should be noted that speking of "understanding" and of "messages" is here metaphorical talk. Its literal meaning is this: In order to come to understand commands (in the literal sense of "understand") I have to learn a language and to react to orders and other messages (in the literal sense of "message") – and this process involves a (physical) impact on my nervous system. My brain becomes programmed to certain reactions to stimuli. This does not mean that the same stimulus will invariably call forth the same reaction. The programming is to a complex of stimuli, and variations in this complex may cause variations in the reactions (responses). On the level of mentalistic talk this answers to the fact that there may exist several reasons for and several reasons against an action and also reasons which, although present, are not efficacious in relation to the action which eventually results from a "balancing" of the reasons for and against.

The upshot of our discussion of the parallelism between the reasons (motivation background) of an action and the neural patterns causally responsible for its physical aspect is thus as follows: To the *understanding* of the reasons (as reasons for or against an action) there answers a programmation of the neural apparatus, and to the *existence* of the reasons in the context of a certain action there answers a stimulation of this apparatus, and to the agent's proceeding to action there answers innervations of some muscles in the agent's body.

Two questions now arise: Do these correspondences amount to identities? And: What is the bearing of these correspondences on the problem of freedom?

6.

I shall here introduce a technical term, *substrate*. And I shall say that the nervous processes under consideration are the substrate of the agent's setting himself to the action. Similarly, I shall call the muscular activity which constitutes the physical aspect of the action the substrate of the action. There is a reason why we cannot identify either the agent's setting himself to the action or the action itself with what I have called their "substrate." It is the following:

We could observe and accurately describe the muscular activity without knowing of which action it is the physical aspect. I see the agent's hands and arms go through certain movements manipulating a lock with a key. What is the agent doing? Unlocking the cupboard? This is one possibility. Or trying to see whether he can open it? (The trick may not be easy.) Or checking whether the key fits the lock? (There are many keys in the bunch, and the agent forgets from time to time which key matches which lock.) These are other possibilities. In order to know which of these actions the agent is performing, if any, we must know what he intended or "meant" by his behaviour. To find this out is usually not difficult. We do this by taking note of what preceded or what

followed the perfomance or simply by asking the agent. But observations, however accurate, on his muscular activity alone cannot give us the answer (at most they may give rise to a surmise), because the substrate of an action does not stand in a one-to-one relation of correspondence to the action. And the same also holds good, of course, for the relation between the innervations and the agent's embarking on the action. Even if the innervations could be identified and described with great accuracy, *they* would not tell us which action the agent engages in.

But are not the muscular activity and the action, after all, the same reality, two different conceptualizations of what is here called "the substrate"? And the same with the innervations and the embarking on the action?

In some sense of "reality" they are the same. I shall call this their *robust* reality. The action is not anything over and above its physical aspect, if by "over and above" one understands some thing or some event in the physical world which one could identify as that which, when "added" to the muscular activity "makes up" the (whole) action. There is no such thing. And similarly for the innervations and their "equivalent" in actionistic terms.

So must we not say then that the action *is* identical with its physical aspect (muscular activity) and the agent's embarking on it identical with the innervations, *i.e.*, with the neural cause of the muscular activity? The answer is No – for the reason already given, *viz.*, that no description of the substrate would be sufficient to identify the action.

7.

What causes the innervations to occur? Roughly speaking: Stimulations of a nervous system which has been "programmed" in the course of the lifetime of an individual (the agent) to respond in characteristic ways to stimuli of the kind under consideration. All this can, and should, be understood in strictly "physicalistic" terms – as soundwaves affecting the auditory nerves, neurons firing, "engrammes" being implanted in the connections of nerve-fibres, *etc.*, *etc.* The response is, in the last resort, the nervous impulses which steer the muscles.

This is a sketchy description of what I propose to call the "substrate" of the motivation background present with an agent in the context of an action.

The overt effect of the reasons in moving the agent to action thus is the same as the overt effect of a physical stimulation of a "programmed" neural system, because either effect consists in that the agent's bodily organs go through certain movements. Does it follow that the reasons are identical with the physical stimuli? The answer is analogous to the answer we gave in order to clarify the distinction between action and muscular activity.

How does one establish that an agent has a certain reason for action, *e.g.*, understands a command, believes that something is a means to an end, wants something and shuns something else? Partly by taking note of what he professes to understand, believe, want, *etc.*, that is, by eliciting from him verbal responses to questions. But these are by no means the sole criteria – just as the reason the agent himself gives for an action need not settle the question why he acted. Further investigations about his past history or subsequent behaviour may be called for, and the results of such investigations may override the verbal testimony of the agent. ("He cannot really believe what he says; he is too well educated for that, and his behaviour on other occasions speaks strongly against this.")

The existence of a reason is not anything which can be pinned down to the obtaining of a state of affairs or the going on of a process at a certain time and place. It is a "global" fact of non-definite extension, a characteristic of the type of logical individual we call a "person."

The observations on behaviour (including verbal responses) on the basis of which we attribute to an agent a certain reason for action do not logically entail the existence of the reason. But they are not (only) signs or symptoms of something the existence of which could be established independently with "absolute certainty" on the basis of some defining characteristics *other* than those behavioural manifestations. This is why I shall call these latter "criteria" of (the existence of) the reasons.[11]

Neural states and processes do not, on the whole, serve as criteria of (the existence of) reasons. Perhaps they would be criteria *among others* if they were more manifest and accessible to inspection and better known than they are at present. But as things are, their epistemological position in relation to reasons is quite different. Suppose we had found out, by anatomic and physiological study of the nervous system, that in many cases there is a correlation between some kind of simple reason (*e.g.*, being thirsty) for some simple types of action (*e.g.*, drinking) and certain neural patterns and processes. We could then frame a *hypothesis* to the effect that this correlation holds also in unexamined cases, if not "without exception," at least with "high probability." This hypothesis could then be tested on further cases. Testing it – like making it – presupposes that we have already established on independent grounds the existence of the reason for the action which is now being "matched" with a corresponding neural state. If the correspondence is well established, the neural state in question may be regarded as a reliable *sign* or *symptom* of the existence of the reason. As long as the correlation remains a scientific hypothesis, the neural state fulfils this rôle of a symptom. Only in the very unlikely case that the hypothesis became so well confirmed that we would be extremely reluctant to drop it when faced with seemingly contrary evidence could we conceivably use the neural state as a *criterion* of the agent's having a certain reason for action.

And even then the criterion would only be one among many, and its usefulness in attributing to agents reasons for their actions would depend upon how well it contributed to our understanding of the agent as a person *and* to the agent's understanding of himself.

The above should suffice to make it clear why the *identification* of the existence of a reason with a correlated neural state is out of the question. And also that this is fully compatible with identifying the impact of the motivation background on the agent with the causing of the innervations which are responsible for the external aspect of the action.

About the nature of the causal mechanism not too much is known at present. More may be known in future. It cannot be regarded as certain that the correlation between a motivation background and its substrate is one-to-one in the sense that the presence of the same reasons will answer to the same neural states and processes causing the muscular activity in each context of the same action unless – which is always possible – one *postulates* the sameness and ascribes the difficulties in establishing it empirically to the play of (so far) unknown or unobserved factors.[12]

8.

I hope I have succeeded in showing why it is no accident that when the reasons move the agent, the causes of muscular activity move his body correspondingly. The idea of something accidental calling for an explanation is produced in us by the misleading picture of two parallel chains of independent and yet (in time) co-ordinated elements, *viz.*, one chain of reasons and another one of causes, both chains converging in the action. From the point of view of their "substrate," *i.e.*, their robust, spatio-temporal reality, there is only *one* "chain."

If man from birth were endowed with a brain and a nervous system functioning in accordance with strict causal laws, and if this system never changed in the course of the development of the individual, then it would indeed be something of a "mystery" how neurological causes could produce somatic effects (movements of a body) in "congruence" with our actions. But this idea of the brain as a system is not correct. When an infant grows up to be a member of a society, learns to speak and do various things, to understand the meaning of challenges and institutions, and to participate in various practices, its nervous apparatus undergoes a simultaneous development partly of learning under the influence of external stimuli and partly of maturation of inborn capacities. The two processes go hand in hand and therefore the congruence between the mental and the bodily aspects of action is a *harmony* established in the course of the individual's life and necessary for its preservation over the span of time allotted to each of us.

That the solution we have given to the problem of congruence is not "materialistic" should be obvious. Less obvious is perhaps that it also involves no commitment to determinism.

Muscular activity is caused by stimulation of a "programmed" nervous system. Might not the stimulation in its turn be caused by events anterior to the context of the action, anterior even to the life-span (existence) of the agent, operating perhaps "from the dawn of creation"? So that then, by transitivity, the physical aspect of an action would be predetermined, in some cases at least, long before the action took place.

We have little reason to believe in such "rigid determinism" – and it is not even certain that it can be given a clear meaning. But let us not now question its possibility nor even its truth. Would this affect our view of the freedom of our actions?

9.

Suppose that the action is one which we cannot connect in the understanding with any particular reason for doing it. We did it "for no particular reason." We cannot account for such fortutious or gratituous actions – except possibly by looking for causes of the movements which constitute their physical aspect. If we can find a cause, we should presumably say that the action was not "free." We would treat it as a reflex rather than an action. If we cannot find a cause we should not know whether to call it "free" or not. Fortuitous actions, as we have observed before, have a peculiar relation to freedom just because they lack that which is the hallmark of free action, *viz.*, to have been performed for some reason(s).

In order to have a clash or conflict between freedom and determinism we must imagine a case when there is both a reason-explanation and a causal explanation at hand which both are, somehow, of "the same thing." To imagine this, *i.e.*, to describe correctly a case of conflict is not at all easy. As we shall see, it may not even be possible.

It is important here to see clearly the different nature of causal explanations and reason-explanations. A reason-explanation is of an action, a causal explanation of the physical (somatic) aspect of an action. A given display of muscular activity does not show by itself of which action it is the somatic aspect. Only in the case of some very simple actions such as, for example, the raising of an arm, may it seem pointless to separate the action from its physical aspect, for example the rising of an arm. What the causal (neurological) explanation can explain is the rising of the arm – and if the action performed was (just) the raising of the arm, one is tempted to say that one has a causal explanation of the action too. If, moreover this action has no other explanation, was performed as we say for "no particular reason," then the causal explana-

tion of its physical aspect is the sole explanation relating to this action which we have – and then, as we know, we may even be in doubt whether to call it an action at all. If, however, the action was, say, that I was reaching out to fetch a book from a shelf, the situation is different. There is no causal explanation of why I reach out for a book, although there may exist a causal explanation of why my arm reached, or failed to reach, the book I wanted (or had) to fetch. (This simple example should make us aware of the danger of using very "primitive" examples when discussing action. Arm-raising is one of the most favoured ones – but it is a *poor* example of an *action*.)

Since causal explanations and reason-explanations have different explananda there can be no "conflict" between the two types of explanation as such. But this does not yet show that there might not be a "conflict" between a reason-explanation of an *action* and a causal explanation of its *physical aspect*.

Assume next that we have these two explanations relating to the same action and assume further that the one makes reference to reasons which are present for the agent in the context of the action and the second to stimulations of the nervous system of the agent in that same context. Then there is no "conflict." *In the context of the action* there simply cannot *be* any "conflict" between the two explanations. On the contrary: we who share the "belief in science" of our century regard it as probable or even certain that if the action has a reason-explanation its somatic aspect has a causal explanation.

In order to give a causal explanation at all, it must have been established – using appropriate experimental techniques – that a certain stimulation of the nervous system *outside the context of any action* results in a certain type of muscular activity. (One should thus be able to simulate the somatic aspect also when no action of which it might be the somatic aspect takes place.)

For there to be a conflict between the two types of explanation we must now imagine a situation in which a certain action is performed and it is known that prior to the context of this action the agent's nervous system had been stimulated in a way which is bound by "causal necessity" to produce the somatic aspect of that same action. ("He had been secretly given an injection.") We must also imagine that the muscular activity occurs exactly when the agent performs the action. If it occurs before, the agent might say something like this: "Strange, I was just going to fetch a book from the shelf when my arm suddenly went up 'of itself' to the desired position." If it occurs again later, he might say: "Strange, my arm did not rise at once when I was going to fetch the book, I had to wait a second."

If the agent himself knew of the operation of the cause he would also anticipate the display of muscular activity consequent upon it. ("Two minutes after the injection my arm will rise.") When the activity occurs he might use the opportunity for doing something for which those movements are required. "When my arm rose, I snatched the book from the shelf." The snatching is then

an action with a physical aspect of its own, *e.g.*, closing my fingers around the book; the rising of the arm was just something which happened to me and "facilitated" the action. But it is also possible, and perhaps more likely, that the agent, knowing what is going to happen to his body, will do nothing at all then.

Assume, however, that the agent does *not* know of the operation of the cause but that *we* know. The agent said that he did something for a certain reason, and we say that the physical aspect of his action would have occured even if he had not acted. Was his action free? Since he had a reason for his action it was what we call "free action." But suppose that we did not only know of the operation of the cause, but that we had ourselves made it operative? ("We gave him an injection.") Shall we say then that the agent had been "manipulated"? This would not be right. His *body* had been manipulated. But since he happened to have reasons for doing an action the physical aspect of which consisted in the muscular activity which we had caused to happen, his *action* was not a result of manipulation. Only by influencing an agent's *reasons* can he be (genuinely) manipulated.

The sort of case we have been imagining is artificial and plays at most a marginal rôle in an agent's life. But more importantly: we have not succeeded yet in staging a genuine case of conflict between freedom and determinism. Have we set ourselves an impossible task then? Let us make this final attempt:

Within the context of the action, could not the cause of the somatic aspect of the action in its turn have a cause operating from outside this context and thus, by transitivity, be itself the cause of the somatic aspect? Such an anterior cause would be a stimulus affecting the (programmed) neural state of an agent either from outside, say in the form of soundwaves, or from the inside, say in the form of cramps in the stomach. In the medium of the understanding these affectations may appear as reasons for actions (to obey an order or to get something to eat) and in the medium of the nervous system they may release nervous impulses guiding the somatic aspects of "corresponding" actions. Whether they will have this effect or not depends upon how the agent and his nervous system have been "programmed": the agent in the form of learning and previous experience, his soma in the form of traces which learning and experience have left on it. Is this a "conflict between freedom and determinism"? I don't see how it could be called this. But the influences (stimulations) to which a person has been exposed in the course of his development (education and life experience) and is currently exposed to in his social and physical situation determine to a great extent the reasons which he will have for actions and thereby also what he will do. This is a fact to which we have to acquiesce. It does not make a man unfree in the sense that he would not be acting for reasons. But it makes any man to some extent a "victim" of the circumstances of his life and sometims also a victim of (genuine) manipulation

by other agents. The circumstances of a man's life, and therewith the reasons he has for various actions, are *also*, however, to some extent his own making.

10.

Is every total somatic state rigidly determined causally by preceding somatic states? The answer is negative, since the somatic states are also causally dependent upon stimuli from outside the body. So the question is whether every total somatic state of the body is causally uniquely determined by preceding states and external stimuli. But even with this obvious supplementation the *meaning* of the question is obscure.

What does the phrase "causally uniquely determined" mean? An answer could be: It means that knowing the stimuli and the preceding states one could predict ("with certainty") the next state. But what is "the next" state? Do the successive total states form a discrete manifold then? And does a state depend causally only on the immediately preceding state, or also on patterns in the succession of (several) preceding states? We shall not even *try* to answer these questions. (Raising them will, however, give an idea of the conceptual obscurity surrounding our initial question.)

Predicting future states of the body on the basis of knowledge of stimuli and past states also presupposes knowledge of connecting *laws*. Such laws would, in the last resort, be generalizations from experience, *i.e.*, from experiments and observations. Let us not question the possibility of knowledge of such laws.

In order to complete the deterministic picture we are drawing we have also to assume that all the stimulations which affect a body have a causal history which is strictly deterministic. We are thus forced to consider not only the total state of a body but much larger fragments of the total state of "the world" – and maybe not only fragments but the unbounded totality. In the end we may have to draw something like the suggestive picture of rigid determinism which Laplace impressed upon the scientific and philosophic imagination in an immortal passage in his *Essai philosophique sur les probabilités*.

But have not scientific developments in our century eroded and made obsolete the idea of rigid determinism in the physical world? At least at the microlevel there seem to exist "margins of indeterminacy" within which bodies can behave (move) freely. Neural states and processes are studied at the microlevel. One talks about "spontaneous activity" in the neural system. And some philosophers have hailed these developments in science as loopholes for "free will."

I hope that I have succeeded in showing that such pro's and con's of determinism are completely irrelevant to the philosophic problem of free action. Even the most rigid determinism in the physical world, which we could

conceive as a logically consistent possibility, would not show that human beings are not free agents or that "free will" is only an epistemic illusion.

Determinism holds good, one could say, to the extent that it works, *i.e.*, we can succesfully predict the future on the basis of past experience and hypothetically assumed laws of nature. Our success in this regard has been considerable. The search for causes and deterministic explanations has turned out to be immensely rewarding. Therefore it has been useful to entertain the idea of determinism as a heuristic maxim for guiding research. In many areas of science the idea is likely to continue to play its classical rôle. In other areas it may have to be modified ("relaxed") or it will be dropped as useless.

11.

Have I wanted to say that study of somatic states and processes is of no relevance to an account of actions in the terms of reasons? By no means have I wanted to say this.

Several of the basic "passions of the soul" have characteristic somatic accompaniments – other than the overt bodily expressions known of old to observers of human nature. This is true, for example, of anger and fear. They are "reflected" in measurable fluctuations in blood pressure or secretion of adrenalin. Observations of such changes may on occasion be relevant also to our understanding (explanation) of actions.

They might, for example, be used as a kind of "lie detector." An agent perhaps denies that he did something because he was afraid of something he wished to escape or because he was angry with somebody and wanted to harm him. He may give an entirely different reason for his action. We doubt what he says – and a medical examination gives support to our suspicion.

Perhaps we can "force" the agent to admit that he was lying, hiding from *us* his real motives. But perhaps he had used a "noble" motive to hide an "ignoble" one not only from us but also from himself. He was "lying to himself" too (cf. above, p. 22). What can our "lie detector" now achieve, if the agent himself was not even aware of fear or anger? Great caution is needed when trying to decide such cases. Perhaps the wise thing is to suspend judgement. But maybe we can make the agent realize that there was something in the situation that he actually feared or that actually had angered him – although he says he did not "feel" fear or anger then. This may make him more watchful (reflective) of his subsequent conduct. In this way he may arrive at a changed self-understanding in the light of which he will also view some of his past actions differently.

12.

Do animals act?

We do not easily say that they do. To say that an animal "performed" this or that action – or omitted to perform one – even sounds a bit comical or ludicrous. It sounds like a "personification" of the animal – such as is common in fables and tales. But animals, "really," are not persons. (Some, however, can be "characters" or even "personalities.")

Animals, of course, *do* a lot of things. But this holds also of many inanimate objects; our language is permeated by "actionistic" ways of talking about things that ("passively") take place.

Yet animal behaviour also has many features in common with human action. Animals *learn* to do various things – which they then do on appropriate occasions. When thirsty they exhibit "water-seeking behaviour," when hungry they "go for food," to use the jargon of psychologists. How like or unlike human hunger and thirst is animal hunger and thirst? This is a philosophically interesting question – but I shall not go into this topic here.

Aiming, intending, can certainly be attributed to animals. Whether we should say that animals "have" aims and intentions is less certain. Animals make choices. They may, perhaps, even be "torn between alternatives," like Buridan's famous ass.

Animals are free when they are not (physically) prevented or restrained from doing what otherwise they would do. But are their doings free in the sense human actions are? In what sense then are human actions free? Free action is action for reasons, I have said. (And action, essentially, is behaviour for reasons; the adjective "free" in "free actions" is redundant except when it means absence of "compelling reasons.") That animals do not act is connected with the fact that they do not possess the self-reflective capacity which "having reasons for actions" is. And this again is connected with the limited linguistic capacities of animals.

Since animals do not act for reasons, *why* do they behave as they do? Descartes thought that animals were machines, automata. If this means that animal action, to the extent that it can be explained at all, must be explained as *reactions to* (inner and outer) *stimuli*, I think that Descartes was right. The other type of explanation of behaviour, *viz.*, in terms of reasons for action, simply does not *apply* to animals.

Human behaviour too – including the physical aspects of actions – may be studied as reactions to (inner and outer) stimuli. Man is no less a machine than animals are. Rather one should say that he is "more" of a machine because his machinery is complex, more developed. It is not by being exempted from the boundage of natural law that man is a free agent. He is this because we can understand him in a way, *viz.*, as a person, in which we – or most of us at least – cannot understand the rest of creation.

NOTES

[1] "A hundred misleading pictures come together here and *this* makes for the difficulty of the philosophical situations. Wherever we put our feet, the ground yields. The 'great', difficult problems of philosophy are this not because of the existence of some extremely subtle or mysterious state of affairs which we have to ascertain, but because in this place a great number of misleading forms of expression are crossing each other." From an unpublished work by Wittgenstein called *Bemerkungen II.*

[2] In Stoutland (1982).

[3] Moore (1912, 131).

[4] *Ibid.*, p. 134.

[5] *Ibid.*, p. 135.

[6] For how my opinions on this question have changed, cf. von Wright (1963, 190) and the papers (1976) and (1981) reprinted in *Philosophical Papers I, Practical Reason.*

[7] von Wright (1963a, 39f).

[8] *Ibid.*

[9] Schopenhauer's treatise on the Freedom of the Will (1841), still very much worth reading, may be regarded as the *locus classicus* for this position. Motives, in Schopenhauer's view, are causes and, as such, necessarily connected with the ensuing actions. Motivational causation he characterizes, interestingly, as "die durch das *Erkennen* hindurchgehende Kausalität." Schopenhauer quotes with approval Hume, who held "that the conjunction between motives and voluntary actions is as regular and uniform as that between cause and effect in any part of nature" (*Enquiry*, § VIII). A later writer in the same vein is Westermarck (1906–1908).

[10] Cf. Westermarck (1906–1908, vol. I, ch. XIII) for a good clarification of the distinction.

[11] The distinction between criteria and symptoms is familiar to every student of the later Wittgenstein. There is a vast literature commenting on the distinction, and many different interpretations have been offered of what Wittgenstein understood by the two terms. We need not add to the exegesis here.

[12] Cf. Wittgenstein (1967, § 608).

REFERENCES

Moore, G. E. (1912), *Ethics*. London, Oxford University Press.

Stoutland, F. (1982), "Philosophy of Action: Davidson, von Wright, and the Debate over Causation," in G. Fløistad (ed.), *Contemporary Philosophy: A New Survey*, vol. 3. The Hague, Martinus Nijhoff, pp. 45–72.

Westermarck, E. (1906–1908), *The Origin and Development of the Moral Ideas*, I–II. London, Macmillan.

Wittgenstein, L. (1967), *Zettel*. Oxford, Basil Blackwell.

Wittgenstein, L. *Bemerkungen II.* Unpublished.

von Wright, G. H. (1963), *The Varieties of Goodness*. London, Routledge & Kegan Paul.

von Wright, G. H. (1963a), *Norm and Action*. London, Routledge & Kegan Paul.

von Wright, G. H. (1976), "Determinism and the Study of Man". Reprinted in von Wright (1983).

von Wright, G. H. (1981), "Explanation and Understanding of Action". Reprinted in von Wright (1983).

von Wright, G. H. (1983), *Philosophical Papers I, Practical Reason*. Oxford, Basil Blackwell.

SENSATIONS AND PERCEPTIONS

I

1.

"I see a bird over there". If this is true, there *is* a bird over there. If not true I, at most, *seemed* to see or *thought* I saw a bird over there. In both cases I had a sensation, I sensed something, but only in the first case did I have a *perception*, did I perceive something. That is: this is both a useful and a common way of distinguishing between "sensation" and "perception".

Sensation language is secondary to perception language. I cannot say that I seem to see a bird unless I can identify something as a bird. Physical language is prior to perception language. Describing one's perceptions presupposes that one can describe the percepts. The percepts are events or objects in the physical world (but not necessarily "material objects", however.)

If I seemed to see a bird over there, then, in most cases, there was something to be seen over there. By looking closer I may discover what it was. Perhaps it was a flower, partly hidden in the grass. But did I really *see it*? Was the flower the object of a perception? If not, *what* did I see then? Some coloured patches in the grass? This is not necessarily a better answer than saying that I saw a flower (if there was a flower to be seen).

Perhaps upon reflection (I am not now considering the case of "looking closer") I can *redescribe* the object of my perception. For example, say that I saw something red and yellow over there. If this is how I express myself upon reflection or when being challenged, and if there really is something red and yellow over there to be seen (a flower, say) then it may be true that I *saw* something red and yellow over there. (I say "may be" because of the possibility that I was not truthful but just made it up.)

Suppose, however, that I can, when challenged, only say sincerely that I saw "something". I cannot say that it had this or that colour or otherwise how it looked. This is a somewhat strange situation and it is not clear what else we can say about it. One possibility is this: since there was something to be seen, *viz.* a flower, I evidently "saw something" – and that is all we can say. Another possibility is to say that, in fact, I *saw* nothing (did not see anything). But I obviously had a *visual impression* (sensation) "from over there". It was to it I gave expression when I said, mistakenly, that I saw a bird. One can

G. H. von Wright, In the Shadow of Descartes. Essays in the Philosophy of Mind, 45–68.

therefore call it "an impression of a bird". This impression was what I *had*, not what I *saw*.

There is a temptation to speak of "objects of sensation". Perhaps what philosophers called "sense-data" were such objects. Some philosophers would presumably have said that a coloured patch is a sense-datum – but presumably that a flower or a bird is not. But it seems to me that if one says the first, one can just as well say the second.

"Sense-data" is an auxiliary construction the aim of which is to describe my sensations, or their "content", that which is given to me in sensing.

Consider again my seeming to see a bird. Someone tells me there is no bird. I take a "second look". Perhaps I advance closer to the thing. Then I see a flower where I thought I had seen a bird. This is a (veridical) perception; there really *is* a flower.

Without visual impressions there is no seeing either. Perceptions presuppose sensations, "sense-data". But the reporting of sensations is often in terms of perceptual objects ("an impression of a bird.") Just as physical language is prior to perception language, perception language is prior to sensation language.

2.

But is it really so? Cannot a sensation be described in purely sensational ("phenomenal") terms? My visual field at a given moment as an arrangement of coloured patches, for example?

What *is* the visual field? It is that part of the (physical) world of which, at a given time, I have an impression. If this is the answer, I can see various things – trees, birds, furniture, as well as coloured patches – in my visual field. These "seeings" are perceptions. And I can also seem to see things in my visual field.

"Visual field" is a technical term and I could also use it for the *impression* I have of a part of the world. Under this terminology I have various impressions in my visual field, for example the impression of a bird. The visual field can then be said to contain, or to consist of, a number of impressions. The impressions are its parts. The description of these parts is in terms of the things, birds, trees, *etc.*, in my visual field in that *other* sense of the term "visual field", meaning a part of the world.

If someone were to tell me that it is "certainly possible" to describe one's visual field as a kind of "rag carpet" of coloured specks in a certain arrangement I should ask him to try. And if he tried and thought he succeeded, the next question would be whether, or in what sense, the terms of the description are purely sensational.

3.

Since I can say in retrospect that I thought I saw a bird where in fact there was a flower, does it mean that a perception is a sensation *plus* a thought? Saying this could be seriously misleading. But one could say, with caution, that a perception has two aspects, a "sensational" and a "cognitive" one.

4.

I saw a bird over there and someone asks what colour was its head. If I answer "red" and this is right, I presumably *saw* the colour. I *might*, however, also have guessed the answer and been right; or I might have known that birds of that kind have red heads and based my answer on this knowledge and not on what I actually saw.

Suppose, however, I answer that I did not see what colour the bird's head was. Did I not see its head then but only some other part of the animal? Perhaps; seeing a thing does not entail seeing every part of it; in the case of opaque three-dimensional objects this is even impossible. But perhaps I *saw* the bird's head without noticing what colour it was. This surely can happen. Someone might then say: You *must* have seen the colour, only you did not *notice* it. Light-rays of that colour entered your eye. Yes – but still it would be wrong to say that I *saw* which colour the bird's head was. It may happen that upon reflection, or all of a sudden, I realize that its head was red. Perhaps this phenomenon has an explanation which depends on the fact that such and such light entered my eye. But that would not mean that I *saw* the colour *then*.

Suppose, finally, that I answer "green". Since this is wrong I cannot say in retrospect that I saw that the bird's head was green. Because that would be true only if the bird's head had been that colour. Suppose I say that the bird's head *seemed to me* green. But can this be *true*? Can something red seem to me to be green? Certainly. The light which fell on the object and was reflected could have been such that something red appeared green. Or there could be some anomaly with my seeing; I had perhaps been staring at something red just before I saw the bird and was in consequence having a green afterimage. If I am red-green colour blind and cannot clearly distinguish the two colours, the case may be counted as a case of seeming. But if my colour blindness is complete one would have to say that I made a mistaken guess or that my memory failed me. There was then a "false cognition".

But what if no such explanation is forthcoming? Then the possibility is not that the red object seemed green to me, but that I did not use the colourword correctly. It could have been a "slip of the tongue". Or my native language is not English and I was confused about the meaning of the English words "red" and "green". Further tests will usually soon settle the question whether I made

a mistake in language or whether something red actually had seemed to me green.

Something which is red may seem green to somebody and blue to somebody else. But something red cannot seem green to somebody and seem red to somebody else. Something may seem different from what it is but not like what it is.

We can imagine a situation in which things very rapidly changed their colour so that no two persons observing them in succession would ever agree in their judgements of the colour of the thing. But if this were the case with all things in the world, subjects could not have learnt to use colour words (as we know them). In such a world there would be no such thing as objects having colour and therefore no such thing as objects changing colour either. Therefore, the situation we "imagine" to be the case with *all* things is not an imaginable situation.

Animals do not use colour words, and have not colour concepts. But some animals can discriminate between objects on the basis of their colour, that is, discriminate between colours *roughly* in the way we do. This fact may be given a "scientific explanation" by pointing to similarities between them and us in our visual apparatus and to facts concerning physical optics and the nature of light. And thus one can say that the fact *about us* that we have colour words and attribute colours to objects depends upon facts and regularities in nature which the study of anatomy, physiology, and optics reveals to us.

5.

I sit in front of the open sea and say: "I see a lighthouse there, far out, sticking up above the horizon". But after a while I see that it is a ship approaching the shore. So I have to correct my previous judgement and say: "I seemed to see a lighthouse there, far out".

What I saw was a ship. I saw the ship but I did not see (realize) *that* it was a ship. (I thought it was a lighthouse.) The first was an impression, the latter is a cognition. Furthermore, a cognition "based on" an impression is a perception.

I see something there, far out, sticking up above the horizon. I say: "It *looks to me* like a lighthouse". "It" means "That which I see", the object of my perception. But it turns out to be a ship approaching the shore. What *did* I see? A lighthouse? No, because I did not claim to see a lighthouse and there was no lighthouse. A ship? No, because I did not claim to see a ship although there was one. A speck on the horizon? Can a speck be a ship, turn out to be a ship? *What I see*, not recognizing the object, can turn out to be a ship. So, once again, what *did* I see? Something? An object? A speck? Calling it a "speck"

may not say anything more than calling it "just something". And then the speck can truly turn out to be a ship.

I say: "I see a speck far away on the horizon". I do not know what it *is* that I see – it does not look like a ship nor as anything else that might *be* there. But I said I saw a speck. So must there then not have been a speck there, far away? It turned out to be a ship. So the speck was a ship, or how? This is what we say and it is intelligible and can be true. So there is a sense in which it can be true that a speck is a ship, and another sense in which this is either false or nonsense. This we must accept. This is how our language functions when we try to say what we sense or perceive.

Could it be that I *seemed to see* a speck on the horizon? Yes. There was a speck on my spectacles which obscured my field of vision. Or a straw of grass sticking up just a few yards in front of me. Then I *saw* a speck, there was a speck to be seen and it was the object of my perception – but I did *not* see a speck on the horizon because there was no such thing to be seen "out there".

Another case. I see a speck out there on the open sea. Perhaps a vessel passing by has spilled oil? I set out to check and find that there is no speck (of oil or anything else) on the water. So, it only *seemed* to me that I saw a speck on the water. What was it then that *looked* to me like a speck on the water? Perhaps the shadow of a passing cloud. Did I *see* the shadow? At least it would not be right to say that I *saw that* there was a shadow on the water. I had a visual impression (sensation) of a speck on the water. But is not a shadow too a speck? That depends upon how we understand "speck". If I meant a film actually on the water, then a shadow does not qualify as a speck, whereas a film of spilled oil does. But if a shadow too qualifies as a speck, then I *saw* a speck on the water, saw *that* there was a speck on the water.

6.

The observations in the last section should make us aware how vacillating perception language is. This is something which must be respected. Introducing greater precision in the way philosophers talk about perceptions and sensations can lead to conceptual distortion. Particularly dangerous is the invention of technical terminology leading to a multiplication of entities. Sense datum terminology is an example. It easily leads one to talk and thereby also to think about the created entities as *analogous* to things with which we are already familiar from our perceptual life, – for example to ask whether sense-data are there ("exist") even when nobody is "sensing them". But one can also make an innocuous and, possibly, even conceptually clarifying use of the notion. The same can be said of talk about the "content" of impressions or about what is *given* to us in sensing or about the distinction between objects of perception and physical objects.

II

1.

Having a sensation is not like having an "object" – a hat, say. Perhaps the best answer to the question what having a sensation *is* is to say that it is a *state* in which a sentient being is. This is not to say that the sensation itself is a state.

The subject can attend to, observe, register, report this state – for example a state of being in pain or of seeing a bird over there. Two subjects may compare their sensations and find that they agree or are different. There is nothing absurd in saying that the pain I feel in my back is exactly like the pain you feel in your back, that they are *the same*. This being so, there is a perfectly good sense in which *I* may feel or have *your* pain, and *vice versa*. – This is not in conflict with the "privacy of sensations".

To describe a sensation is to report a state – not to describe an object. Therefore one should be cautious speaking of the "description" of sensations. That way of talking suggests a view of sensations as a kind of "objects of perception". This is a confusion.

Reporting a sensation which one has can be called making *expressive* use of language. "Expressive" is then contrasted with "descriptive". Reports on sensations are also sometimes called *avowals*.

Sensations, though never objects of perception, can be objects of *recollection*. Reporting on one's sensations from memory is not expressive use of language in the same sense as reporting sensations when one has them. But nor is it description in the sense of describing, say, objects of perception. I shall call reports on sensations from memory "echo-avowals". When two persons compare their sensations this is usually on the basis of such "echoes" of how they "feel" when they have the sensation in question.

2.

I have injured my foot and feel pain. You have had a similar injury. We compare our respective pain sensations. Perhaps I say "This is exactly how I feel". But perhaps there is a slight itch in my pain but not in yours. Then the pain I have now is not the same (kind of) pain as the one you experienced then.

"But this is only 'indirect' comparison" someone might say. In order to tell whether my pain "really" is the same as yours one must have access to both "directly". What would that be?

One can imagine that my nervous system became "extended" so that *I* could *sense* pain caused by an injury in *your* foot. Then I might say: "The pain in your foot feels exactly like the pain in my own foot". But this is no guarantee that our respective pain sensations are similar (identical in kind). Because I do not yet know how *you* feel the pain in your foot. Perhaps you feel it quite

differently. In order to find out we should again have to rely on our respective avowals. And then decide whether our pains are the same, or different.

But what if the extended connection from my nervous system were, not to your foot, but to the "pain centre" in your brain? Could I then not literally feel *your* pain? No, because there would still be no guarantee that the "pain-message" from your foot when thus transmitted from your brain and felt by me is the same sensation as the pain felt by you.

So none of these imaginings give us what we want: a "direct" comparability of our sensations. Your pain is yours, even if *I* feel pain in *your* body. And the same is true of *my* pain. This is what "the privacy of sensation" comes to: they are sensations of *different persons*. In order for me to have your sensations I must be you!

Could one make this point also by saying that my pain and your pain although they may be the same in kind (generically identical) are yet "numerically different"?

This would not be very illuminating. It conjures up the picture of the two pains as two objects. This is a mystifying and useless picture. Already the plural "pains" is suspect when it does not refer to two *kinds* of pain. (Headache and toothache are two different pains.)

"Two individual pains" means the pain of two individuals. And their pain may be the same or different in kind. If different, one can say they have different pains.

One can say that the factor which individuates sensations is the persons or sentient beings who have them.

Assume that two persons always felt the same pain at the very same time. Then one could say that they share a sensation, that their sensations are numerically the same (one). This is an innocuous *façon de parler* as long as one does not build a mythology round it. One could also call this a "sympathy of souls". There *is* such a thing.

3.

The comparability of sensations which we discussed presupposes that the individuals concerned have a language in common so that they can communicate and understand each other's "avowals". If one masters a much more refined vocabulary for sensations than the other, comparison may be difficult. (One of them is perhaps an artist or a psychologist, or is much more given to "introspection" than the other.)

Animals, at least those of the higher species, are sentient beings. But they do not speak to one another about their sensations. It follows that the "comparison of sensations" of the kind we discussed is not possible between animals or between humans and animals. But this does not mean that comparison of every

kind is excluded. Beings, including humans, communicate their sensations also in other forms of expressive behaviour than the expressive use of language. We can tell that a dog suffers pain. We can usually tell whether its pain is mild or grave, perhaps tell that it is a "stabbing" kind of pain, – and we see or conjecture the affliction which causes the pain. Also in the case of humans behavioural "avowals" other than linguistic ones (moaning, contortions resulting from pain) are part of the evidence on which we attribute sensations of a certain kind and quality to other individuals. We also rely on non-linguistic evidence when we are not sure whether a verbal avowal (of pain, say) is genuine or feigned.

4.

I see a cubical box in front of me. Unless it is of transparent material I do not see all its sides nor its interior. So what do I "really" see? In order to answer the question we construe a new perceptual object which consists of those and only those parts of the box of which I have a visual impression, *i.e.* of which I can say, in reply to a question, that I see *them*. This object could be the top and two sides of the cube.

When I said "I see the box", was this overhasty, or an "exaggeration"? I would say No. I saw the box. *It* was my object of perception. The *whole* box? I could answer Yes – meaning that I saw *the box*. But I can also answer No – meaning that I did not see every *part* of the box. The meaning of the question "Did you see the whole box?" is not clear by itself. It could mean, for example, "Did you notice that there is a hole in one of its sides?". Or it could mean: "Which sides of the box did you actually see?". Depending upon my position in relation to the box and also on the size of the box, the answer may vary.

The box ("the whole box") can be what I see, an object of perception. A part of the surface of the box can also be my object of perception. Can they both be it at the same time? Again we must ask for the meaning of the question. One might introduce a notion of "attending to what one sees" and then, perhaps, say that one cannot "attend" to more than one object of perception at a time. Saying this may have a point – or it may be pointless.

Asking what I "really" see when I see the box is a challenge to me to attend to my *sensation*, to the sense-datum or -data I have of the box. My description of these data, however, refers to parts of the physical object, the box.

5.

Perceptual objects exist in what we call "the external world". They are what I propose to call *physical phenomena*.

Some objects of perception are *material objects* (bodies, things). For example: tables, birds, mountains. But all objects of perception are not what we call "material objects". Sounds, shadows, mirror images, smells, for example, are not.

Objects of perception, moreover, are not necessarily "thing-like". Also changes, events, ongoing processes can be genuine perceptual objects.

III

1.

I can report my sensations when I have them and I can describe them from memory – for example to the doctor who, I hope, will cure my back pain. There is surely a sense in which I can be said to observe or watch my sensations – for example whether my pain increases when the doctor touches such and such a place in my body.

Can I *test* my sensation reports? In some sense Yes. I was in grave pain but is it still there? I "feel" whether it is still felt and say perhaps that I hardly feel it any more. This is testing – but it is not testing the veracity of the report I gave two minutes ago. I cannot test my sensation reports in the same sense as I test my perception reports. But when I say I have a pain I *mean* that "that which I have" is pain. (That is: unless I am feigning or lying.) My sensation has a content and I am reporting which it is. I cannot be mistaken about what the sensation *is*, but I may be mistaken as to what it is called – for example, if I make my report to the doctor in a foreign language.

But there is also another possible type of mistake. Some sensations are "conceptually vague". One sometimes mistakes a sensation of a certain kind for a sensation of a different kind. For example: a weak pain may be "mistaken" for a tickling sensation. Further "observation" may make us decide that it is a pain and not a tickling sensation. Or that it is a "tickling pain". This is one of the many cases which encourages us to think of sensations as having a real object which we can scrutinize by "feeling" it.

2.

One can turn away one's attention from sensations – as one can close one's eyes to a sight. But the two cases are very dissimilar. I can, for example, ignore a slight pain so that I do not feel it. But the physical cause of the pain in my body may still be operating uninterruptedly. (Some process in a tooth, say.) As soon as I turn my attention to the fact that I felt pain I feel it again. It is tempting to say that the pain was there all the time, "subconsciously", "to be felt", only I did not feel it. Just as the sight, to which I had shut my eyes was

there "to be seen" all the time. Talk of "subconscious sensations" is innocuous, one could say, only as long as one remembers that they are not *sensations*.

If the extra-neural cause of pain was operating although I did not feel the pain, then, presumably, the intra-neural processes which "mediate" or "transfer" the feeling from the pain-nerves to the centre, thereby calling forth the sensation (of pain), were somehow "blocked". The idea is that there is a difference *at the neural level* between the case when I feel and when I do not feel the pain (when I feel the pain and when it stays "subconscious"). This is, of course, sheer hypothesis – but it seems to me very plausible, because the thing we call "turning one's attention from something to something else" or "concentrating on something rather than on something else" is surely an activity "of the mind" which *also* has a bodily aspect. Its presence would *explain* why it is that, by "attending" or "concentrating", the uninterrupted operation of the physical cause of pain can be "felt" or not "felt" as a sensation. One is even tempted to say that without some such explanation the phenomenon which we have been discussing would be unintelligible. This is admittedly a "philosopher's dogma" – but is it unwarranted? Cannot demands of "scientific intelligibility" exist, so to speak, in their own right?

3.

It has been a matter of dispute whether one can, in a genuine sense, be said to *know* that one has such and such sensations, for example that one is in pain.

The sentence "I know that I am in pain" has an intelligible use. I can address it to somebody who doubts whether I really have pains or only pretend. But under this use it is only an emphatic way of saying "I am in pain". For this reason one may say that the use of "I know" here is not "genuine".

What then is a "genuine" use of the phrase? One could suggest the following answer: "I know that ---" is used in a genuine way when it tells something about *me* which is not already told by the sentence "---" following the phrase. Something like that I have ascertained (investigated) the matter, or been taught or told it by someone (whom I trust), or perceived (seen) or otherwise witnessed it. If one refuses the use of terms like "ascertain" and "observe" in first person statements about sensations one also has reason to refuse the use of "I know". But once one sees clearly that the uses in question are "analogical", one can also allow the use of "I know" here.

What epistemologists have called "direct" or "immediate" knowledge ("the given") seems in many cases to be such "knowledge" by the sensing subject of his own sensations. This concept has an interesting feature from the logical point of view. It is a kind of *necessary knowledge*. One can argue:

"Surely I know whether I am in pain or not. (Barring cases of conceptual vagueness.) This is necessarily so." It does not follow that I necessarily know

that I have pain or necessarily know that I have not. But I know necessarily which one of the two is the case. This is not true of "genuine" knowledge. There necessarily is a bird in the tree over there or there is no bird. But I do not necessarily know whether there is one or not. If, however, I *look*, then I necessarily know whether I see a bird there or not, whether the object of my perception is or is not *X* (a bird).

These observations on the logic of knowledge in relation to perception and sensation seem to me interesting. They can be related also to ancient ideas of an omniscient being.[1]

4.

If I "really" am in pain there is no verification by me of the statement that I am in pain, no ascertaining or testing by me whether I am in pain or not. My *saying* "I am in pain" is expressive rather than descriptive use of language. One could say, with Wittgenstein, that it is a form of pain-behaviour. Unlike the "natural" expressions of pain such as crying or screaming or contortions of the face and other parts of the body, saying "I am in pain" is *acquired, learnt* behaviour. It is acquired with the learning of language and the acquisition of the *concept* of pain.

The statement *about me* that I am in pain is open to testing by others. The testing relies, primarily, on the criteria which my pain-behaviour provides. In normal human communication the criterion provided by the person's verbal report is decisive. "If he says he is in pain, he is in pain." Sometimes, however, we cannot trust (just) his words. Then further observations about him are necessary. Most of the things which fall under the concept "pain behaviour" can be simulated. Can we ever be sure whether he is in pain or not? *As a matter of fact* we often *are* sure – even in cases when our judgement disagrees with the "sufferer's" own. Does it mean that in such cases of disagreement he is lying, insincere, simulating? Not necessarily.

The *behavioural criteria* of the truth of the statement that a person has a sensation of a certain kind must be distinguished from what I propose to call *causal signs* or symptoms.

Sensations, as well as perceptions, have a causal origin in some physical event. In the case of bodily sensations, the cause is internal to the body of the sensing subject but external to his nervous system. Toothache, for example, is caused by a "bad tooth". The states or processes which make teeth bad can be subject to observation from "outside" (by a dentist). These observations may be such as to remove any doubt that the person who complains of toothache really has toothache – even in the absence of all other "pain behaviour" but the verbal statement. Or the doctor finds nothing wrong with the tooth and concludes that the patient at least cannot have *tooth*ache. He may, because of

other criteria, not doubt that the patient is in *pain* but conjecture that the cause of the pain, which qualitatively may resemble toothache, is some process internal to the nervous system itself.

Unlike the "natural" behavioural criteria, the causal symptoms of a sensation presuppose the *concept* of the sensation in question. In order to know, for example, that a certain affliction of the body *causes* pain we must know what pain *is* (*what* it is that is being caused). This we know on the basis of pain behaviour: verbal complaints, cries and contortions, the sufferer's retreating from the perceived cause of the state he is in, *etc.*

By "retreating from the cause" I mean such reactions as that we withdraw our hand from a burning hot object which we happen to touch. This is a *spontaneous* (reflex) reaction. It does not amount to "knowing the cause" of the pain. The fact that the reaction occurs is a feature of what it *means* to feel pain. When, at a later stage, we *shun* (avoid contact with) hot objects, or fire, *etc.* an attitude which might be called "causal knowledge" is being built up in us. When fully developed in humans it may be said to *presuppose* the concept (of pain), and not be *constitutive* of it. But the border between spontaneous avoidance of something and learnt shunning is not sharp.

Nor is the distinction between behavioural criteria and causal signs sharp. If, for example, by a conscious "effort of will" a person suppressed all manifestations of pain, both verbal and natural, we may still be certain that he is in pain because of the affliction he has suffered and because of what an examination reveals about his nervous system. Our evidence would overrule his denials.

It is a conceptual feature of the behavioural criteria of sensations that they are individually neither necessary nor sufficient to establish the occurrence of the sensation. Nor do they form a definite body of behavioural traits which collectively amounted to a necessary and sufficient condition of the sensation. What the criteria under given circumstances show depends upon the circumstances. And normally we agree about their testimony. We cannot doubt that he is in pain. Or we are sure that he is only pretending. Or we do not "know what to think".

Yet it is this openness (as I shall call it) of the behavioural criteria of sensations which nourishes, on the one hand, scepticism about the possibilities of ever "knowing for certain" whether another person has a certain sensation and, on the other hand, the dogmatism of indubitable introspective knowledge by the subject himself of his sensations.

5.

Assume that I am in *grave* pain. "Surely I know, 'by introspection', that I am in pain then." What information does saying this convey? For example: that this is not a case of "conceptual vagueness" in which I may hesitate whether

the sensation is one of pain or of tickling. It is a "clear case" of pain; a case in which I necessarily know, "by introspection", whether I have pain or not. Such cases, moreover, are the normal or typical cases; if they were not this, then our *concept* of pain would be different from what it is now. This is important. Because my certainty, in a "clear case", that the sensation I have is one of *pain* is my certainty that I know (master) the correct use of the *word* for it (in English "pain"). I can use it for asking for help or for soothing or pity – or for explaining why I cannot now perform a certain action which is asked or expected of me. And I can "embellish" the use of the word with cries and moanings; sometimes I am forced to do so (cannot resist doing so) because of the effects which the affliction has on the rest of my bodily reactions (spontaneous pain-behaviour). I know the spontaneous bodily reactions (pain-behaviour) which "match" the use of the word.

"Knowing that I am in pain" – "Knowing *by introspection* that I am in pain". What does the phrase "by introspection" add to the content of the sentence? It may refer to an act of *reflecting* on the sensation. I say to myself "this, surely, is *grave* pain", or "the pain is still there although hardly perceptible". And one can think of other functions for the phrase.

"I know that I am in pain" – "I am in pain". What do the words "I know" add to the meaning? Perhaps the same as the words "I introspect" might add. Or, as in the case discussed above, they can express my certainty that I use the right word to describe my sensation. In many cases, however, the words are just "empty".

But – at least when the pain is grave – do I ***not know*** that I am in pain even in the (unlikely) case that I have not learnt a word for the sensation? Surely I "know" *what I have*; it is *this*. Don't you hear my cries and see my tears? Help me, help me! What is it then that I "know"? Well, I *am* in (*have*) grave pain – and I either yield to the "natural" ways of expressing it, or make a "heroic effort" to conceal it. In no other way is now my pain an object of "knowledge" for me. Outside observers, however, may be certain that it is pain which I suffer even when I try to conceal it. To them *my* pain is a (genuine) object of knowledge. (Although they do not *have* it.)

6.

Consider the chain: affliction of the body – neural events – sensation of pain – pain behaviour. One would think of it as a chain of successive causes and effects. It contains a link, however, which is "philosophically suspect", *viz.* the sensation of pain. It is something "mental", all the other links being "physical". Can something physical cause something mental, and something mental something physical? This is an old problem.

In order to establish that there is a causal relation between some neural events and a sensation of pain one might try to produce pain by stimulating the nerves. This requires that the person (or animal) on whom the experiment is performed somehow or other "acknowledges" the effect. He reports it verbally or cries out or jerks as if to rid himself of his discomfort. Here it seems quite natural to say that what happens in the nerves *causes* these various forms of pain behaviour – and the "mediating" causal rôle of the sensation (the pain) seems to drop out as not being needed. The reason for this is not just that, by transitivity, the cause of the pain becomes the cause of the pain behaviour. The reason is rather that, in order to establish that the nervous events cause the pain we must *first* connect them with various forms of pain behaviour. The sensation (the pain) has no independently establishable causal function to perform – neither as the (mental) effect of a (physical) cause, nor as the (mental) cause of a (physical) effect. The talk of causal ties between the two realms, the mental and the physical, becomes otiose. With this also the problem how such causation is "possible" vanishes.

But can I not testify that it is *because of the pain* that I cry or jump about – and also that it is because of the pain that I say I am in pain? If it were not for the pain I would not behave as I do. True enough – but is this a *causal* "because"? Or is it a conceptual "because"? That is: is it more akin to a reason than to a cause?

If I cry and someone asks Why? a possible answer would be "Because it hurts so terribly". I am told "Don't cry. It does not help. Control yourself". Maybe I stop crying and say "Yes, it is no use crying". Or maybe I do not stop but say "I can't, I can't, it hurts so terribly". In the first case I view my crying as an *action* for a certain *reason* and therewith as something which I am, as we say, "free" also *not* to do. In the second case I just react to the pain "spontaneously" as we say. Even then I can, usually, *try* to control my reaction. I, for example, clench my teeth, pinch my arm, beat my breast with closed fists, breath regularly and slowly. These are changes in my *bodily* condition which relieve the pain or, at least, "distract" me from feeling it.

When we react spontaneously to the pain we do not react to it (act on it) as a *reason*. Still there is something odd about calling the pain a cause of, say, my crying. What caused the tears to stream from my eyes was, I imagine, some processes in my body which had their causal origin in the "affliction" which (also) gave me the pain.

In order to do justice to the meaning of "I cry because I am in pain" when crying is spontaneous we must, I think, acknowledge a "because" which is neither causal nor conceptual. I shall coin for it the name *expressive*.

Expressive behaviour or movements have traditionally been studied in the psychology of *emotions*. Joy and sadness, anger and fear, affection and tenderness have characteristic bodily expressions. Like pain behaviour and other

expressions for sensations they can to a large extent be simulated so that the expressions may be there without the "corresponding" emotion (feeling). But it is also well known that these expressions have a "reinforcing" effect on the emotions. This observation was exploited for theoretical purposes in the famous James–Lange theory of emotions. "We do not cry because we feel sad but feel sad because we cry", *etc.* Applied to sensations we should then have to say "We do not cry because we are in pain but are in pain because we cry." There is a grain of truth in both formulations. The truth is that there is a *conceptual* relation between, on the one hand, the emotions and sensations and, on the other hand, the bodily expressions. The latter are the criteria relying on which we come to know the existence (presence) of the former. But there is also something wrong with the quoted formulations. What is wrong is that they appear to deny what I have called the *expressive* relation between the two types of phenomena, those of the "soul" and those of the body.

(Like perceptions, emotions, too, have a sensational element. Love and hatred, joy and sadness are also things we "*feel*" – and in this they are like bodily pains and pleasures.)

"I cry because I am in pain" is not a physiological explanation of my tears. I assume that they *have* a physiological explanation. This will tell how certain neural processes of peripheral origin were "propagated" to the tear-glands, activating the glands and making them produce a liquid which then emerges from the eyes. This story is for the neurophysiologist to tell. But even if there *were* no such story to be told, I can usually relate my tears to some affliction which I (my body) suffered; I fell and broke a leg, say. This too is a "physiological explanation" of a sort. If I give it, people will understand it as an explanation, although they will sometimes doubt whether I am speaking the truth. Also when I do not myself know the nature of the affliction, what gave me the pain, I am prone to think that it must have its origin in something which happens to or in my body. We are strongly disinclined to think that pains occur without a bodily cause. We think that this must be true at least of everything which can be called "bodily pain". Pain without a cause in some affliction of the body we should presumably hesitate to call a "sensation", and classify with such emotions as grief or anger.

7.

Perceptions can be shared in the sense that the perceptions of different perceivers can have the same object. Sensations too can be shared in the sense that they are qualitatively the same. (Cf. above p. 50f.) But one person cannot have another person's perceptions nor another person's sensations. This follows from the *distinction* between them as persons (p. 51). This distinction again is connected with there being different bodies. By abstracting from differences

between some bodies one can form various concepts (ideas) of collective or corporate bodies and persons. One can, for example, regard the members of a nation or religious community as the body of one (super)person – and speak, for example, of its afflictions and sufferings. One should not regard such locutions as purely metaphorical. The sufferings of a people too is a "mental phenomenon" and I see no meaning in the talk that it can be "reduced" to the sufferings of several or all of its individual members. The eminent degree of privacy which *we* attribute to the "inner life" of individual human beings is a product of our individualistic and subjectivistic culture. To say and reflect on this is not to yield to bottomless cultural relativism.

IV

Colour Inversion

1.

I should like to take up for discussion the adolescent fantasy that my "subjective experience" of x is similar to your "subjective experience" of y, and *vice versa* my experience of y similar to yours of x.

The fantasy has traditionally been entertained in the form of an imagined "inversion" of peoples' colour spectra. We both call the same things "red" and "blue", say. But your sensation of red is similar to my sensation of blue, and *vice versa*. Is this (logically, conceptually) *possible*?

One could raise the same question about sound, for example. What you hear as high-pitched tones I hear as low ones, although we agree about whether the tones we hear are to be called "high" or "low".

2.

Could findings of neuroscience give evidence in support of the "inversion" of colour spectra? Let us consider the following possibility:

We have a general description of the visual apparatus V of adult humans. Anomalies and defects apart, normal humans have a visual apparatus of this description.

When the apparatus is stimulated with light of a certain colour, certain reactions occur in V (the eye and optic nerve) which are different with lights of different colour. Now it is found that in the V of some subjects, stimulation with red light causes the reaction R to take place, and stimulation with blue light the reaction R'. With other subjects it is just the other way round; stimulation with red light causes R' and with blue light R. This is a possibility. Assume now that it gives rise to the following speculation:

3.

There are two subjects *S* and *S'*. The *V* of *S* reacts with *R* on red and *R'* on blue; the *V* of *S'* again with *R* on blue and *R'* on red. Does this not support a hypothesis that *S* sees red as *S'* sees blue, and *vice versa*? The hypothesis is that the "subjective experience" of *S* of red is similar to the "subjective experience" of *S'* of blue, and *vice versa*. They agree in their use of the colour *words* but differ in their experience of the *colours* themselves.

What does it mean that *S* "sees red" in the same way as *S'* "sees blue", *i.e.* that their "subjective experiences" of the two colours are similar?

What could be the standard of comparison here? It is *not* that *S* calls things "red" which *S'* calls "blue". None of them mistakes things for having another colour than they have – as can happen in the cases of colour blindness. Remember that they agree in their colour judgements!

Seeing red things may call forth certain reactions in *S*. He is perhaps an irritable person and seeing red things makes him angry or aggressive. We can imagine that *S'* reacts in the same way when he sees blue things. If there is a remarkable similarity in the reactions of the two persons to things of the two colours we might say that *S* sees red as *S'* sees blue. Saying this would have a clear meaning and one could ascertain whether the truth-conditions of the statement are satisfied. So one can *make sense* of the statement that a person's "subjective experience" of one colour is similar to another person's experience of another colour – and also of the statement that the first person's experience of the second colour is similar to the second person's experience of the first colour.

Is this "colour-inversion"? Call it this if you like.

4.

In the absence of all other indications that the two subjects "see" the two colours "differently", the obvious thing to say would be, in the first place, that we have found a difference in their respective visual apparatus. If this difference is such that, say, the *V* of *S* is like that of (most) normal people in that it reacts with *R* on seeing red and with *R'* on seeing blue, we should say that we have discovered an "anomaly" with the *V* of *S'*. This anomaly could be very common, so common perhaps that about half of the human race has a *V* which reacts on seeing (stimulation with) red (light) in the *S*-way and half has a *V* which reacts in the *S'*-way. And then we should no longer call it an "anomaly". Would it mean that the subjective experience of about half of all humans when they see red things is the same as the subjective experience of the other half when they see blue things (and *vice versa*)? We are back at the question what it *means* that I "see red" as you "see blue" (and *vice versa*).

As already said, one can make sense of this. You and I may have different reactions to the two colours although we agree in our colour judgements (attributions of this or that colour to given objects). In the absence of such differences in reactions one can only say that there is a difference in the way their *V*s react to red and blue objects – and, moreover, that the difference has the symmetry of an "inversion".

5.

Let us assume that there *are* such characteristic differences in the reactions of the persons in addition to the differences in the reactions of their *V*'s to the colours.

Assume that there is some generality about this. So that persons generally, who, say, dislike red things and like blue things have *V*'s which react to red in the *R*-way and to blue in the *R'*-way, whereas persons who like red and dislike blue things have *V*'s which react to blue in the *R*-way and to red in the *R'*-way. This would indicate a correlation between their different "subjective experience" of the two colours and a difference in the functioning of their *V*'s, *viz.* that they react to the two colours with different *emotions*. Would it indicate that they "see" the two colours differently? Yes, if by "seeing differently" we mean having different emotional attitudes to the two colours. This is a respectable sense of the phrase "see two colours differently". And if someone said that persons of the one category see red as persons of the other see blue and *vice versa*, and what he means is that they react to the two colours with the characteristic emotions – this would be in order.

If there is no such generality, the thing to say would be (only) that there are some persons whose *V* reacts to red light in the same way as the *V* of other persons reacts to blue light. If there are only very few persons of the one kind, we might speak of an "anomaly" in the function of their *V*. If there are about equally many of each category we would record that there are two ways in which persons' *V*'s react to red and two in which they react to blue light.

6.

Nothing we have so far said will satisfy one whose mind is captivated by the "inversion argument". The difference in "subjective experience" which he has in mind is not a difference in, say, emotional reactions to colours but an alleged difference in the very *colour sensations*. "The red sensation of *S* is similar to the blue sensation of *S'*, and *vice versa*", he insists.

What is he then insisting on? This is *very* far from clear. There is this picture: If *S* could *see* what (how) *S'* sees when both are looking at the same coloured thing, *S* who sees the thing as red would find that *S'* sees it as blue

although he calls the colour "red". *S* has a red sensation and *S'* a blue one. (The very verbal expression is here faulty, since *impressions* are not coloured. "A red impression" means an impression of (something) red.)

But how could *S* see what *S'* sees? This does not now mean "see the same object" but "see *S*'s impression (visual colour image)". But to *see* *S*'s impression must mean to *have* *S*'s impression. The question, therefore, is whether *S*, in addition to his own impression of the seen thing, can also have *S*'s impression of it. What does *this* mean?

S can look at the same red object from the same, or nearly the same place as *S'*. This is a sense in which *S* can be said to have *S*'s impression of the thing. We sometimes express ourselves thus. There is nothing "mystical" about this. But it is not what our "subjectivist" philosopher is after. So what *is* he after?

7.

Let us introduce a third person *P*. Could he check or compare the impressions of *S* and *S'*? Assume he could "peep into" the brains of *S* and *S'* (and know *where* to look). He can then see what happens in the brain of the one and also of the other when they look at the same object (their *V*'s being affected by light from it). What *he*, *P*, then sees are brain events, not colour patches. He may notice that what happens in the two brains (the two *V*'s) is the same kind of process – but he may also notice some difference, even a difference which he would describe by saying that what happens in the brain of *S* is similar to that which happens in the brain of *S'* when *S'* sees a blue thing, and what happens in the brain of *S'* is the same as what happens in the brain of *S* when *S* sees a red thing. Would this observation speak in favour of a difference in their "subjective experiences"? No, unless there is some *other* indication that *S* sees red as *S'* sees blue, and *vice versa*.

Thus peeping into the brains of persons cannot establish that their subjective experiences (of red and blue) are "inverted". In order to establish this one would, as it were, have to "peep into" their *minds*, and not their *brains*. The idea is, roughly, that *P* inspects (observes, sees) the visual images of *S* and *S'* and finds, to his amazement, that the visual image of *S* of an object they both (all) call "red" is blue – and the visual image of *S'* of an object they both call "blue" is red! We can extend this uninhibited play of imagination even a little further and imagine that *P*'s visual image of the colours of the two objects agrees neither with that of *S* nor with that of *S'*. He would then say, for example: "Objects which I see as red and blue, *S* sees as yellow and green and *S'* as green and yellow. But we all agree to call them "red" and "blue" respectively. This is what we have all been taught are the names of their colour!"

8.

Why is this nonsense? Because the sensation, impression, visual image is not anything we *see*. It is something we *have*.

But saying this does not yet settle the matter. We can watch, make observations on our visual images – as distinct from watching the seen object. Is this not a way of "seeing the impression itself"? And is it not conceivable that we *have* the impressions of others?

Observing an object and paying attention to my (sense-) impression of the object is not the same. I say, for example, "I do not see the details clearly". "I cannot take in the whole thing at one glance; if I look at the centre the edges seem blurred", *etc.* With such remarks I report my sensations of the object. One can call it observing and describing the *sensation* as distinct from observing and describing the *object* (itself). But this does not mean that I *see* my sensation (visual image). It makes no sense to talk about one's sensation of one's sensations.

We indicated earlier (II, 1) in which sense two subjects can have the same impressions. One can call this a *sharing of impressions*. If somebody said that they "have each other's impressions" – why not. Or that the impression of *S* is also the impression of *S'*, and *vice versa*. We express ourselves thus. But what we mean is not that which our "subjectivist philosopher" is after.

9.

A last "desperate" attempt.

Can we not think that the brains of *S* and *S'* are thus connected that *S* can "have" the impression which *S'* has when the *V* of *S'* is stimulated with red light? And now, lo and behold! When the *V* of *S'* is thus stimulated a *blue* patch appears "behind the retina" of *S*. (Perhaps *S*, in order to have this sensation from the *V* of *S'*, must close his eyes so as to shut out stimulations of *his V* from outside.) *S* asks *S'*: "What colour do you see?" *S'* answers: "Red". *S* says: "I see blue". Then *S* looks at the object from which light is reflected (emitted) to the eye of *S'* and notices that this object is – red. Shall he now say: "Evidently the impression of *S'* of red is similar to my impression of blue"? On the basis of his observations *S* can say this – and mean exactly what the observations tell him. This is in order, as long as he does not put unwarranted interpretations on what he observes. One such unwarranted leap would be if *S* said that he *has* the impression which *S'* has. What *S* has is an impression (of his own) the physiological basis of which are processes in the *V* of *S' and* in the "link" connecting the *V* of *S'* with the *V* of *S* himself. If one is struck by the discrepancy in colour impressions one would have to make comparative studies of the anatomy and physiology of the two *V*'s and, in particular, of the linkage. And one may, or may not, find an acceptable

explanation for the fact that when the *V* of *S'* transmits red light to the brain of *S'*, then *S'* sees red, but when the "impulses" in the *V* of *S'* are transferred to the brain of *S* then *S* sees blue. If the two subjects tried to describe to one another their impressions they would describe them differently. *S* would perhaps say that his impression is that of the colour of strawberries, whereas *S'* would say that his is that of the colour of the sky on a clear day. They would agree that their colour impressions are different.

10.

It makes good sense to say that two persons compare their impressions (sensations) of something.

One can ask whether two persons' sensations of, say, red, are qualitatively similar. One way of testing whether they are, would be to make the persons classify other objects as being of the same colour as a given object or not. If their classifications are practically identical we should say that they see red in qualitatively the same way. If there is a noticeable difference one would say that they "see red" differently. They may still agree that the object which is the basis of the comparisons is red. (One of them classifies, say, also orange objects as of the same colour with the red object. He is "less sensitive" to difference in colour one would then say.)

By seeing red in the same way or differently one can also mean that one "evaluates" the quality differently, that the emotional reactions of subjects to the colour differ. In this sense of similarity and difference one can even speak of "colour inversion". (Cf. above p. 61.)

One can compare the impressions which different subjects have of the colour of the same object. But one cannot compare the colour of the impressions which different subjects have of an object. "Impressions of colour are colourless" one could say. If our "subjectivist philosopher" does not admit this, he is confused about the concept of an impression (sensation), about the "grammar" of the words "(sense-)impression" and "sensation".

This will suffice about the fantasy of colour inversion.

V

On Mirror Images and Echoes

1.

I look in the mirror. What do I see? My face. Or shall we say: a mirror image of my face?

One could say both – and be right. But is not saying the second another, and easily misleading, way of saying that I see *my face* in a mirror?

I wish to argue that mirror images are not *percepts*, and therefore not "physical phenomena" in the sense I have previously explicated this concept. This may initially sound strange, but I hope to be able to dispel the impression of strangeness.

I see the mirror image of my face. Another person, looking in the same mirror, sees it too. So, the mirror image is there to be seen by others too (and not only by me). I close my eyes or turn away my head and do not see the mirror image. Then I look in the mirror again and there the image is. It is, as it were, there to be seen.

Speaking in this way about mirror images does not strike me as unnatural. Speaking thus means speaking about mirror images as if they were percepts. But is this so?

Where is the mirror image? In the mirror? What would this mean? Presumably it would mean that the image is on the surface of the glass or reflecting surface. But this is not where we see the image. In "visual space" it is located *behind* the mirror, at the same distance from the surface as your face is.

But surely no image "exists" behind the mirror at that distance. The mirror may be hanging on the wall; so in order to find the image one would have to penetrate the wall! It is said that dogs (who are supposed not to "understand" mirror images) sometimes when they see themselves in a mirror look for another dog on the other side of it. ("Understanding mirror images" is a philosophically interesting notion.)

Shall we say then that the mirror image is *nowhere* (in physical space)? Then it would not be a percept and therefore also not a "physical phenomenon". Could one then say even that it *exists*?

As anyone with minimal insight in geometrical optics knows, mirror-images are said to be "virtual" images (only). They cannot, for example, be projected onto a screen, as can photographic images (slides). A photographic picture (of a face, say) is a *real* image. It is a percept, a physical phenomenon.

When the water is calm I see the passing clouds and the surrounding wood reflected in the lake. I turn away, and nobody else is there to enjoy the sight. Is the reflection still there, to be seen also when nobody is there to see it? The best answer seems to be No. The clouds may still be passing and the wood is there independently of the spectator. One can see the *clouds* and the *woods* reflected in the water, *i.e.* by looking in the direction of the water. But one does not *see* their mirror images.

Do not be dogmatic! Surely one "sees" their mirror images, if one takes a look. Yes – but what this means is that one sees *them* by looking not at them but into the water.

Thus to see a mirror image of something is to see this thing reflected in a mirror. The percept is the "something" which one sees. *It* is a physical phenomenon.

So what then is the mirror image "itself"? It is a visual sensation. Since its object has to exist in the physical world (otherwise light from it could not be reflected from a mirror) it is also a visual perception.

Sensations and perceptions may be called *mental* phenomena. A mirror image, therefore, is a mental phenomenon.

Mirror images do not "exist" ("belong") in the physical world. But an account of how mirror images originate belongs to geometrical optics and involves principles of the reflection of light. Light is a physical phenomenon.

A mirror image can give a distorted picture of the seen object. The object, for example, appears smaller or larger in the mirror than it would appear if looked at from the same distance as the mirror is from the eye. And it can be distorted in many other ways, too. These are familiar phenomena.

One can give an account of the way a mirror image is distorted. This account belongs to geometrical optics and is partly understood in terms of geometrical properties of the reflecting surface (mirror): whether it is concave or convex, *etc.*

Assume that the mirror enlarges. Do I then see in it an "enlargement of my face"? It is better not say this. What I see is my *face*. This is the percept. But it appears (looks, seems to me) bigger than it is. What does it mean that something *looks* bigger than it *is*? It is worth reflecting about the answer. An object which, when *looked at* ("directly") would look that big would be much bigger than the object reflected in the mirror.

To look at things in a mirror is an *indirect* way of looking at them. Mirror images could therefore also be called "indirect(-ly obtained) perceptions". It is interesting (philosophically, psychologically) that such phenomena should exist.

Some things which I see in a mirror are things which I can also see directly. For example, by turning my head. But my face (my eyes) I can only see indirectly. This too is interesting.

2.

It is instructive to compare mirror images with echoes.

I stand on the shore of the lake on a calm evening and shout "hallo". After a short while I hear the sound reflected from the cliffs on the opposite shore.

What did I hear then? My voice, the sound I produced? Or the echo of it? If I say that I hear the echo, what I say is that I hear my *voice* reflected.

The case of echoes is analogous to that of mirror images. The percept is the sound which is reflected as an echo. An echo is an acoustic sensation. If it is a "real" echo it is produced by a "real" sound and is thus a perception of a

sound. The echo, therefore, is a "mental" and not a "physical" phenomenon. But an account of how echoes occur, originate, belongs to physics (acoustics) and makes reference to laws about the propagation and reflection of sound.

This is a difference between echoes and mirror images: I cannot see my eyes (face), but I can hear my voice (or any sound I produce). I shout out something, and hear it at once. After a few seconds I hear it again reflected as an echo. Did I hear one or two sounds: the "original" one and the reflected one? I think one should answer: I hear the same sound twice over. But the two acoustic sensations can be rather different. The one is perhaps louder and more distinct.

One is tempted to say: The two perceptions have the same percept (the sound, an individual physical phenomenon), but they differ in their "sensational component". We perceive the percept differently.

3.

Of mirror images and echoes the principle *esse est percipi* holds true. Berkeleian "epistemological idealism" could be characterized as a view according to which all things which we perceive are like reflections. But a reflection without something reflected is a self-contradictory idea. So is also this version of idealism.

NOTE

[1] Cf. the note on "Knowledge and Necessity" in my book (1984, 68–71).

REFERENCE

von Wright, G. H. (1984), *Truth, Knowledge, and Modality*. Oxford, Basil Blackwell.

THING AND QUALITY. SUBSTANCE

I

Physical Phenomena – Physical Things. What are They?

1.

A prototype of physical things are (solid) material bodies in three-dimensional space. What about liquids and gasses? They, too, exist in three-dimensional space. But they have no shape "of their own". They may be contained within solid boundaries and then assume the shape of the container or vessel.

Solid material bodies are composed of parts which are also solid. Are not liquids and gasses likewise composed of "solid material bodies" (molecules, atoms)? Perhaps – but not in the same sense in which perceptible or *macroscopic* solid material bodies are thus composed. The "material bodies" which compose liquids and gasses are *microscopic* and not perceptible (in the normal way).

Also the smallest of macroscopic *solid* material bodies are composed of microscopic bodies – in principle like those which compose liquids and gasses. Thus it is according to the atomistic conception of matter. This conception is much older than the atomic theory of our science.

But molecules and atoms, we are told, are *not solid* material bodies. They are more like "systems" of smaller bodies (electrons, protons, *etc.*). Are these components "solid"? Here the *concept* of solidity (of being a solid material body) is already becoming obscure.

Is an atom more like a solid body than it is like a liquid or a gas? It is not easy to grasp the meaning of the question.

2.

We create for ourselves a *picture* of the microworld: it is a world of particles ("elementary particles") *like solid material bodies in various constellations in space*. An atom is perhaps like a mini-mini-miniature of the solar system.

But this picture may be very misleading. Are elementary particles "really" solid material bodies? A physicist would probably brush the question aside.

But in a microscope we can *see* small material bodies which are not visible

G. H. von Wright, In the Shadow of Descartes. Essays in the Philosophy of Mind, 69–82.

to the naked eye. Are they not "solid" then? Some are, others perhaps not. They may have a definite position relative to some other small bodies, or they may move in observable ways relative to one another.

What are the criteria of a body being "solid"? We distinguish *solid* and *hollow* bodies. A hollow body, however, has a crust which, in the normal case of macroscopic bodies, is solid matter, *i.e.* which we can break up into *solid* material bodies.

The notion of solidity is logically prior to the notion of being hollow.

Are atoms and elementary particles more like hollow than like solid bodies, then? This question, too, should be brushed aside. An atom is not bounded like a mini-balloon.

The notions of "solid" and "hollow" and therewith the notion of material body apply primarily to macroscopic physical bodies. It is not clear how to apply them, or whether they apply at all, to (very small) microscopic bodies. But they certainly apply also to some bodies other than those which are visible to the eye.

It is not at all surprising, but is on the contrary to be expected, that applying the notions of spatial position and speed of movement to bodies "in the atomic dimension" should lead to *conceptual* difficulties.

Saying, for example, that an electron has a position and a velocity but that they cannot be determined (measured) with exactitude because the measurement (observation) will interfere with the measured, is misleading. The picture of the electron in a definite position is a confused picture.

We are here in the neighbourhood of the classic debate over the "reality of atoms". *Of course* atoms are real. But their "reality" is rather different (conceptually) from the reality of macroscopic, hollow or solid, material bodies.

3.

Are the things I see in a microscope but which I could not see with the naked eye *percepts*? For example, cells. Yes, but why? Because I can say about that which I see in the microscope very much the same things as I can say about macroscopic things which I see. For example: I may be able to tell their colour, or their shape, their position in a group (colony) of such things, whether they move or are stationary, *etc.* I may not be able to touch them with my finger or smell them. So they are not *exactly* like "normal" macroscopic percepts. But they are sufficiently like them to be called "percepts".

Looking into a microscope is like peeping into a different world. Like a world of imagination and legend – and yet it is "real". To understand that the things I see are the same as those which are there, right in front of us, within reach of our hands, already requires some rudimentary familiarity with

geometrical optics and instruments like spectacles and telescopes. For a child or a representative of an "alien culture" this may not be altogether easy.

But what about molecules or atoms – or electrons and other "elementary particles". Are they percepts? Does what I see in a cloud chamber or an electronic microscope amount to seeing *them* or only to seeing "traces" which they leave in perceptible surroundings? Most things which I can say about macroscopic objects on the basis of looking at them, touching them, *etc.* I can*not* say about, *e.g.*, atoms. Therefore it is unnatural to call them "percepts". But I can perceive a good many things which I may relate to them. I may, for example, be able to trace their trajectory in a closed space. But this I can only do because of theories which we have built and which are such that with their aid we can account for perceptions which we have had or predict perceptions which we shall have in *macroscopic* circumstances.

Denying that atoms are possible percepts is not to deny that they possess "reality", or to deny that they exist in the physical world.

4.

Macroscopic material bodies can be seen by looking at them and felt by touching them. We may also smell them and taste them. But can we *hear* them (by listening to them)?

Hearing is peculiar – and it is to be regretted that traditional philosophy concerning sensation and perception has concentrated heavily on the optic (visual) and the haptic (tactile) – particularly on the former.

We say of some things that they make sounds. "The incessant roar of the sea." "The shriek of the siren."

Sounding things are also said to *emit* sound. *Where* is the sound which they emit? In the air? "The sound fills the air" we say of some sounds. It occupies space – just as the sounding thing does. But the sound and the sounding thing occupy different portions of space.

Comparisons can be made with colours. Something is coloured. The colour *is* on the surface of the thing. We also say that the coloured thing emits (reflects) *e.g.* red light. There is a similarity between *light* and sound in relation to space – and there is a similarity between *colour* and sound in relation to (coloured and sounding) things. (But there are also differences.)

All the time the thing is coloured, but we cannot see its colour because it is dark. All the time the thing is sounding, but we cannot hear it because of the distance.

But *is* the red thing *red* also when it is dark? Does it still have this *quality*?

5.

I *hear the sound* which a sounding object emits. But, normally, I do not *see the light* which an object reflects. I see the *colour* of the object.

I can, however, also see *coloured light*. The coloured light may have a visible source, *e.g.* a lamp. The colour of the light and the colour of the lamp are the same *quality*. But they are not qualities of the same *thing*.

That light is coloured means that things *in this light seem to have its colour*. ("Its colour" = "the colour of the light".) Here "seeming" plays an essential rôle. The colour in which the objects now appear is not their "real" colour. Their real colour is the colour in which they appear in noncoloured light.

(These things are more complicated, conceptually, than one may be inclined to think.)

6.

"The whole room was drenched in blue light." "The whole space was rosy." This can happen.

"Thunder filled the whole space. The *whole space*? No, of course not the whole. But so that everyone even miles away could hear it."

The object is red. *Where* is its colour? In the case of ordinary macroscopic bodies one might answer: The colour is on the surface of the thing. Or: the surface is coloured.

The object is sounding. *Where* is its sound? One would not (easily) say: In the object. One would rather say: In the air. Sound is not "attached" to the thing, a quality of the thing *in the same way* as colour or shape or hardness (softness).

Describing a sound. A sound can be intermittent or continuous, steady, increasing, decreasing, high, low, roaring, whispering, penetrating, *etc.* We then attribute *qualities* to the sound, treat the sound as if it were an *object* (percept).

Similarly with colour. The colour of the object is red. The red colour of the object is bright, is dark, *etc.*

This takes us to the distinction between object (thing) and quality (attribute, property).

7.

We construct for ourselves the following picture: the thing is a relatively small solid body. It is coloured, has a certain shape, also smell, taste, and touch. Let us even imagine that it is sounding.

Now we strip it of its qualities. Perhaps we can wash it so that it becomes colourless. Perhaps in the process (or in some similar intervention) also its

smell and taste withers away. And if it was sounding it is now mute. But what about shape and touch? As a solid body it has *some* shape even if it is, as we would say, quite "shapeless", like a lump of unmoulded clay. Possibly when we touch it, our finger sinks into it. Or it may be hard (as Newton said atoms are). *Some* tactile quality it will have.

If we imagine the object to be liquid, it has no shape "of its own" but assumes the shape of the vessel. And if gaseous it may not even have tactile qualities.

Shape and touch thus seem to be in a different category from, say, colour, taste, and smell. They were sometimes called *primary* qualities and contrasted with *secondary* qualities. The distinction has played a great rôle in philosophic thinking. It has also been criticized. That there is *something* to it is certain.

One thing which makes shape and touch different from other qualities is their connection with our notion of an individual material body. The *identity* of a body (and its identification) requires that it has shape. It must occupy a place "of its own". Unlike liquids and gasses which "take the place" of a vessel or container. The shape may be changing (now this, now that, perhaps indescribable in simple geometrical terms), but if the thing breaks up in parts or floats out in space beyond certain limits it loses its identity as *one* material body. A material body cannot be shapeless in the sense in which it can be, for example, scentless or colourless.

Touch is in a similar position. If there is nothing there to touch there is no material body. Liquids too can be touched, have tactile qualities. It is not clear what it means to "touch" a gas. But a gas may have tactile qualities – for example, be hot or cool.

Touch seems to hold an even more basic position than shape among (primary) qualities. A gas which, in addition to being invisible, shapeless, smelless, and tasteless, does not have any tactile quality, cannot even "exist", one feels tempted to say. On the other hand, it may have a smell and by this *alone* "vindicate its existence" to the senses. So having "touch" is not necessary. The gas which has an odour but has no other sensible qualities is still a (macroscopic) object of perception.

What of a gas which is completely imperceptible to the senses but which is, say, poisonous? Inhaling it may even have catastrophic consequences. Is it an object of perception? No. To give an account of its "existence" would involve a descent to the microlevel and is not possible without getting involved in physical *theory*. We can easily imagine a society in which no-one ever entertained the idea that poisoning could result from the inhalation of an imperceptible *gas*.

Of shape one may raise the question whether it *is* a (sense-) quality at all. Shape is connected with geometrical and topological relations. A body is spherical if there is a point inside it which has the same distance to every point

on its surface. But a spherical body is also *round*, and roundness, straightness and some other shapes must be counted as genuine visual (optic) qualities of things.

8.

So much for the *picture* we create for ourselves of the object-quality relation. It is founded on everyday experiences with macroscopic solid bodies and the changes which they may undergo (without loss of identity). When placed in a broader perspective the restricted relevance of this picture should become clear. Awareness of this again should make us ask what the distinction in question "really" amounts to.

The object-quality distinction is also related to the *grammatical* subject–predicate distinction. The grammatical subject stands for something of which the grammatical predicate is "predicated" – like a quality of a thing. The analogy is clear in the sentence "the ball is red". It is less clear in "the weather is bad". How thing-like is the weather? "It is raining." What is here the thing of which something is being predicated? The word "it" in the sentence is sometimes called the "formal subject". In some languages it is not expressed at all. In Finnish "it is raining" is translated by the one-word sentence "sataa".

A classic question which has troubled philosophers is this: What "remains" of a thing if it is "stripped" of *all* its (perceptible) qualities? And a classic answer is: Nothing remains. The thing just *is* a "bundle" or "complex" of qualities.

This answer is *not yet* tantamount to saying that a thing is a bundle of *sensations*. Or that *esse est percipi*. But it is, so to speak, an answer which points in the direction of the position known as "phenomenalism" or "sensationalism".

Can one make sense of the idea that a material thing is a bundle of perceptible qualities without falling into the pitfalls of phenomenalism? I shall make an effort to show that one can.

9.

The distinction between thing and quality is a distinction between *individual existence* and *generic attribute*. ("Qualities are universals.")

In the physical world, the individual existent occupies a certain *space* at a certain *time*. (Space and time are *principia individuationis*.) The same existent thing may at different times be differently placed in space. But it cannot at the same time be in more than one place in space. Its place in space and in time need not have sharp boundaries. Questions like "is it still here?", "is it already here?", "exactly where is it now?" cannot always be answered. (Think of

individual existents which are gaseous.) But in a good many cases the questions can be answered *to the satisfaction of the questioner.*

10.

Now think of a sound as an individual existent.

A sound fills (takes up) a certain (portion of) space. When we are inside this space we hear the sound (if we have normal hearing); if outside we do not hear it. If no-one is inside, no-one hears it. But the sound *is there* ("to be heard").

The sound usually fades away after some time. But the *same* sound may appear again, in the same place or in a different place. There are usually good and uncontroversial criteria of identity. They would often, but not always and necessarily relate the sound causally to a source, *i.e.* to another individual existent which "emits" the sound.

We can describe the individual sound in terms of its qualities. It is high or low pitched, strong or weak, continuous or intermittent, rolling (like thunder), piercing, shrieking, *etc.* These qualities would also count when judging the identity of the sound over a stretch of time. If it changes at once in all its qualities we would presumably speak of two different sounds. But one and the same sound may also "wax and wane".

In what sense then can a sound which nobody hears be said to exist? Someone might now answer: a sound which nobody hears exists, "is still there", in the form of waves in a certain volume of air (or other medium). This is what sound "really *is*", waves of a certain kind.

As everyone with an elementary knowledge of physics knows, sound is usually produced by vibrating bodies and its propagation through space involves wave-motion in the spatial medium surrounding the vibrating body, *i.e.*, normally, in the air. Of these movements a developed scientific theory exists which correlates features of the wave movements (frequency, wave-length) with such acoustic qualities as pitch, intensity, timbre, *etc.*

The acoustic qualities are not properties of the medium in (through) which sound is propagated. For example: air or water. The sound-*waves* are properties of the medium. They happen (occur, take place) in, for example, air or water. The movement in the air is, ultimately, an invisible "dance of the molecules". It is not perceptible to the eye. The access of perception to it is highly "indirect". This does not mean that the waves are just "theoretical constructions". They are as "real as anything" in the physical world. But they do not "look" like waves (for example in the ocean).

Of what then *are* the acoustic qualities "properties"? The question has a straightforward – but perhaps at the same time a slightly puzzling answer. The acoustic qualities are properties of the *sound*. But what is the sound apart from

its acoustic properties? Answer: Nothing. How can nothing ("something which is nothing") have properties (qualities)?

One can answer the question *where* the sound is. The sound is in the place (space) where it is audible, can be heard. This space is not necessarily identical with the space in which the sound-waves can be recorded. But it is *roughly the same* space. (The waves, I presume, may be traced to below the audibility threshold.)

I feel tempted to say: The *substance* of the sound is the *space* it occupies and the time it lasts. The substance is *not* the *medium* in which the sound is propagated, for example air. The medium has of course substance too, *viz.* the space *it* occupies. And the substance of the medium may be the same as the substance of the sound. But the medium and the sound are not the same.

11.

The frightfully confused notion of substance. There is a prototype picture of substance: a solid material body. Or: solid matter. Or simply: matter. This picture, it seems, is basically *haptic* (tactile). (See sect. 7 above.) Substance is something which we can touch. It may be transparent and colourless like a piece of glass or like air. But it makes a perceptible *resistance* when we touch it or it "touches" us. We may be able to *penetrate* it – but this only means that parts (portions) of it are pushed away or pushed apart in space. *Matter is impenetrable* we have learnt. Two different "items of matter", material objects, cannot occupy the same space at the same time. But this is only a grammatical remark about a picture which we entertain.

12.

The substance of a perceptible individual existence is, one could say, the "*it*" to which perceptible qualities are attributed (or which *has* perceptible qualities).

One could say that the "ontological status" of this "it" is the "*togetherness*" of some qualities in space and time, *i.e.* their occurrence in a certain (not necessarily sharply bounded) place in space and time. Or one could say that this "it", the substance, is that *location* itself in space and time. Or one could say that, as a "thing", the substance is a "bundle of qualities" in space and time. But one could also say that this "it" has no ontological status at all, that it is a *grammatical* and not an *ontological* category. It is the (grammatical) *subject* of a sentence, of which the name of a quality is the (grammatical) predicate. The predicate names something which has "ontological status", *viz.* a quality.

It is on purpose that I have repeatedly used the phrase "one can say". One can say all these things and find them illuminating of the notion of (perceptible, physical) substance. This notion is a "philosopher's creation". It is an idea

which forces itself upon us when we philosophize. At the same time it seems inevitably to lead our thinking astray, to call forth misleading pictures.

13.

We must learn to talk about these matters in a completely relaxed way, to put up with their inherent ambiguities and obscurities and not try to create an artificial order in this conceptual chaos by introducing suitable "definitions". Only then can we attain the clarity for which we are striving.

14.

"The box is red." "The box" is the grammatical subject. Does it mean that the box ("itself") is the substance of which redness is a quality? One could say this. But remember that then the substance is a material thing.

"It is a box." This could be the reply to a question "What is this?". "It" is the grammatical subject – but what is the *it* of which "boxiness" is being predicated? A material thing, the box? Are we saying, then, that the box is a box? Or that a "bundle of qualities", optic and tactile, occupy ("come together" in) a certain spatio-temporal location? If the first, the substance as the bearer of qualities is thought of as a persisting material object; if the second, the substance is an "empty" region in space and time. Is the difference important? In the case of the box it is not. But in the case of a sound it is. Because if substance is material then the "substance" of the sound would have to be the medium in which sound is "propagated" and the acoustic qualities would be qualities of the medium, say air. And this, I have tried to argue, is a mistake.

15.

I started by asking the question What are physical phenomena? What I call percepts hold a pivotal position among them. Some percepts are material bodies, others (gasses) are material but not what we ordinarily call bodies, others still are not even material. Percepts appear in regions in space and time and they are bearers of perceptible or sensible qualities. I have called them "a togetherness of qualities in space and time". The insight embodied in this phrase, however, is easily confused or mystified by conceptual pictures or models which we make ourselves of its meaning.

16.

One who maintains, as a *philosophical thesis*, that percepts are "the togetherness of qualities in space and time", or that they are (nothing but) "bundles of qualities" is apt to mislead others – and probably also himself. And yet there

may be important insights implicit in what he maintains. Some such insights I have tried to make explicit. But they cannot be captured, without risk of misunderstanding, in one short formulation or "thesis".

Physical things and other phenomena are something objective, real. This is true also of those physical phenomena which are *not* percepts. But among physical phenomena percepts hold a pivotal position in the sense that it is only "through them" that we have access to, or know of, those physical phenomena which are not themselves percepts.

II

1.

I have wanted to say that the *primary physical phenomena* are *percepts*, *i.e.* that which we see, hear, taste, *etc.* Also: that the *primary material or physical things* are "a togetherness of qualities in space and time".

Suggestion: the *primary physical world* is the totality of percepts. But then it must be remembered that different subjects have different perceptual capacities. There are blind, colour-blind, and deaf people. A sound is there "to be heard". Some people happen to hear it, others not. Some people *can* hear it (could have heard it), others not. And it still may be *there* also when nobody hears it. It is essential to the notion of a physical phenomenon that we should admit this last as a possibility.

2.

What about sound which no human subject can hear, although there are animals which can hear it? There are familiar cases of this. I had a friend[1] in my youth who investigated birds with this capacity. By producing a conditioned reflex in the test animal he could decide whether the bird heard a sound which *he* could not hear. That is: wave movements in the air which, when affecting his ears did not call forth an acoustic sensation nevertheless made the bird react in a noticeable way. We are inclined to say that the bird *heard* the sound. That the sound was there "to be heard". That the bird was aware of an "acoustic quality" which is not perceptible *to us*.

But was what the bird perceived (registered, reacted to) "qualitatively like" a *sound*? We are tempted to ask something like this – but the meaning of the question is obscure.

Suppose somebody were to suggest: The bird did not *hear* anything at all. But its nervous system was affected by sound-waves of a frequency which the human ear cannot hear and which caused the reaction to which the bird had been conditioned. The conditioning in the experiments to which I am referring

took place as follows: the bird was given a light electric shock immediately after a whistling sound had been produced and as a result of the shock the bird got "startled". After some trials, the bird became startled when the whistle was sounded. The experimenter could then find out whether the bird could "hear" sounds of a frequency which the human ear could not "hear". But was the "hearing" of the bird anything over and above the reaction to the stimulus (it becoming startled)? I think we must say that it is conceivable that it was nothing more than this. If this *was* the case then there is no *sound* which the bird but not the man can perceive. But this, although possible, is not necessarily so.

3.

An animal reacts to light and sound. Does it follow that it sees and hears? It is not clear what the answer is. It depends upon how like or unlike us the animal is – for example whether it has sense-organs even remotely like eyes and ears. A bird is, in this regard, "sufficiently like us". But what about a worm?

What is it to react to light and sound? Even this is not clear in itself. *Primarily* it means reacting to optic and acoustic stimuli which are perceptible to (normal) humans. Science has taught us to correlate light of different colours and sound of varying pitch with ranges of frequencies of "waves" (of certain kinds). "Waves" with frequencies beyond these ranges cannot be perceived as light or sound. We know, however, that there are such waves and, in many cases, also how to produce them. It is therefore possible to correlate these waves too with optic and acoustic qualities. These qualities are not perceptible to (ordinary) humans, but we can imagine beings with a sensory apparatus "good enough" to enable those beings to see and hear them. We can also imagine – indeed we know of – beings who react in various ways to those to us "invisible and inaudible" waves. Then it is natural to say, for example, that those beings see ultraviolet light or hear supersonic sound. This is a *secondary* meaning of saying that a being reacts to light and sound.

This secondary meaning (use) we can now disentangle from the correlation with qualities. We can say: Whether or not those beings perceive the waves as optic or acoustic phenomena, they react to light and sound also outside the ranges in which the waves are perceptible to humans. This mode of speaking seems logically acceptable. There is no objection to applying it also to inanimate beings, artefacts and machines. Such application has become common in our "electronic age". It *can*, but *need not*, lead to obscurities in our thinking.

Suppose now that we tried to bypass altogether talk about (perceptual) light and sound. We say: The animal reacts to light- and sound-*waves* in the ranges of "visible" light and "audible" sound in characteristic ways which make us say

that it sees and hears *and* also to waves outside these ranges in ways which are sufficiently analogous to the earlier ones to make us still say that it sees and hears something *we* cannot see or hear.

Saying this makes sense. But have we therewith admitted as a possibility that the animal which reacts in those ways to light- and sound-waves has no visual or acoustic sensations, only characteristic behavioural reactions? Would not admitting this amount to entertaining a "Cartesian" view of animals as machines? Moreover: if this were not merely a possibility but also true, then no optic or acoustic qualities would exist outside the ranges of (to us) visible light and audible sound. What would decide whether this is so or not?

Let us go a step further. Humans react to light- and sound-waves in a certain range in characteristic ways which make us say that they see and hear. We cannot deny the existence of optic and acoustic perceptions, *i.e.* perceptions of light and sound. But to what is it that humans *react*? To the waves or to the sensations?

4.

We would not see light and hear sounds unless light waves and sound-waves affected our eyes and ears, triggered nervous processes in them which are propagated to the seeing- and hearing-centres in the brain. And without this affectation of the sense organs there would not be any reactions on the perceiving subject's part to light and sound either.

But are the perceptions essential to the reactions? We describe the reactions as reactions to acoustic or optic sensations. For example: We turned round because we heard a sound; we stopped because we saw the red light appear. Similar descriptions we are inclined to apply also to animal reactions to "unperceivable" light and sound. But our suggestion was that we might be able to describe the animal reactions without referring to animal sensations. Is this so? – and, if so, why not also human reactions?

To *what* does a person react who is startled by a sound? We have already said it: he reacts to the sound. But what *is* this? He jerked, his body "jumped". The jerking was caused by outgoing impulses from his brain which in turn were caused by ingoing impulses which had been caused ("released") by sound-waves entering the ear and affecting the cochlear nerve. But what we then describe is how *his body* reacted to the sound-*waves*. In this description the sound perception, his hearing the sound, does not enter ("as a link").

Can we distinguish between his *body's* reaction to the sound-*waves* and *his* (the person's) reaction to the *sound*? Yes, for example by describing his body's reactions as ("mere") *movements* and his reactions as (intentional) *actions*. In the first case the bodily movements are described as *caused* by (the processes started by) the sound-waves. In the second case the sound was a *reason* for the

person's actions. And the sound being a reason requires, it would seem, that the sound was *heard* by the person. In this case the perception enters ("as a link") in the description, is essential to it. To the perception there is an answering percept, the heard sound, a physical phenomenon.

5.

The jerk is a reflex, not an action. I can explain it by making reference to the sound (which I heard). But this is to give a cause, not a reason. And here it is possible to say that it was not my hearing the sound which caused my body to jerk, but that it was the sound itself (the sound-waves) which did it.

Might it not happen that I jerk in response to sound-waves even though I do not *hear* the sound? I am perhaps so "engaged" in doing something that I do not notice the sound. Don't say that I "heard it subconsciously"! I just did not *hear* it. But it nevertheless affected me. And what about deaf people? Can a sound startle them even though they do not hear it? It would be interesting to know the answer.

Suppose that I react to a sound by turning my head (to look) in the direction from where it came. This would (normally at least) be an action for which I can give a reason. "I turned my head because I heard a sound." This is *not* a causal explanation. "My head turned because my ears were affected by a sound (sound-waves)." This is a ("first sketch" of a) causal explanation of my head's turning. "My head turned because I heard a sound." "I turned my head because my ears were affected by sound-waves." These are "logically confused" modes of expression. One could understand them either as "ill-formed" rational explanations of why I turned my head, – or as "ill-formed" causal explanations of why my head turned.

Is it *irrelevant* (not essential) to the causal explanation whether or not there is an acoustic perception? Is it *irrelevant* (not essential) to the rational explanation whether or not hearing is produced by sound-waves affecting the ears?

I am inclined to think that the answer to both questions is affirmative.

We further ask: Is it *essential* to the rational explanation that there be an acoustic perception? Is it *essential* to the causal explanation that there are sound-waves affecting the ear?

I think the answer to the first question is Yes – with one reservation. I must have had an acoustic *sensation* in order correctly to explain, making reference to a sound, why I turned my head. But the sensation need not have been the *perception* of a sound from an outer source. It might, for example, have been caused by some process internal to the sense organ of hearing. If this becomes known to me later, I could say that I turned my head because I *seemed to hear* a sound (from that direction).

Also the answer to the second question has, I think, an "affirmative core". The undulatory nature of sound cannot be a conceptual necessity. But it may be a necessity that sound, in order to be perceived, must be transmitted through space to immediate contact (contiguity) with the sensory organ and that this transmission must have the form of changes in a material medium (for example, air) and that the changes must consist in displacements (movements) of some material particles composing the medium. When I say that this "may be a necessity" I mean that it may be a *conceptual* requirement on *causal explanations* of bodily reactions to sound (and, *mutatis mutandis*, to light).

NOTE

[1] Olof Granit, a cousin of the famous Nobel laureate physiologist, Ragnar Granit. O.G. fell in Finland's Winter War. The results of his research were published posthumously by his teacher, Professor Pontus Palmgren, under the title "Beiträge zur Kenntnis des Gehörsinns der Vögel" in *Ornis Fennica* XVIII, 1941.

AN ESSAY ON DOOR-KNOCKING

I

1.

I hear a knock on the door, rise and walk to the door, seize the handle, turn it, and push or pull the door open.

I think this is a *good* example. It describes a typical action situation. A special virtue of the example is that it also includes mention of how the agent *came to have* the reason on which he acted.

What was the reason then? The knock? My hearing the knock? The question: Why did you open the door? could be answered equally well: "There was a knock on the door" or "I heard a knock on the door" or "Somebody was knocking". It is understood that if there was a knock I heard it – or else it could not have been a reason for my action.

But suppose nobody was knocking. Perhaps somebody just happened to hit the door. Or I had an illusion. Becoming aware of this, I must modify the statement of my reason and say: It *seemed to me* that there was a knock or I *thought* somebody was knocking. But at the time of my action I need not have had any "thought" of that sort.

2.

What is a knock? A sound produced by someone with the intention of being let in. A knock is "an intentional phenomenon".

What is a sound? A sound is something objective, "purely physical", "out there" in the world. Somebody hears it, another not. Or nobody hears it.

3.

But how can hearing a knock or a sound, a "mere" perception or sensation, be a *reason* for doing something?

It can be this only by "entering the understanding", by being *understood* to be (understood as) a reason for action. What is this?

An agent understands a knock as a reason for opening a door if, and only if, on hearing a knock he, normally, proceeds to opening the door unless he is physically prevented or has some stronger reason against doing this.

G. H. von Wright, In the Shadow of Descartes. Essays in the Philosophy of Mind, 83–96.

This essay was first published in the German periodical *Rechtstheorie* **19**, 1988.

This statement requires several comments.

The word "normally" is meant to say that the response need not follow invariably even in cases when the agent is not prevented and has no reason against the action. But if, under such circumstances, the agent frequently fails to respond it becomes doubtful whether he understands the meaning of door-knocks.

If the agent is prevented, for example cannot get up from the chair, he can *excuse* himself for not responding by referring to the fact that he was prevented as a *reason* why he did not respond. This sort of reason, however, is not a reason *against the action.*

The clause "unless physically prevented" could also be omitted by including it under the meaning of "normally". If the agent is physically prevented the situation is not "normal" – one could say.

Assume that we know that the circumstances are normal and that the agent heard the knock and understands its meaning but did not respond. Then we may wonder *why* he did not respond. And there will normally be an explanation for this, giving a reason for the agent's passivity. I say "normally" – because it sometimes happens that the agent "for no particular reason" remains passive. If such cases were frequent we may come to doubt his hearing or his understanding, or suspect that there are "hidden reasons".

What has been said amounts to a "logical connection argument" of a sort. It is part of the *logic of the situation* here that response follows if things are as we have described them. Whether response is successful, *i.e.* the responding action completed, is another matter. If the agent fails in the performance – no reason for interrupting it having cropped up before completion – we would describe what he did in terms of *trying*.

If I respond to a knock by opening a door it does not follow that I want to let the knocking person in although *he* wants to be let in. I may just be curious to know who is there. Or wanting to tell the knocker that I have no time to see him now. Or I may have responded without having any such "mental attitude" at all.

Suppose we explain why a person opened a door by saying that he heard a knock. Or that he himself explains his action thus. What are the truth-grounds of the statement that he opened the door because he heard a knock? They are the following three, it seems: First that he heard a knock. Second that he understands the meaning of knocks on doors. Third that he did not do what he did for some other reason or for no reason at all. One might argue that the second truth-ground is included in the first because the acoustic perception of something *as a knock* presupposes that one understands the meaning of knocks as reasons for action. But one may also keep the two grounds separate.

The first two grounds establish that the person *had* a reason for doing what in fact he did. But they do not establish that he did not have some other reason

as well. If he had some other reason we must establish that he did not do the action because of *it*. The case when he acted for no reason at all we can lay aside as "marginal". We can therefore say that, assuming that the first two grounds hold true, then, in the absence of reasons against, the suggested explanation of the agent's action is (almost certainly) valid. This is how we *understand* his action: as done because of (performed for) that reason. Nothing further is needed. The words "almost certainly" should cater for the exeptional possibility that he acted for no reason at all, although he had one.

4.

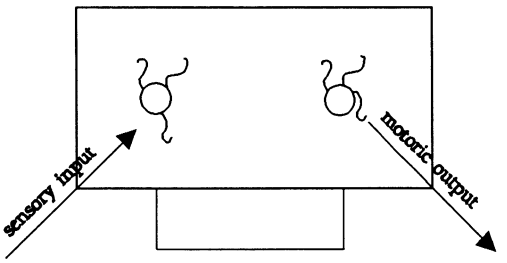
We shall later return to the topic of "other reasons" and "reason against". But first we must make an "excursion into neurophysiology".

That I heard a knock presupposes that vibrations of a door were produced by somebody hammering on the door, thus calling forth waves in the air which were propagated to and entered my ears affecting in them the auditory nerves, *i.e.* "releasing" in the nerves physico-chemical (electro-chemical) processes propagated along the nerves to the auditory centre in my brain.

That I got up from my chair, advanced to the door, and opened it presupposes that in the motoric centre of my brain originated physico-chemical (electro-chemical) processes propagated along motoric nerves to muscles in my body causing the muscles to contract and relax thus making my body and some bodily organs go through a complex pattern of movements.

These descriptions are very rough and oversimplified. Thus, for example, the nervous impulses "steering" my bodily movements are constantly modified by "signals" from outside my body emanating from seen or touched objects which I encounter in the course of performing the action.

We can draw a picture illustrating the above story as follows:

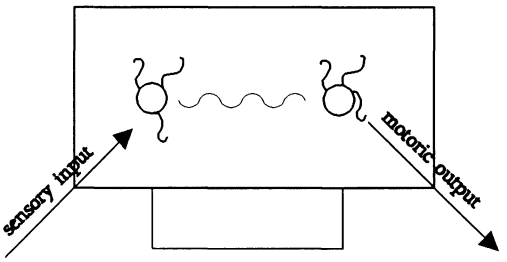


It is tempting, not to say compelling, to think that there is a *causal relation* between input and output. *Not* in the sense that the input causes the entire output, however. A knock on the door cannot possibly (alone) cause the complex behaviour of opening it in every detail – which it would do if it caused the entire complex of neural events constituting the output. But it may

have started the motoric process, "released" the nervous processes which were then in various ways reinforced and readjusted in the course of the performance of the action.

An affectation of the sensory centre in the brain by a knock or an equivalent stimulus is not innately connected with an out-going response of the kind which initiates movements answering to an action of door-opening. Just as the agent has to learn, get to understand how adequately to respond to the sound of a knock, similarly his brain has to "learn" to respond with the adequate motoric output to the sensory input effected by the knock. Wherein does this learning of the brain consist? And how is it related to the learning of the agent?

One can make a picture of what happens in the brain as follows: The incoming sensory impulses are "transferred" from the sensoric to the motoric centre through the region separating the two. The effect of the learning process on the brain is that this transference, which does not take place "innately", now happens. We can imagine that connections between initially unconnected nerve cells have become established, thus linking the two centres. The situation, learning having taken place, is like this:



In the picture there is a connection between the sensorial and the motoric centre in the brain. Perhaps one can also imagine that the effect of learning was some kind of change in the "total" state of the brain (or part of the brain) which, once achieved, enabled the sensory input to effect the motoric output. (But it is not easy to think of such a change if it cannot be localized to a region *linking* the input to the output.)

Common to both pictures which we suggested of what it may mean that the brain learns the adequate motoric response to the sensory stimulus is that they presuppose *structural changes* in the brain. The brain of the agent before he has learnt to understand the meaning of knocks as reasons for action and his brain after he has acquired this understanding exhibit some (in principle) verifiable difference in structure. The question is: Is it a demand of "scientific intelligibility" that it *must* be so?

Could we not also think about the matter in the following way: When the agent has learnt the adequate motoric response to the sensory stimulus and then, in an individual case, upon hearing the sound proceeds to opening the door

there is indeed a "transfer" from the sensory to the motoric centre and this transfer takes place in a regular way which can be observed and described – but there are no structural changes in the brain which enable the transfer to take place now when the learning process is completed? This would mean that the brain has got a new *functional* capacity, *viz.*, that of releasing a motoric reaction in response to a sensory stimulus. This capacity it did not have before but has now acquired, without any structural alterations of a more permanent nature having taken place.

Let us go even one step further. Upon the sensorial stimulus there follows a motoric reaction which did not occur before, *i.e.* prior to the agent's having learnt how to react to door-knocks. Now it occurs with great regularity under the normal circumstances. But this regularity, switch from sensory to motoric, is the "whole of the transfer". There are no structural changes in the brain (no new connection between nerve fibres established) which could be attributed to the learning process. Nor are there any characteristic processes which "mediate" between the sensory input and the motoric output. We can still say that the brain has acquired a new function or capacity which it did not have before, *viz.* the capacity of "switching" from the sensory to the motoric, – but this capacity consists just in the fact that the "switch" occurs with great regularity. Learning has left no traces of any kind in the brain.

With these three possibilities in mind, let us raise the question: Is the transfer from the sensorial to the motoric, which we are considering, an example of a *causal* (*nomic*) relationship; sensorial cause – motoric effect?

There is a strong temptation to speak of a causal relation when the transfer also involves structural changes on the brain. If again the transfer is just the regular sequence: sensory input – motoric output and nothing else "connecting" them or "mediating" between them, the relation between the two neurological terms seems no different in nature from the relation between the sound understood as a reason for an action and the initiation of the action. If the former is causal, the second presumably is causal too. But is this so? Let us inspect the case a little more in detail.

5.

In our description of the regularity on the neurophysiological level we have repeatedly made use of the phrase "normal circumstances". The same phrase occured in our previous description of the way the agent responded with an action to a stimulus. There the phrase meant that the agent who heard the knock on the door and understood this as a reason for action proceeded to that action (of opening the door) unless he had some stronger reason against performing it. His reaction to the knock was a sign (criterion) that he understands the

"meaning" of door-knocks. There is a *logical connection* of a kind between his hearing the knock and proceeding to the action. (Cf. above p. 84.)

What then are the "normal circumstances" to which reference is made in our neurophysiological parallel story? Exactly the same, it would seem. Let us assume that when the transfer takes place it happens in some regular manner which can be observed on the brain – whether or not there is also some anatomic (structural) foundation which enables it to happen. The regularity is an empirical generalization from observations. How would one test it? The subject must be in a situation in which door-opening is something he can do, and he must not know that an experiment is being performed on him. If he knew this he might change his motivation so that he refuses to react to a knock in a "normal" way. In other words: if the result of the test is to be relevant to the hypothesis that the sensoric-motoric transfer takes place in such and such a regular way, the situation of the agent, including his motivation for performing the action under consideration, must be "normal". It must be the one in which he normally reacts to the stimulus in the "adequate" way.

We can now return to our question (p. 86f.) how the learning of the brain (to respond to the sensorial input with motoric output) is related to the learning of the agent (to respond to hearing a knock by opening a door). In both cases we have a stimulus-response relation. The second is a case of acting for a reason and looks like a logical or conceptual relation in that the occurence of the response is a criterion of learning having taken place. The first is a case of regular sequence of certain neural processes and looks like a causal or nomic connection. But this picture of it as a causal relation is "blurred" by the fact that the relation is restricted to "normal circumstances" which are defined in the terms of *reasons* for or against action.

6.

After this "excursion into neurophysiology" we return to the topic of reasons for an action.

I hear a knock on the door. I have learnt and am thoroughly familiar with door-knocks as reasons for action. If this is so and if I respond to the knock by an adequate reaction what I do does not require, stand in any need of, any explanation. It is obvious why I do what I do. But if I do not respond in this way an *explanation* is needed if we are to understand my failure to react.

The explanation could be "physical" (causal). Perhaps I was lame or for some other reason unable to raise from the chair. But these cases we shall now ignore.

There may have existed a reason why I did not open the door. Some reason *against* the action. What could it have been? Perhaps I was busy doing something else and wanted not to be disturbed. Or I was too tired to see a

visitor. Or I feared an unpelasant encounter with somebody and wanted to give the impression that I was not in the room. Or maybe I had been forbidden to open if somebody knocked. In all these cases I had an intelligible (understandable) even if not necessarily acceptable *excuse* for not reacting in the normal way to the knock.

It may happen that I had one or several of these reasons against the action but that I nevertheless acted. We must distinguish between *having* a reason and *acting* for a reason, between the *existence* and the *efficacy* of a reason.

We must also distinguish between *giving* and *having* reasons. I can give as a reason for not opening the door that I had been forbidden whereas the "real" reason was that I did not wish to be disturbed. I can sincerely think that I did not open because I had been forbidden, the real reason being something else. The phenomenon called "rationalization" occurs when I give reasons which were in fact not the reasons for which I acted. The psychology of such cases can be very complex – and their clarification an important and interesting task.

If in spite of the reason which I had for the action I did not act (strictly: did not proceed to action) then the explanation for my in-action, if it has an explanation, is that I did not act because of the reason or reasons against the action which I had. A reason against was stronger than the reason for the action, or several reasons against together weighed heavier than the reason for. This judgement of relative strength or weight depends logically on what I do. Roughly speaking: the fact that I act or remain passive *shows* which reason is stronger. Only in some oblique sense can it be true that I acted although I had a stronger reason against it. This can mean, for example, that in somebody else's opinion I should not have acted. Or that, in retrospect, I regret my action.

The possibility cannot be excluded that in a situation where I have a reason for an action but do not act I remain passive *for no particular reason*. This can happen even in a case when I admit that, in addition to the reason for the action, I also had reasons against it. But it were not they which led to my in-action. If so, my in-action just has *no explanation*. This can be the case with an action, too. But such cases of inexplicable acting or omission to act are marginal. If they were frequent they would tend to upset our very notions of action and of reasons for and against an action.

7.

Let us inspect somewhat closer the various reasons which there may be against opening a door in response to a knock. They are a rather mixed bunch.

I was too tired to open the door. Someone may wish to say that this is no "reason" at all. Tiredness "overwhelmed" me; I could not bring myself to get up from the chair. Saying "I was too tired" would at least be an *intelligible*

explanation of my inaction. In this sense it was a reason. It would not have been this, however, if it were not the case that people often do not do what they have a reason for doing or ought to do because they are, as they would say, "too tired".

I wanted to be left in peace, undisturbed and therefore I did not open. Objects of want, surely, is a standard example of reasons for, or against, actions. Learning to act, one could say, is in a large measure learning how to achieve one's objects of want, coveted "ends of action". There is therefore a *conceptual connection* between wanting (to do, to get, to have) something and "going after" that thing. And similarly for the opposite of wanting which is shunning. If I want not to be disturbed by visitors then I will not let them in by opening my door in response to a knock, unless I have some overriding reason for letting some visitor in "after all". (The overriding reason could, of course, be the very fact that somebody knocked. The wish to be left in peace need not override polite behaviour.)

I had been forbidden to open. One has learned not to do forbidden things, and to anticipate some unwanted consequences for oneself if one does not observe the prohibition. This makes prohibitions reasons (against certain actions). It would hardly be right to say, however, that there is a conceptual connection between things being prohibited and things not being done. A prohibition which I habitually neglect, or one given by an authority whom I do not acknowledge as one who has a "right" to issue prohibitions to me, simply is *no reason* (motivation) for me not to do the action in question. If then in response to the question, why I did not open, I answer that I had been forbidden to do it, this would be mere pretence and not an explanation.

So much for reasons which may override the reason which a knock on the door is for letting the knocker in.

8.

We go back once more to the neurophysiological story. Tiredness is a state of the human body which, for all I can imagine, also effects the brain. The tired man is less alert, less prone to respond to various stimuli from outside which challenge him to action. The connections in his brain get "blocked" so that, for example, the transfer of the sensory input from a knock on the door to the motoric centre does not take place or is too weak to call forth (all) the movements of the action of opening the door.

Similarly, the man who wants to be undisturbed will "resist" disturbances either by taking appropriate actions or refusing to act. Can we not imagine that his preparedness to turn away disturbances has a characteristic "reflection" in the state of his nervous system such that, for example, his resistance to the

challenge of a knock on the door is reflected in the failure of the sensory input from the sound to be "transformed" into the motoric output of an action?

Awareness of a prohibition can be a reason against a certain action. Is not that like a stop-signal in the brain, something blocking a passage which would otherwise be free, – for example a sensoric input being transferred to a motoric reaction?

In summary: reasons against an action are factors inhibiting certain reactions which, had it not been for the inhibition, would otherwise have followed in response to some reason *for* doing them. Can we not think of all this also in the terms of blocking within a neural system? In other words: Could there not be an explanation in neuro-physiological terms in addition to or "parallel with" the explanation in intentionalist terms of the agent's failure to act which refers to the reason he had for not responding to the knock? This parallel explanation would tell us why the sensory stimulus was not followed by the normal motoric response. *How* would it tell us this? By pointing to some changes in the state of the brain produced by the "cropping up" and presence of those reasons which inhibited the agent's reaction to the knock. (His getting tired, his coming to want something (else), his being prohibited, *etc.*) The changes need not be structural, but they must be somehow accessible to measurement or some other kind of observation.

I cannot see why this "fantasy" could not be reality. Could not, moreover, such explanations be interesting? The answer depends on what one is interested in. It is difficult to imagine, however, that such explanations, if they existed, could be of much "practical importance".

II

1.

A person hears a knock and proceeds to opening a door. (Whether he also has reasons for not doing what he does is not now important.) We said that to his understanding of the knock as a reason for his action there answers an acquired capacity of the brain to transfer the sensory input from the knock to the motoric output of the action. If this is to mean anything over and above the (mere) fact that the person, having learnt the meaning of knocks, normally, unless he has some overriding reason against, proceeds to the action, we must assume that the learning has made some independently observable, structural or functional impact of some permanence on his brain. A structural impact could be that some connections have been established which were not there before; a functional impact that some processes in the brain can be identified as a transfer of sensory input to motoric output. To entertain some such idea seems "logically permissible". If it is *true* one way or other, it is a *contingent* fact

which has to be ascertained and established by empirical investigation. *What* the neurological equivalent *is*, is not philosophically interesting, – nor even *whether* there *is* an equivalent. The "philosophical" question is whether there *must be* an equivalent. Let us next consider the nature of this question.

Someone (a philosopher) says: Unless there is an equivalent we could not *understand* how the agent has learned to respond to knocks or we could not *explain* how teaching has had this effect. It is a demand of "scientific intelligibility" that there must be such an equivalent.

One sometimes wonders how it is that a certain person has or has not learnt a certain thing, does or does not understand this or that. And it is usually clear what answer or information will satisfy us. "Nobody taught him" or "He is too stupid" could be satisfying answers. We can be amazed that a person has learnt or got to understand something so quickly and say "I cannot understand how it was possible". And maybe no further information will remove our bewilderment.

Neurophysiology can also be relevant to cases of the type at which we just hinted. Perhaps the reason why a person cannot learn or understand something is that he suffered a disease which is known to affect the brain. Human beings thus afflicted are incapable of certain things.

But all of this understanding and not-understanding belongs in a different dimension from a philosopher's postulation of a neural correlate to mental phenomena in order to make these phenomena intelligible.

Someone says: it would be unintelligible that one can hear sounds unless acoustic sensations have a neural substrate. In response should be asked: What was it then that one did not understand before one knew something about the neurology of hearing? Did one not understand what it *is* to hear? Or how it is *possible* to hear?

What is it to understand what it "is" to hear? One has learnt a word for the phenomenon and can, in normal cases, tell whether a person is deaf or not and whether he heard a certain sound or not. One knows how to ascertain this in doubtful cases. One can also report one's own acoustic sensations. Neurology can provide us with some additional methods for deciding doubtful cases – but use of these methods presupposes that we already have a concept of hearing and can handle it in linguistic communication. Roughly speaking: Unless one already knows what it *is* to hear one cannot make discoveries about the neural basis of hearing. (Or one would not know that these discoveries are of relevance to *hearing*.)

2.

Is it or is it not right to say that what we know about the neural basis of hearing, about the physical phenomena which are the objects of auditory

judgements of perception, has made us understand better what hearing *is*? If the aswer is Yes, does it mean that before these scientific discoveries were made we did not know what hearing or sound "really was", that these discoveries as it were revealed to us the "essence", the "true nature" of those *concepts*?

We *have* these concepts and know how to handle them in linguistic communication. This is so whether or not anything is known about the "neural basis of the mental". Unless we had the concepts we should not know how to identify the phenomena falling under them, and if we could not identify them independently we could not establish their correlation with other phenomena (for example "brain events") either.[1] To acknowledge this is not to deny that knowledge of the correlations may have a modifying influence *also* on the concepts themselves.

3.

It is important to distinguish between *concept* and *phenomenon*, – between the phenomena of (a person's) hearing sounds, understanding something, having reasons for and against an action *and* the concept of hearing sounds, of understanding, or of reasons for action. In order to classify phenomena as cases of such and such we must already have the concepts "such and such".

The psycho-physical correlations are, primarily, correlations between two realms of phenomena – between things which we observe and register in the neural system and things such as (persons) hearing sounds or seeing colours or feeling pain but also believing or remembering or understanding something or intending or wanting to do this or that. In order for us to correlate the phenomena the concepts must be there and have at least so much stability that classification under them is, in most cases, unproblematic. To put it slightly differently: We must be able, by and large, to *identify* the neural phenomena and the psychological phenomena *independently*, on the basis of different criteria.

The criteria according to which we classify or identify different phenomena as falling under either neural or psychological concepts are, however, neither absolutely stable nor do they always permit univocal classificatory judgements. Therefore they are not sharply separable from one another either. It may happen that we have to rely on neurological findings or tests for deciding, say, whether the person "really" heard (could have heard) a knock – or appeal to the person's report of his sensations in order to decide whether our reading some instrument which registers brain events was reliable or was "an error of measurement". Many such cases we simply leave undecided.

The occasional appeal to criteria associated with one type of phenomena for classifying phenomena of another type may have repercussions on the concepts themselves. This may happen also with the psycho-physical correlations.

Whether there are examples of its having happened I do not know. But the thing is conceivable. For example: I can imagine that one could correlate minute shades of the same colour which usually pass unnoticed with very conspicuous differences in the neural reaction to stimulations with light of the two shades of this colour and that this fact would make us better aware of the shades (as seen), so much so that we would coin *names* for the two shades and henceforth teach people a *conceptual* distinction which did not exist before.

But even if such influence on the psychological concepts from neurological observations is possible it is certainly not important in *our* system of psychological concepts.

What holds for hearing and other perceptual and sensational concepts also holds, *mutatis mutandis*, for understanding and other cognitive notions. That persons understand the meaning of knocks as reason for opening doors means that in normal cases they react to knocks in the adequate way unless they have overruling reasons against the action. This is a *conceptual* observation and it provides us with criteria for identifying, testing, verifying that or whether a person has (acquired) the understanding in question. It gives us the criteria for the *phenomenon* of understanding.

4.

An advantage of our example of the knock on the door is that the description of the action situation includes mention of how the agent got the reason for his action. The knock gave him the reason but it also affected his nervous system and thus *literally* "put him in motion". The motivational mechanism of his action and the causal mechanism of his bodily movements in executing it were both activated by he same event in the world, the knock. Hence there is no "mystery" connected with the simultaneous operation of the two "mechanisms". If we call action *free* when performed for a reason, there can be no conflict between "freedom" and "determinism".

The agent's action for a reason and his body's reception of the reason-giving stimulus and subsequent reaction to it are two aspects of *the same* (complex) *event* in the world. One could call them, with caution, the bodily and the mental aspect of this event. *Their* "connection" is *not* contingent: the knock would not be a reason for action unless it also were perceived and the agent's proceeding to opening the door would not be an action unless his body went through the appropriate movements. The connection of the two aspects is, to this extent, conceptual and therefore logically necessary.

Is this "psycho-physical parallelism"? One could call it thus. But it is hardly that kind of parallelism which supporters in modern times of the "parallel-theory" or the "identity-theory" have had in mind. Their concern has been the *contingent* relation between acting for a reason on the one hand and a correlated

chain of events in the agent's neural system on the other hand. The nature of this chain is still, *i.e.* at the present stage of development of science, largely hypothetical – but there seems no reason to think that it will not one day be known in much greater detail. But since neither the bodily nor the mental aspect of acting for a reason can be reduced to it, it is difficult to see why *this* "parallelism" should have much interest to the philosophy of action. Its philosophical interest seems to be gone when the "reductionist illusion" associated with it has been dispelled.

III

The example of the knock should make us aware of the fact that reasons for action normally are things which an agent *gets*, which "happen" or "occur" to him. Usually they enter his "stream of life" from "the outer world" in the form of perceptions the meaning or significance of which as reasons for action he has already learnt. The stage, so to speak, has been set for them as potential reasons. This is not to deny that reasons and motives of action may "crop up" in an agent spontaneously, also without any apparent cause in things which happen to him. But it seems to me that they are exeptions rather than the rule; and I think that philosophy of action has suffered under a tendency to look for the "forces" which move us to action exclusively in the "inner life" of agents.

Those things which happen and thereby provide an agent with reasons for action may in their turn have a causal history. Their occurence may require an explanation; in any case one can ask *why* they occured.

Why did somebody knock on my door just then? Knocking is an action. Somebody wanted to see me. Why? There may be an answer to that question too, *i.e.* to the question what gave that other person his reason for coming to see me. (Perhaps it was a policeman coming to arrest me for some crime I had committed. So ultimately it was I who had given the reason for this particular knock on my door.) Just as there is no conflict between freedom and determinism in the case of my opening the door in response to the sound and my body going through certain movements caused by an affectation of my hearing nerve, similarly there is no conflict between his knocking on my door in order to be let in and his fist hammering on the door under the influence of outgoing motoric impulses from his brain.

The reason-giving event could of course also have been a "purely physical" event which prompted me to do something – say, close a window. It in turn may have been caused by something else, and so forth (perhaps far back in time). But this causal determination of the event which constituted for me a reason for an action does not make my action unfree. ("The noise was unbearable, I was *forced* to close the window." This may be true, – but in as

much as what I did was an action for which I had a reason I was free to do or not to do what in fact I did.)

Causal determination of events in nature cannot constitute a "threat" to the freedom of human action ("free will"). To think that it could is to be guilty of conceptual misunderstanding. The misunderstanding removed, the problem of human freedom appears in a new light. It is then a problem of the agent in relation to the reasons he has (gets) for his actions. A man has reasons for and against an action, or reasons for incompatible actions, or reasons stemming from his spontaneous wants and inclinations and such which his social commitments and duties give him. Acting for a reason is free action, but the complexity and multiplicity of reasons may be felt by the agent to be constraints on his freedom, tearing him in different directions, confronting him with agonizing "existential choices" or enslaving him under the practical necessities imposed by his social duties. The greatest freedom may coincide with complete unfreedom. In the contemplation of these paradoxes and puzzles we are confronted with the "real" problem of human freedom. It is a problem of *social* philosophy, rather than of epistemology or metaphysics.[2] And it falls outside the orbit of this paper.

NOTES

[1] Cf. below p. 147 about the epistemic priority of the mental in relation to the neural.

[2] Cf. my book *Freedom and Determination*. North Holland Publishing Co., 1980; *Acta Philosophica Fennica* **31**.

NOTES ON THE PHILOSOPHY OF MIND[1]

BEHAVIOUR

The notion of behaviour is related to the notions of bodily movement and muscular activity, and also to that of neural processes. A basic difference between behavioural reactions and neural processes is that whereas the first are macroscopic and manifest or overt ("visible and audible") the second are microscopic and not manifest to the senses in the same direct way as the first.

The macroscopic reactions are sometimes called *molar*, the microscopic ones *molecular* behaviour. Calling both types of bodily reaction "behaviour" seems to me misleading and should therefore be avoided.

The behavioural reactions proper consist in *movements* of parts of the body: usually of the arms (hands, fingers) and the legs and the tongue. The movements of the tongue together with such in the chest and larynx produce sound when the being in question is crying or speaking or singing. Some behaviour, however, is actually motionless – for example when it "consists" in pressing the hand against the door to keep it closed, or in some other form of bodily resistance to physical constraint which, if not resisted, would force the body or parts of it to move.

The manifest movements of parts of the body are in their turn caused by contraction and relaxation of muscles inside the body. Muscular activity again has a remoter cause in impulses "propagated" from motor centres in the brain through nerve fibres ending in the muscles.

In the case of "motionless behaviour" the muscular activity causing it is, it seems, best described as *tension* of the muscle(s).

That the muscular activity causes the manifest behavioural movements presupposes that the muscles are connected to (bones in) the moving limbs (*etc.*). By cutting these connections we could have muscular activity without corresponding manifest movements of parts of the body. In this sense the two are *logically independent* of one another. In a similar sense the neural activity and the resulting contractions and relaxations of muscles are logically independent of one another.

It is of interest to consider here those behavioural reactions which are *vocal*. A person (or animal) screams, say. We would describe his (its) behaviour as screaming. Screaming is producing a sound. A sound is a phenomenon (an event) in the physical world. It is not "bodily movement" – and in that sense "behaviour". *Producing the sound* is behaviour. This happens by means of

G. H. von Wright, In the Shadow of Descartes. Essays in the Philosophy of Mind, 97–123.

changes in the being's body: pressing a stream of air through a larynx which has been "formed" so as to cause the sound to appear.

Does the behaviour of screaming consist in the bodily changes which are responsible for the (occurence of the) sound? (Remember: the scream (as a sound) is not behaviour; screaming is.) One could say so. And one could say that my behaviour in raising my arm consists in my arm rising, that my behaviour in getting up from the chair where I am seated consists in my body changing its shape from bent to upright, and that my behaviour in walking across the room consists in my body moving over that distance.

One can say all this and be right – and at the same time something speaks against it. *My arm* rises but *I* raise my arm; my *body's shape* undergoes a change, but *I* get up from the seated to the upright position; *my body* moves through the room, but *I* walk across it. What *I* (in those cases) do is *behaviour*; the accompanying bodily movements are only its visible manifestations. Or, as I have said above, they are what the behaviour "consists" in.

Is behaviour then something "over and above" its "visible manifestations"? The answer is: one can make a *distinction* between behaviour and bodily movement. We make this distinction when we say of a living being (human or animal) that it behaves in a certain way and when we say that *its body* or some part of it *moves* in a certain way. The distinction is between the living *being* and its *body*. Is the being then something "over and above" its body? Perhaps a body plus a "soul"? If we say this, we are making a *conceptual* distinction.

How deep does this distinction cut into what we call "the animal kingdom"? A zoologist studies an amoeba in a microscope. He observes various reactions (movements) in response to some stimuli (or "spontaneously"). Shall we say that the amoeba "behaves" in a certain way? Perhaps one should not trouble one's head too much over this question.

I notice an irregularity in my heartbeat. I say to the doctor: "My heart behaves in a funny way". Is my heart then a being who "behaves"? One can say so. But one may also think it unnatural. Finding it unnatural means that we sense a conceptual difference here (between behaviour and "behaviour").

My lungs breathe in the normal way. Is this behaviour? It is certainly bodily movement. I fill my chest with air at the doctor's request in a medical examination. This is how *I* then behave.

INTENTIONALITY

Here it is useful to bring in the notion of *intentionality*.

Normally, when we say that a (living) being behaves in a certain way and do not wish to identify what we say with a statement about changes in that being's body, we think of the behaviour as intentional, as something the being *does*, and of the bodily movements as having been intentionally *performed*. I would

say that the use of the word "behaviour" in which a distinction between behaviour and mere movement is implicit is a *primary* use of the word – at least in our culture – and that other uses are analogical extensions of this primary use and would sometimes be called "metaphorical". If *I* kick somebody or something this is usually an intentional movement, *i.e.* behaviour in the primary sense, something I do. But if I am hit under the knee and *my knee* performs the movement characteristic of a kick this is "behaviour" only in a metaphorical sense.

"When the ship was about to enter the harbour, the needle on the mariner's compass began to behave in a funny way". This would be an entirely correct and acceptable use of the word "behave". I would hesitate to call it "metaphorical". This is so because there is (for us) no question of intentionality being at play here. The movements of the needle are not intentional *of the needle*.

Plants have life. We do not attribute intentionality to them. And we do not too often speak of their behaviour – except in a clearly metaphorical sense. This is obviously connected with the fact that plants *on the whole* do not *move*. The Drosera rotundifolia closes round the fly and devours it. This *resembles* animal behaviour – but it seems unnatural or even out of the question to call it intentional.

I said that we should not trouble our heads over the question whether the movements of the amoeba are behaviour. Would calling it "behaviour" then mean that we regard it as being intentional? We should not worry ourselves about this question either. Animal (also human) behaviour which we can study (explain, understand) exclusively under the aspect of (reflex-) reactions to physical stimuli – without feeling that something is left out of consideration – we do not call intentional. That is: we do not then attribute intentions to the being which behaves. Attributing intentions to it is a way of *conceptualizing* its reactions.

(It is also good to remember that "intention" and "intentionality" are rather *technical* terms!)

When movements in or of the body of a living being are understood or "seen" or described as intentional, I shall say that they are being *conceptualized under the aspect of intentionality*.[2] It seems to me a plausible suggestion that we should reserve the term "behaviour" when applied to a living being for its bodily reactions thus conceptualized. In this way we keep clear a distinction between behaviour and "mere" (bodily) movement – and also between behaviour and neural activity. Muscular activity is a borderline case. Sometimes it is behaviour: "He braced his muscles".

Here a warning is in place. "Conceptualization" must not be understood to mean, necessarily, an *interpretation* of our immediate impressions of the bodily movements of living beings. It is probably right to say that our spontaneous ("primitive") understanding of such movements usually is that they are

intentional. Sometimes, however, we are unsure and sometimes mistaken. We see a person fall in the street. Did she throw herself onto the ground or did she stumble over something on the pavement? If, however, we see somebody waving his arm in the street we would usually immediately think that he must "mean" something by his behaviour – even though we may not understand *what* his intention is.

LIVING BEINGS

The notion of intentionality which we brought in for purposes of distinguishing between "mere" or "reflex" movement and behaviour is closely linked to the notion of a living being – and therewith to the very notion of life. Only to living beings do we attribute intentions.

What is a living being? Cannot an artefact, a machine be a living being? We certainly commonly speak of them as though they were. ("The engine refuses to start", *etc., etc.*).

We need not here dig into the question what distinguishes "dead" from "living" matter, nor try to decide on which side of the demarcation line this or that organism falls – viruses, for example. Least of all need we enter into the question of the origin of life. Such questions may be of great interest to science – but it is also of importance to take note of the fact that they are not relevant to that aspect with which we are concerned here, the question what constitutes a living being.

A living being has a (material) *body*, but it is not identical with its body. Its body may "survive" it, *i.e.* still be there for some time after the being itself has died. We then call the body a corpse. The body comes into existence through a series of transmutations of matter. It is another disputed topic which does not (should not) concern us here, *i.e.* at which point in this process of transmutation we speak of it as the body of a living being. (The "abortion problem".)

One might turn the matter round and say that a living being is (for us) one the movements of which we conceptualize under the notion of intentionality. (Not *all* its movements, but some at least.) In this way the notions of intentionality and movement become defining ones in relation to a living being. Plants can be alive or dead but we do not normally conceptualize their movements (and other changes in them) under the aspect of intentionality. For this reason we may hesitate to call them living *beings*.

The notion of a living being which I here try to capture could be called a *grammatical* category (in a Wittgensteinian sense of the word "grammar").

TELEOLOGY

Intentionality is related to teleology. Behaviour, *i.e.* movements conceptualized under the aspect of intentionality, is *in many cases* aiming at an end or goal. Is

this perhaps in an extended sense true of *all* behaviour?

Some behaviour is *expressive* – for example of astonishment, delight, disgust, fear or pain. (These are in a broad sense "emotional attitudes".) Such behaviour often borders on reflex reactions. Then it is doubtful whether it is intentional. For example: I get frightened and scream. We say "I scream" and not "my body produces a screaming sound". But if, as we say, I could not help screaming then (which may be true), it is doubtful whether the scream should be attributed to *me*. I am not "responsible" for it. A reason why still it is natural to say that I did it is, I think, that screaming *can* be, indeed normally is, clearly intentional. I sometimes scream *in order to* call for help or give warning. I may even deliberate whether to scream or not. In such cases screaming is not only intentional, but is also goal-directed, teleological.

On the other hand, it should be also noted that when expressive behaviour is "reflexive" and not "intentional", as screaming or weeping may be, it retains a teleological aspect. Screaming and weeping, it seems, are biologically meaningful or purposeful reactions. They serve, say, to avert a threat or to relieve tension. But they need not be undertaken "for the sake of" such an end.

An interesting case is presented by laughter. Only man has this capacity, it is said. Laughter can, usually, be suppressed. In this it resembles intentional behaviour. But it is seldom, if ever, undertaken with a view to an end. There is nothing obviously purposeful or teleological about it.

The answer to the question raised at the beginning of this section is thus a qualified No.

MEANING

The notions of intentionality and teleology are related to that of meaning. This also holds for behaviour when understood as intentional movement of bodily organs.

(The purposeful we spontaneously call "meaningful", the intended end we say is "meant".)

Also of (intentional) behaviour other than verbal it is natural to say that the being in question *means* something with its movements. It is, for example, reaching out for something, say food; "seeking food" is then what its movements mean. Saying that its movements *have this meaning* is simply a shorter way of saying that it (the being) *means* this with those movements. Both locutions are connected with the danger of being misleading: the first because the bodily movements by themselves do not mean anything, and the second because the being need not "think", be conscious of what it is doing.

Something similar holds for behaviour, other than verbal, which is *expressive*. His groaning, the contortions of his face, *mean* that he is in pain. His smile perhaps meant that he was amused (at somebody's remark, say).

One also says that his groaning is a *sign* that he has pains, his smile a *sign* of amusement. This suggests a separation between the *sign* and the *signified*. Here the sign is something physical, the signified (its "meaning") something mental, psychical.[3]

A groan or a cry is a sound. We do not perceive it as movement (waves in the air). Nor do we usually perceive those bodily movements (in the larynx for example) whereby it is produced. One could, however, say that it is those bodily movements which, in a primary sense, are signs or mean that the being is, say, in pain. There is nothing unnatural about saying this – and there is no conflict between calling the sound produced and also the bodily events which produce it the meaning-carrying sign.

A being's hands and arms move in a certain way in contact with a window (which then opens). We conceptualize (understand) those movements as an action of window-opening. This is what *the movements* mean. But we may also ask: Why did he open the window? An answer could be: He meant to ventilate the room. This was the *purpose* of his action, what he was aiming at. Talk of purpose and aim is more natural here than talk of "meaning". But we must not be pedantic with language. It is the conceptual distinctions which matter. We could also say that the purpose of his movements was to open the window, that the movements were aiming at this. The conceptualization of the movements is now in the terms of purpose and aim. Without some such conceptualization "under the aspect of intentionality" the movements are merely movements. As such they may have causes and also effects. They are "purely physical".

VERBAL BEHAVIOUR

Speech acts are intentional (behaviour). When I say that, unintentionally, I said this or that, "unintentional" usually refers, not to *the act of saying* something but to some not foreseen or intended *effect* which it had.

Is verbal behaviour goal-directed behaviour? Does one always say something *in order to* achieve something? Not in an ordinary sense of "in order to". Could one say that whenever I say something I say it, at least, in order to be understood? This may be true of all speech acts which have a communicative character. But not all speech acts are communicative. (Talking to oneself, for example.) So the answer to our question is No.

Language when used intentionally normally *means* something (has meaning, a semantic dimension). One can, however, also talk (complete) nonsense. Sometimes, one does this for a reason, something is "meant" by it.

In the case of verbal behaviour, the bodily movements produce sounds or signs (on a sheet of paper, say). When the sounds and signs are conceptualized under the aspect of intentionality we understand them as being *words* and *sentences* of a language. The agent who produced them *said* or *wrote* those

words and sentences. The sounds and signs produced by his bodily movements were meant to be words and sentences.

Word-like sounds may be produced "involuntarily", "unintentionally", "automatically". Moreover: sounds may be more or less articulate or inarticulate, *i.e.* it may not be clear which words and sentences, if any, they are meant to express (be); and the same holds for scratches on paper.

The relation of verbal behaviour to meaning is complex. One must distinguish between what *the agent means* by some words and sentences and what those *words and sentences mean*. The person who says "I am in pain" does not necessarily mean to say that, at the time of saying this, he suffers pain, although this is what his words mean. He may just have uttered the words for no purpose at all or for some quite different purpose from announcing his suffering. (In acting a rôle in a play, for example.)

The meaning which words and sentences have "by themselves" can be called their *conventional* or *lexical* meaning. It is explained in dictionaries and translated into words and sentences of another language which are assumed to have the same conventional meaning.

The relation between words and sentences and their conventional meaning in a language is, I should say, the prototype for what we understand by a "semantic relation". It is the semantic relation *par exellance*.

To clarify the nature of this relation is *the* main problem of a philosophy of language. Throughout the history of thought there has persisted a tendency to conceive of it as a relation between sign and the "thing" signified or meant, as something like the relation between a (proper) *name* and its *bearer*. It may be easy enough to realize that "meanings" are not the bearers of linguistic labels. But it is anything but easy not to be misled, again and again, by false analogies when philosophizing about meaning.

BRAIN AND BEHAVIOUR

My body and parts of my body can be moved, pushed and pulled, by external forces acting upon it. Such *moving* of my body we do not normally speak of as bodily movement, still less as behaviour. Also genuine bodily movement (whether intentional or not) requires that it should have (an immediate) cause in something which happens in the *neural system* (of my body).[4] The same holds for the higher animals. When we descend to the lower forms in the animal kingdom their neural system becomes progressively unlike ours. It is at least a plausible conjecture that unless in a living being we can recognize something sufficiently *analogous* to our neural system we do not conceptualize movements within or of its body under the aspect of intentionality. But the word "sufficiently" should here remind us of the relativity and vagueness of "the limits of intentionality".

There is a certain analogy between the brain (nervous system) in relation to bodily movement and the agent (person, subject) in relation to behaviour (action). If behaviour is movement conceptualized under the aspect of intentionality, could one then not say that the subject itself is the neural processes conceptualized under that same aspect? One *may* say so, but only with the greatest caution. It may be useful as a rebuttal of ideas of the mind, the soul, the I, as some substance which exists independently of the body and yet in some shadowy manner is quasi-corporeal. (Cf. Wittgenstein, *Philosophische Untersuchungen* §36: "Wo unsere Sprache uns einen Körper vermuten lässt, und kein Körper ist, dort, möchten wir sagen, sei ein *Geist*.")

We would, normally, not call a being's reaction to a sound intentional (behaviour) unless we think that the being in question *heard* the sound. In marginal cases its reaction may be to sensed, though not "heard", vibrations – and in still other marginal cases the reaction may take place at the "subconscious" level. In cases of this last kind we may hesitate whether to call the reaction "intentional" or "reflex". In all the cases, however, there is an affectation of a sense organ by the (physical) sound. The sense organ is a peripheral part of the nervous system. Its affectation by the sound is an "input" in the system. The input is, somehow, "propagated" towards the centre, where it "releases" ("calls forth") an outgoing reaction which is causally responsible for the bodily movements which constitute the being's reaction to the sound.

The nervous system thus *mediates* between a stimulus and a response to it which we call intentional (behaviour). The stimulus may come from outside (as in the case of a sound) or from inside the body (as in the case of stomach pain) or from inside the nervous system itself. In this last case one can ask whether the stimulus must have a more remote cause outside the system, or whether "spontaneous" activity in it can cause bodily movements which constitute (intentional) behaviour. I do not know the answer.

This chain: input – nervous system – output, may be viewed as a chain of purely physical events, "drained of intentionality". The description of it which is easiest to give is, however, usually in intentional terms. "He heard a sound and turned his head looking in the direction from where the sound came." When "drained of intentionality", the description becomes: "Movements in the air affected his cochlear nerve, called forth such and such processes (movements) in his nervous system, causing his head to turn so that it faced the direction from where the air-waves had emanated". This description is highly provisional and incomplete. It is not even certain that we in practice can give it in full. But we feel – rightly in my opinion – that there is in principle a description of the chain of physical events which is sufficiently detailed to make us understand the way in which the successive phases (terms) in it are causally linked to one another. In order to have this effect on our understanding, the description would have to be heavily "theory-loaded". The mere succession:

input – nervous processes – output does not tell us that the output (the bodily movements) is the "outcome" of the input which has been "sieved" through the nervous system as medium.

There is a sound and a consequent turning of the head. Is this a *causal* relation? There is a tap to the knee and a consequent kicking movement of the leg. This, presumably, *is* a causal relation. (The patellar reflex.) That the latter is a reflex is related to the fact that the reaction to the stimulus is not a *learned* reaction. A conditioned reflex, however, is a learned reaction. Could it be that turning the head in response to a sound is, if not an unconditional, a conditioned reflex? Is it true to say that all learned reactions are, *as bodily movements*, conditioned reflex reactions – as so many "behaviourists" have thought?

Perhaps we cannot and shall never be able to give a causal account of the successive events in this chain "in minutest detail" (which would mean in terms of microphysical processes). But I fail to see that this is not "logically possible". We can, without contradiction, entertain the *fiction* of its possibility. Entertaining it is perhaps of great heuristic value. Surely we can investigate nervous processes under a purely causal aspect, *i.e.* without conceptualizing the object of study under the aspect of intentionality.

We should take care not to tie what has been said above to some restricted (Cartesian or Humean, push-pull) ideas about causation. Causation in the brain may have a "global" or "integrative" character. Whether event E causes event E' to take place may depend on the "total state" of the brain at the time when E occurs. It may also be the case that causation in micro-dimensions has peculiarities which cannot be captured by ideas about cause and effect on the macro-level. Also: "strict determinism", whatever it means, may not prevail in the brain. (Spontaneous activity, for example.)

In spite of these limitations and restrictions one can take an overall view to the effect that what happens in the brain is, together with physical stimuli working on the nervous system "causally responsible", as we say, for all those bodily movements which, under the aspect of intentionality, constitute behaviour.

To summarize: in the material (physical) world there is a stimulus S which activates neural processes N which result in a response M. To say that S activates N or that N results in M are both *causal* statements. Whether *all* phases in the chain S-N-M are causally related may be left open. This is also true of the more precise nature of the causal relations of "activating" and "resulting in". Further: the stimulus S may have an intra-neural origin; if this cannot be traced to an extra-neural cause, we would have to acknowledge spontaneous activity in the neural system. M, finally, is movements of limbs and other parts of the being in question's body. Also bodily (muscular) tension belongs here.

STIMULUS AND RESPONSE

Let us go back to the example where *S* is a sound and *M* a turning of the head. We say: he heard the sound and turned his head. Then we have conceptualized the situation under the aspect of intentionality. "Hearing" is a mental event (a sensation). "Turning" is (here) an action, something *he*, the hearing person, not his body, does.

How do we know that the person heard the sound? "Since he turned his head, he must have heard it". Normally, this would be a reply which satisfies us.

How do we know that turning the head was something the person *did*? Could it not have been a reflex – perhaps released by some different stimulation of his nervous system? "Since there was a sound and turning towards it is an understandable (meaningful) reaction to hearing a sound, we take the movement of his head as being (the result of) his action of turning the head."

Taking the movement of his head to be an action of turning it thus is a "criterion" of his having heard the sound – and his having heard the sound makes us think of the movement of his head as an action of his. This is of course circular, but not in an obnoxious way. The intentionalist conceptualization is of the whole *context*, of *S* and *M* at the same time.

M is now no longer "mere movement" – but the movements understood as an action of an agent (a person). When understanding the movements to "mean" an intentional action we, as it were, "ascend" or "leap" from the world of matter to the world of mind ("the spirit").

Similarly, *S* is no longer the soundwaves affecting the subject's hearing organs – but this stimulus understood as a sensation which the person has. Here, too, a leap takes place from the physical (matter) to the psychic (mind).

The terms "stimulus" and "response" are still good at the mental level but in order to distinguish them as something mental from their physical "counterparts" I shall denote their new sense by *S'* and *M'* respectively.

The relation between *S* and *M* is causal we said. *S* through the intermediary of the (working of the) neural system *N* causes *M* to appear. The relation between *S'* and *M'* can be called causal too. "He turned his head because he heard a sound." True – but we also say that his hearing the sound was the *reason* why he turned his head. The reason and that for which it is a reason we can distinguish as *ground* and *consequence*. The relation of ground and consequence is analogous to and yet different from the relation between cause and effect when it holds between events (processes) in the material world. (We shall not stop here to consider wherein this difference consists.)[5]

We can be in doubt whether the subject *heard* the sound after all. "Did you hear it?" we ask him. He says: "Yes". "So that was what made you turn your head?" He can answer "Yes" or "No". If the latter, we look for another explanation of why he turned his head – or why his head turned. This we

would also have to do if he had said "No" in reply to our first question, denying that he had heard the sound. – And whatever answer he gives to our questions there is always a possibility that he is not speaking the truth.

Are also the nervous processes criteria for somebody's hearing and doing something? Normally they do not have this function. Long before one knew anything about nervous processes and had access to them by observation, one could say confidently of a person that he heard a sound and, because of this, turned his head in a certain direction. But it is conceivable that we are in doubt whether the person really heard (could have heard) a sound and that we come to think, because of some neurological finding, that he did not hear it. Then we would have to reconstruct our initial explanation of why his head turned. Also neural events can be among the criteria which we conceptualize "under the aspect of intentionality". But such cases are surely marginal.

MIND AND MATTER – THEIR CONCEPTUAL RÔLES REVERSED

There was a sound and a head turned – and there were processes in a nervous system somehow linking the two phenomena causally. Assume that I heard the sound and saw the head turning. My seeing and hearing this is reflected in my bodily reactions (including the processes in my nervous system). But I did not conceptualize *them* when I said what I heard and saw. They were not *my* criteria of seeing and hearing. I just "saw something and heard something". But if somebody else said of me that I heard and saw this or that, *he* would be conceptualizing my bodily reactions. They would be *his* criteria for *my* hearing and seeing. And they may be deceptive.

Thus, it looks as if my hearing and seeing something are the *criteria* on the basis of which I judge, come to know, that there was a sound and that a head turned. Here the rôles of the mental and the physical are reversed. In the example we discussed earlier some events connected with a being's body, bodily "affectations and movements", were the criteria for attributing to this being the mental phenomena of having a sensation (S') and acting intentionally (M'). But now, we say, some mental phenomena, sensations, are in fact the criteria for judging or coming to know that those physical phenomena, S and M, occur.

The attribution of mental phenomena to a person depends on a conceptualization of some physical phenomena under the aspect of intentionality. Similarly, one could say that the attribution of qualities to physical phenomena requires a conceptualization of some mental phenomena *under an aspect of materiality*. Just as bodily movements are "signs" of mental phenomena, sensations are "signs" of physical phenomena. Mental things and events thus have *behavioural* criteria and material things and events *sensational* criteria.

This double relationship between "mind" and "matter" seems to me

remarkable. It is easy to misunderstand it. One could say, exaggerating a little, that the history of philosophy after Descartes is to a great extent the history of these misunderstandings. They traditionally take the form of "reductions" or "false identifications". There are two main types of such misunderstanding. There is the *materialist* misunderstanding, which reduces the mental to the physical. A modern variant of this is "classical" *behaviourism*. And there is what I propose to call the *idealist* misunderstanding, which reduces the physical to the mental. A variant of this is "classical" *phenomenalism*.

Both behaviourism and phenomenalism have traditionally defended themselves against such charges of "false identification". An interesting case in point is offered by early logical positivism, which had to defend itself both against materialist charges of idealism – and against idealist charges of materialism. The defended position usually took the form of a "neutral stuff monism", like the empiriocriticist position of Mach and Avenarius or Russell's position in *The Analysis of Mind*.

Does not rebutting the misunderstandings leave us with a *dualist* position, with some kind of revived Cartesianism? There is the material world and the world of the mental (consciousness, "thought"), and the two exist, irreducibly, "in their own right".

This may be said. But a dualist position, too, invites misunderstanding. There is a temptation to "substantialize the mind" by analogy with matter. To attribute to the mind a kind of "shadowy existence" as an immaterial and yet *somehow material*, "ethereal", thing. This temptation is at the root of the question whether the existence of mind(s) is independent of the existence of matter, and whether there can be such a thing as a "disembodied mind".

A notorious difficulty connected with a dualist position is, moreover, how to account for the connection between mind and matter. A causal interaction between the two seems to conflict with deep-rooted ideas about causation in nature. If one rejects interaction one is left with the uneasy task of giving an account of the correspondence or "parallelism" which seems, somehow, to exist between "what happens in the brain" and "what happens in the mind". The Occasionalists saw in this parallelism a (from the natural point of view) contingent fact guaranteed (or "necessitated") by the will of God. Leibniz, in a similar vein, viewed it as a "pre-established harmony" between mind and matter; in one of his favourite similes, as the synchronization of two clock-works. The "mysticism" of these positions denying interactionism may strike us as a weakness.

In the view which I am here advocating, there is no causal interaction between the mental and the material. The sound waves cause neural processes and these in turn may cause macroscopic bodily movements which we understand as meaning that a subject has had a sensation and intentionally reacted to it. The subject reacts to the sound by doing something. His hearing

the sound may be said to cause his doing in the sense that it is the reason for it. But his intention (volition) in doing this does not cause the bodily movements in which his action may be said to "consist". These movements were caused by nervous processes which in turn were caused by sound waves.

There is thus no causality in the sense of either "material cause – mental effect" or "mental cause – material effect". But there is still a connection or, if one wishes, parallelism between the two. This "parallelism" is there by virtue of the criterional ("semantic") relation between, on the one hand, behavioural and mental (intentional) phenomena and, on the other hand, mental phenomena (sensations) and things and events in the physical world.

But if there is no causal relation between the material and the mental, how then does it come about that we understand bodily movements as signs of something mental – and sensations as signs of material things? To the extent that this question has a clear meaning I think the answer is as follows: the understanding results from the fact that we are members of a community of living beings with whom we can communicate and whose reactions we, up to a point, understand. It is essential, moreover, that this community should be a *language* community. Only within it can we talk of "conceptualization" and of the mind-matter *distinction*.

The material processes (bodily movements) which we conceptualize under the aspect of intentionality are, probably without exception, the effects of neural causes. But their *conceptualization* is not an effect of this. If it be called an effect of anything at all, this would be of "socialization", of the fact that we grow up to live with and understand our fellow humans and also a good many other living beings.

We must not lose sight of the fact that what we are trying to clarify is a *conceptual* distinction and therefore something which "exists" only for members of a human community. A hen easily learns to *discriminate* between grains which are edible and those which are not, I am told. But the *difference* between mind and matter does not exist in *its* world. Nor is it anything *we* have "learned" from experience.

If the mind-body distinction is a result of conceptualization, is my view then that the "real" is, in itself, neither mind nor matter but something "neutral", out of which mind and matter are, somehow, our "constructions"? Am I, in other words, propounding a form of *monism*?

Perhaps – but also a monistic position invites misunderstanding. The "neutral stuff" is not "a third thing", neither material nor mental. But the view of mind and matter as two *aspects* of what is real is perhaps the best way to view the double relationship of mind to matter and of matter to mind. This view is – or so it seems to me – not entirely unlike that of Spinoza's.

A PARADOX

Mental phenomena have behavioural criteria and material phenomena have sensational criteria. Thus – though with caution – mind can be said to depend, conceptually, on matter, and matter on mind.

Is there a vicious circle hidden in this mutual dependence? The failure to accept mutuality which has resulted in materialist theories of mind and phenomenalist theories of matter respectively can perhaps be said to reflect an intuition that one cannot "preserve" both dependencies without vitiating one's position from a logical point of view.

And perhaps this is how things stand: materialism and phenomenalism are both false, but one can reject them both with the right arguments only at the cost of acquiescing in a paradox. It would be like saying that reality (the real) is neither mind nor matter and that it is both mind and matter. It is not clear to me that this is necessarily an unacceptable position.

CRITERIA, SYMPTOMS, AND SIGNS

It has become current to distinguish between *symptoms* and *criteria*. The connection between a symptom and that of which it is a symptom is "empirical". This means, roughly, the following: if *A* is a symptom of *B*, the existence or occurrence of *A* may make us anticipate, expect or predict, the occurrence or existence of *B*. But whether *B* actually occurs or exists will have to be established *on independent grounds, i.e.* on grounds which do not themselves make appeal to (the occurrence or existence of) *A*. These independent grounds, moreover, are sometimes, but not necessarily, what we call criteria (as opposed to symptoms) of *B*.

An example: I have a slight headache, my forehead feels hot, I sometimes shiver as if I were cold. These are typical symptoms of having a temperature. I look for a thermometer, "take my temperature", read the thermometer and find that I have, or perhaps that I don't have, a temperature. ("Having a temperature" would mean, *e.g.*, that my bodily temperature is over 37°C.) In judging whether I have a temperature or not, I make no further appeal to the symptoms ("how I feel").

Temperature over 37°C could be called a *criterion* of the state we call "having a temperature". How do I know whether the criterion is, or is not, satisfied? By "taking one's temperature" and reading the thermometer. But the instrument may be faulty or the measurement not "appropriately" performed. I say perhaps "I feel so damn feverish that there must be something wrong with the thermometer, since it showed only 36.5°". Then the symptoms of fever are taken as *symptoms* of something being wrong with the thermometer. I proceed to test the thermometer – for example by comparing what it shows about my bodily temperature with what other thermometers show. When judging the

thermometer reliable or not I make appeal to some *criteria* for its reliability. In this process a good deal of "physical theory" is already taken for granted. Without it the very notion of having a temperature as distinct from feeling a certain bodily discomfort simply would not exist. The symptoms would not then be symptoms of anything – but be "the very thing itself". (And this, I assume, is the case in communities which are not too much touched by our scientific civilization.) If it were given a *name* and we called it, say, "fever", the signs *we* regard as symptoms would then be criteria of the thing thus named.

The relation between a criterion and that of which it is a criterion is conceptual, logical or semantic. (The three adjectives are here used as "rough synonyms".) Criteria can be necessary or sufficient. If A is a necessary criterion of B, then the absence of A on a given occasion entails the absence of B, too; if it is a sufficient criterion its presence entails (secures, warrants) the presence of B. If the presence of the necessary criterion – or, if there are several, their conjunction – is not also a sufficient criterion of the thing B, we say that there is a *residue of meaning* associated with the name *"B"* which is not captured by the criteria. And similarly, if the sufficient criterion – or, if there are several, their disjunction – is not also a necessary criterion of B.

Now the criterional relationships in which we are here particularly interested, *viz.* those between behavioural manifestations and mental phenomena, seem to be exactly of this kind when the name of the thing in question has a residual meaning which the criteria do not capture. The meaning of this must next be explained.

All necessary mental criteria of a mental state P can be there without P being there. On this is based the possibility of *faking*, not only primitive states such as being in pain, but also "higher" emotional and cognitive states such as grief, suspicion, belief or expectation. The dramatic arts exploit this discrepancy between "outward signs" and "inward state".

But it may also be the case that P is there without sufficient behavioural criteria to establish its presence. On this is based the possibility of *hiding* the mental by *suppressing* the behavioural signs which normally warrant its attribution to the subject. Such hiding may be easier the more complex or "spiritual" P is – whereas it may be next to impossible to hide, say, a state of suffering grave pain.

On these two discrepancies between mental phenomena and their behavioural criteria are founded various ideas about the "inaccessibility" of the mental to observation from "outside".

What is here said about the mental in relation to its behavioural criteria holds, *mutatis mutandis*, also for the physical (material) in relation to its sensational criteria. All the necessary sensations (of a material thing) may be there and yet the physical phenomenon absent. When this happens we speak of sense *illusions*. The reader is reminded of the rôle which the so-called

Argument from Illusion has played in various phenomenalist and also scepticist arguments and controversies.

But physical phenomena can also occur, material things exist, in the absence of anybody's awareness of their occurence or existence. On this possibility is founded the notion of the existence of an external (physical, material) world independent of mind.

In the concluding section of this essay I shall return once again to the "meaning-residue" separating the mental from the physical, and *vice versa*.

PSYCHO-PHYSICAL INTERACTION

I have argued that body and mind do not interact causally. The chain *S-N-M* is a causal chain connecting three physical (material) "systems" – for example, sound waves, neural processes, and bodily extra-neural reactions (movements).

In the chain *S'-I-M'*, *S'* is a mental phenomenon which for the subject, I, who "has" it is a *reason* for doing something. The mental phenomenon *S'* is "called forth" by the physical stimulus *S* and what the subject does results in the response *M*. If we wish to call *S'* the cause of *M'* – for example, a sensation which is the cause of an action – this is innocuous enough and not contrary to the way we talk about these things. We could then say that the chain *S'-I-M'* is a causal chain, too, connecting three psychical (mental, immaterial) "systems" – for example, hearing a sound which makes a subject turn his head.

In this picture, however, there is yet a third "causal pair". The physical stimulus "calls forth" something mental and the subject's reaction to it "results" in some bodily movements, we said. What is the nature of these relations of "calling forth" and "resulting in"? Shall we say "physical cause – mental effect", "mental cause – physical effect"? In some innocuous way this, too, could be said – and yet not conflict with the sense in which I wish to deny mind-body causal interaction. But we must be aware that in saying this we plunge into a conceptual morass.

Part of the difficulties here have to do with the relation between the middle terms, *N* and *I*, of the above two chains. I call it the relation between the brain and the I (the subject).

The brain is thought of as being, somehow, the *locus* of the mind. (Descartes and the pineal gland.) When the sound waves affecting the hearing-nerve have been "propagated" to the cortex, sensations "mysteriously" originate *there*. And it is *from there* (or so we think) that the volitional impulse resulting in intentional action also originates.

We normally say that it is the stimulus (*S*) which "causes" the sensation (*S'*). The sensation is *of* the stimulus. Had there not been this stimulus there would, probably, not have been this sensation either. By reproducing the stimulus we

can, normally, produce in the subject a qualitatively similar sensation. The words "probably" and "normally" here mean that there are exceptions. But they do not invalidate the rough regularity we are considering.

That which happens in the link *N* is supposed to produce, be causally responsible for, not only the sensation but also all those bodily signs on the basis of which *we* attribute the sensation to *him*. In the case of human subjects we largely rely on verbal reports. They result from macroscopic movements in the subject's body which produce sounds or jottings on a bit of paper which we understand to mean that he had, or failed to have, a certain sensation. If *N* does not "mediate" any such "output" *M* in response to the "input" *S* we could not attribute to the subject any sensation either. Our further reaction to the situation, if any, would be to look for some "breakdown" or "defect" in *N*.

What causally results from the processes in *N* are, we have said repeatedly, the *criteria* on the basis which we attribute a sensation to the subject – for example behaviour typical of listening or looking, or simply his saying "I hear ---" or "I see ---".

"But *in* the subject (*I*) those processes (*N*) produce a sensation." What does "in" mean here? In his brain, a part of his body? In the brain material processes go on. *There* is no "room" for anything mental.

The *subject has* the sensation. Nothing mental is "in" him. In his *body* there are various things. They are all material.

But why not say that the sensation is what happens in *N* "conceptualized under the aspect of intentionality"? Then they would be in some sense "the same".

What is straightforwardly wrong with this is that there simply *is* no such "conceptualization". Nor does the sensing subject, except in marginal cases, "interpret" the bodily signs which are *our* criteria for attributing to *him* sensations and other mental states. The subject simply *has* sensations, beliefs, thoughts, intentions, *etc.* (And many of these things he has as the effects of various external and internal stimuli affecting his nervous system.)

In order that the subject might *interpret* what is going on in his brain, he must sense or otherwise be aware of those processes. Could the sensation itself, *S'*, which he has of a sound, say, be called awareness? There is an immediate inclination to say No. In having *S'* he is aware of the sound – not of what goes on in his nervous apparatus. But before rejecting the suggestion out of hand, we must once again go over the whole conceptual terrain.

PSYCHO-PHYSICAL PARALLELISM. INTROSPECTION

The two chains *S-N-M* and *S'-I-M'* constitute a kind of "psycho-physical parallelism". *S* is, for example, sound waves affecting the cochlear nerve (belonging to *N*) and *S'* an acoustic sensation (which *I* has). *M* is, for example,

a turn of the head (of the body in which *N* is located) and *M'* the intentional action (of *I*) of turning the head.

The above straightforward sense of "psycho-physical parallelism" must be distinguished from another, related idea of parallelism. This is the idea that the neural events called forth by *S* in *N* somehow "correspond" to the mental phenomena *S'* and *I* and, further, that the neural processes in *N* which cause *M* somehow "correspond" to the mental states in *I* (his intentions, volitions) which make us speak of the bodily movements *M* as the doing of *M'* (by *I*). Taking this view, the parallelism is not between the two extreme terms of the three-termed chains, *S-N-M* and *S'-I-M'*, but between the middle terms *N* and *I* – between what happens in the neural system of the person's body and what happens "in his mind".

This idea of parallelism or counterparts or equivalents in the brain to what we register as mental phenomena is anything but straightforward. It is on the contrary exceedingly obscure. And yet there is "something to it".

The parallelism which we are now trying to get hold of is, somehow, between neural events and the mental things "themselves" and *not* their criteria in the physical world. But how do we have access to those inner things "themselves" if not through their outward criteria? There is a classic answer: through *introspection*. This is a kind of self-observation the results of which are reported in what may be called *phenomenological descriptions* of the mental. These reports are, essentially, linguistic. For this reason it may be denied that animals other than humans are capable of introspection.

If I report what I perceive – a sound for example – my report is of a percept. The percept is that of which I am aware in the perception, its object. The veracity of the description (report) is checked by observing the percept. This is some intersubjectively observable phenomenon in the physical world. My report was not introspective but extrospective.

The object of an introspective report is a sensation or some other "mental phenomenon" for example a recollection, a feeling, or a thought. If the sensation happens to have an extra-mental intentional object, the veracity of its description is *not* now checked by observing the object, but depends on whether the reporting subject actually *has* a sensation of this description. Normally, we take him "on his word". But sometimes we have doubts. Then we have to rely on futher behavioural signs (in addition to his verbal reaction) which speak for or against the veracity of his report.

"I heard a sound" can be called an introspective report. It seems unnatural to call it a description of a sensation. I can qualify the sound I hear – and also qualify my hearing of the sound. I say, for example: "it is a sound of a drum", "it is a shrill sound, it hurts my ears", "it is barely audible". Are these descriptions of my sensation? The answer depends on how we understand them. If "I hear a sound of a drum" means "I hear a drum beating over there", I do

not describe my sensation but report a perception. But if it means "the sound I hear is like that of a drum", then I describe my sensation. Similarly, "the sound is barely audible" can be an objective statement about its loudness or a phenomenological description of my impression. Another person near me hears the sound quite clearly. Then presumably he has better hearing than I. This difference between us may in its turn be reflected in our respective neural equipment. If so, the difference in the phenomenological descriptions of our sensations can be said to "run parallel" to a difference between our nervous systems. *This* parallelism, however, is grounded on another one, *viz.* on the correspondence between a detail of my description of my sensation (calling it "barely audible") and some fact or feature of my nervous system.

Setting aside questions of veracity, we now ask: Does every detail of a phenomenological description (such as calling a sound "barely audible" or "shrill" or "like that of a drum") answer to some identifiable structural or functional fact about the person's neural system (brain)? The question is intended to be empirical, a matter for science to answer. If the answer is affirmative there is "parallelism" between neural events and "subjective experiences".

THE CASE OF THE AFTER-IMAGE

Assume the agent reports on a visual after-image he is having. He sees it "projected" on the wall in front of him. When he moves towards the wall its size shrinks ("seems to shrink"); when he steps back, its size increases. Since the retinal image remains unchanged, how is it that his experience of the size of the thing he sees there, on the wall, changes? One is tempted to say: the difference must be due to the way the subject conceptualizes his neural state. A constant retinal image at varying distance from its location in physical space, *viz.* on the wall, must "mean" that the percept, *viz.* the patch, changes its size.

Is this not a refutation of psycho-physical parallelism, however? How could a *constant* optic stimulation of the brain answer to *variations* in the subject's optic experience?

Without brain processes of a determinate nature there can be no after-images. These processes can be investigated in minute detail – at least "in principle". Do we perhaps find among them also a cause of the change in apparent size of the after-image in spite of the constancy of the initial optic stimulation? I do not know the answer. But I can imagine the following possibility:

The fact that my distance changes from the background where I see the after-image patch is something which I notice. I approach this background or recede from it; alternatively it approaches or recedes from me. These movements, too, have effects on my brain. For example: if I step forward, the

movements of my legs are caused by nervous impulses from the brain. The movements are, moreover, what we call "intentional". Thus, changes in (the state of) my brain correspond to the changes in my visual after-image. Is this not all that is required for the correspondence between the phenomenological description (report) and the neurological observations, between the subjective experience and the neural events?

Let us assume that the possibility I envisaged is a fact. What happens then to the above talk about "conceptualization"? Surely, the subject itself, the I, did not observe what happened in his brain and *quasi* "took it into account" when "forming" the after-image. But there is a temptation to say that his "perceptual apparatus" took into account the variations in distance from the background and reacted to them in a meaningful though mistaken way, "arguing" that to a retinal projection of given size corresponds a bigger object at a longer distance and a smaller object at a shorter distance and then "saw" the object (the patch on the wall) accordingly as bigger or smaller.

This, naturally, is sheer "anthropomorphism"! The fact of the case is simply that certain changes in the seen after-image correspond to certain changes in the neural apparatus.

The phenomenon of changes in the after-image itself is interesting. That things "look smaller" at a greater distance is a well-known fact. It is "meaningful" from the point of view of estimating what we call the "objective size" of things. Objects which are approaching us do not, on the whole, grow in size when approaching. (Sometimes of course, they do.) This is an experimental fact about the physical world. Has the perceptual apparatus "learnt" to react to it? Or is it an "innate" capacity – perhaps a result of the evolution of the species?

It is inviting to speculate about these questions. I shall not do it here, however. I shall only make the following conceptual observation on the matter:

When saying that the real size of objects on the whole does not change with their distance from the seeing eye, we are assuming that things in the physical world *have* a "real size". This assumption, however, involves a conceptualization of *perceptual* data, of experiences we have under certain "optimal conditions" of distance, illumination, stability, *etc.* "Real size", moreover, can be measured. This involves perceptual comparisons with objects, the "real size" of which is already taken for granted.

The case of the changing after-image (constant physical stimulus – changing mental response) is thus not a counterexample to the idea of parallelism between "what happens in the brain" and "what happens in the mind". When thinking the case through, the notion of parallelism becomes, it seems to me, more intelligible than it might initially have appeared to be.

In other cases the correlation of details of a phenomenological description with neural events appears less problematic than in the case of the after-image. Thus, for example, when we describe a pain we have or mood we are in (how

we feel). The pain is, say, "stabbing" or "piercing", intermittent or continuing, now stronger now weaker. It would be surprising if these variations of our *experience* did not answer to determinate variations in some neural processes. So that, by reproducing experimentally those processes in the brain one could reproduce the same variations in the experienced sensation.

THE CONCEPTUAL INGREDIENTS OF PERCEPTUAL EXPERIENCE

There are a number of perceptual phenomena of which is characteristic a *discrepancy* between the perceptual "content" and the sensorial stimulus. After-images are a case in point. We see the image "wax and wane" with variations in distance, although the projection on the retina remains constant. In other phenomena the reverse is the case: the (mental) impression remains the same although the physical impact on the sense organ varies. For example: the impression of roundness of a coin (which *is* "objectively" round) remains, up to a point, unaffected by a change in the angle from which we see it. The impression of whiteness of a sheet of paper (which *is* white) is to a certain degree "insensitive" to changes of illumination. To the same group of discrepancies between sensorial "input" and perceptual "output" also belong various "sense-illusions" and so-called *Gestalt* phenomena.

What is the significance of these psychological phenomena to our problem here? This is a question which the Finnish philosopher and psychologist Eino Kaila interestingly debated. Kaila appears to be practically unique in this respect – which may be due to the fact that he was not only a distinguished philosopher but also a good psychologist.

It is natural to say – as did Kaila – that in all these cases a *conceptualization* takes place of the "material" offered by the senses (sense organs). It is as though our perceptual apparatus, when "transforming" the impact which the stimulus S makes on N into the perception S' of something (S), "took into account" what it "knows" or "fancies" about the objective character of the stimulus. The coin is round[6] and therefore it is *seen* (perceived) as round also when its projection on the retina is elliptic. In this case, persumably, the perceptual apparatus was "right". In the case of the after-image, however, it was "deluded": the object seen farther away is not bigger than the one which was seen nearer.

It is temping to say of these phenomena that the "mind" argues, reflects, thinks about the situation in N and then presents it in the form of a "conceptualized perception" S' of S. But this of course is only a metaphor – and moreover a dangerous one. One can still say that the "perceptual apparatus" does the "conceptualization" – but this apparatus is *the neural system* (N) itself and the conceptualization means that the character of our perceptions is determined, not only by the "crude" impact of the stimulus on the sensoric

receptors but by what has been called an "integrative activity" of the entire system *N*. It is, so to speak, the "cleverness" or "wisdom" of the *body* (*N*), not that of the *mind*, which we witness in these phenomena. It is the "meaningfulness" and sophistication of the body which impresses us.

To what extent are these capacities of the neural system acquired and to what extent are they innate? Although "innate" in the individual, they are presumably acquired by the species in the course of evolution. In this acquisition their "purposiveness" manifests itself; they have shown themselves to have "survival value". The problems which crop up in this region are a mixture of biological (scientific) and conceptual (philosophic) questions.

THE CASE OF BELIEF

In the case of many mental phenomena the notion of a phenomenological description (of what it is like to have them) is unclear. Take a belief, for example. It can have some likeness with mental states. "I feel absolutely sure", I say. But how similar to "feeling pain" is "feeling sure"? The first is a prototype of a mental state. But is a belief a "state"? Is it even "mental"? It may be said in reply: *faute de mieux* we call it a "mental state". It is surely not a "physical phenomenon".

How does it *show* that I have a belief? In that I am likely to answer certain questions "Yes" and "No", to do certain things and refrain from doing others, make some preparations, warn people, *etc.* This is how my belief "shows itself". It would be slightly odd to say that this is what my belief *is*. But not odd to say that this is what my *believing* is. To "have a belief" is, in some ways, more like something we do than something we, like a sensation, "suffer" or experience.

Suppose someone tries to describe what he feels when he affirms that he believes something. Whatever he describes would not be his *belief*, but at most something "sensation-like" associated with it. The feeling is sometimes stronger, sometimes weaker, say. To the extent that this is a genuine phenomenological description there is as little reason to doubt that it has neurological correlates as to doubt that his feeling of pain has such correlates. The other things again in which his belief "shows itself", are the effects of neural causes. So why not say that the mental phenomenon we call his believing something has a counterpart on the neural side?

But would this not imply that if we could experimentally (artificially) reproduce those neural events in him we could "make him" believe that thing? Before labelling this as nonsense let us stop to think wherein such "reproduction" could conceivably consist.

In order to have beliefs at all, a human being must have reached a certain stage of maturity and acquired a certain amount of "life experience". This

maturation consists partly also in bodily changes, among them structural changes in his brain and nervous system. In order even to be "able to believe", thousands of "connections" between nerve-fibres in his brain must have been established – partly, perhaps even mainly, as the effect of various outside stimulations through the sense organs. If these "connections" are damaged, a belief once acquired and held may also be correspondingly damaged, *i.e.* distorted or lost – for example, the person fails to take certain precautions which normally would go with having the belief in question.

We all know, roughly, what it means to "induce" a belief in a person or "make him believe" a certain thing. We subject him to various influences from outside – stimulations which affect his body, including his brain and nervous system, as well as his "mind". Could these changes in his brain be induced artificially, *i.e.* not *via* those stimulations but by some sort of neurological surgery "knitting together" the nerve-fibres so as to *make* his brain like that of a person who has in the normal way acquired the belief in question? In practice surely not. But "in principle"? Could we make him "believe", for example, that it will soon be going to rain although he did not see the black clouds which made *us* believe this? Perhaps we could make him respond to the question (through physical stimulation of his ears) "Do you believe that it will soon start raining?" by saying (*i.e.* producing the sound) "Yes", and to the question "Did you see the black clouds?" by "No" and (even) to the question "So why do you believe this?" by yet another "grunt", which we perceive as a sound with a meaning.

We need not doubt that all his reactions have neural causes which result in bodily effects. Are these causes "the neural counterparts" of his belief? I would answer: *Call it so, if you wish*. But does the person really *have* the belief or does he only react *as if* he had it? Both could be the case – and we know, roughly, how to find this out.

But we are not now interested in the neural causes and causal conditions in *N* of the bodily reactions which are our criteria for attributing to a person a belief, but in the correspondence between what is going on in *N* and a person's *having* a belief.

We *attribute* to a person *a belief* on the basis of a complex pattern of his bodily reactions (and "dispositions" to react). This is a conceptualization of observations on physical phenomena under an aspect of intentionality. The person does not "attribute to itself" this belief on that basis. He "just" believes, has the belief. He may admit this if asked, or profess it without being asked. Or he may deny this – openly with an intention to deceive, or latently, thus deceiving himself.

Is there then no such thing as a "belief feeling", a touch which singles out a certain mental phenomenon as a belief? People (philosophers) have entertained such a fantasy. And perhaps there is, in some cases, such a feeling,

introspectively recognizable and susceptible of a phenomenological description. Would the details of this description, or changes in it such as varying intensity, correspond to something in the brain? I do not see why this could not be the case – nor why one could not by experimental interference with the brain produce the variations thus described. But our only control that it is so is the *phenomenological description* itself – and this is a verbal (behavioural) reaction of the subject to stimulations in his brain (nervous system).

THE DISEMBODIED MIND – A COUNTERFACTUAL FICTION

Something similar holds good also for the case of the after-image. What we were "matching" there, were the details of the phenomenological description of the sensation with states and processes in the nervous system. The phenomenological description is itself a verbal reaction to stimuli. The reaction "consists" in the bodily changes which have, we believe, sufficient causes in the brain (nervous system) when duly stimulated from outside. But the mind-body parallelism of which we are in search, is not between brain events and the *verbal reaction* but between brain events and (changes in) the after-image ("itself"). Can we not skip or eliminate the bodily reactions which, so to speak, "stand between" the brain and the mind?

Cannot a person experience an after-image without reacting to it with any of those bodily signs (outside the nervous system) on the basis of which *we* attribute such an image to him? Of course he can. But if asked what his *having* the after-image means, the answer must be that it means that he *would have* reacted in the "appropriate" way *had* we put him to the "appropriate" test. The "subjective content" of the phenomenological description, the "purely mental", can only be captured through such a *counterfactual* move.

What we have here said about the mental has a counterpart in a well-known philosophic answer to a question about the existence of material things. Someone says: "There was once a material thing, big as a tennis-ball, white in colour. But it perished without anybody having ever seen or touched it. And it left no traces at all in the world." Is it nonsense to say that such a material thing existed? I think the answer is No. But what does it *mean* that it "existed", if not that if some living being with sense organs *had been* there to sense it, he *would have had* the appropriate sensations of size and colour, *etc.*?

This "unsensed" material body and the "disembodied" mental content both "shrink" to an unreal "counterfactual point". In both cases it is the double dependence of matter on mind and of mind on matter, in combination with the residues of meaning of the two in relation to one another, which tends to confuse us.

When speaking of these topics *language* is constantly trying to pull our leg. The difficulty is to resist this.

A WORLD WITHOUT MIND – A REALIST POSITION

It seems easy enough to "think away" all consciousness, leaving the world "physically intact". "There it still is – unseen, unheard, untouched – with the lakes and the woods, the blue of the sky, and the murmur of the billows when they reach the shore." The lakes and the woods, yes. But what of the murmur and the blue of the sky? This last is not even the colour of a surface – how can it "be there" when there is nobody to see it?

But wait. Saying that the blue of the sky is there *means* that those same atmospheric conditions prevail which make the sky appear blue when I look at it. And similarly: that the sea is there although nobody hears it *means* that the molecules in the air above and surrounding the water move in a characteristic way. These are objective features of the world and their presence explains why a spectator, if there were one, would see the blueness of the sky and a listener, if there happened to be one, would hear the murmur of the billows.

Is this so? We are then assuming that there is – "independently of consciousness" – air and water and some properties of those things which are "causally responsible" for certain optic and acoustic sensations which a sentient being would have if he were present. But what does it *mean* that there is water, say? Water is a tangible liquid stuff, without any shape of its own, colourless, transparent, *etc.* Air has many of these properties too, but is not even tangible. Wherein does this difference between air and water "objectively" consist? In some properties of theirs which affect our sense of touch differently. What then is the "objective reality" of those properties? The answer, presumably, requires a "descent" to the submicroscopic level of the structure of the two elements.

By pursuing such questions as those above we seem "pushed" to a conception of what "there is" in a world without consciousness which leaves us with molecules and atoms and whatever there chance to be "inside" them. Are these the "most real" material things then? Some philosophers of the highest ranking, have doubted that those things are real at all!

As long as one thought of atoms as indivisible, very small solid bodies it was also natural to think of them as having a determined geometrical shape, for example as small spheres. Newton said that atoms were "hard". But these pictures of the atomic world are no longer considered very appropriate. Are electrons "solid"? Do they have "shape"? *Some* concepts which are related to those of solidity and shape may still be applicable also to electrons, *e.g.* the concept of size or position in space. But *many* pictures which we associate with these concepts in the macroworld may not be applicable to electrons or to the microworld generally.

(What holds of our pictures of microparticles holds, *mutatis mutandis*, also of our pictures of microwaves. We do not see them undulate like billows on the water – except in our imagination.)

The reality of the microworld manifests or reveals itself in the macroworld

in the observed results of experiments which are designed on the basis of highly sophisticated theories about "the atom". Agreement and discrepancies between anticipated and observed results also serve as a test of the theories. The observations are of macroscopic percepts, and the theories may be called a conceptualization or interpretation of these percepts (objects of perception) – *somewhat similar* to the sense in which talk of intentional action as a conceptualization of observations of the (macroscopic) bodily movements of living beings.[7]

One might question the adequacy or truth of various theories of microphysics – on the ground, for example, that predictions from them disagree with the results of experiments or that they cannot explain some such results. One can also question the appropriateness of various "pictures" which we make ourselves of microphenomena – such as the partly conflicting pictures of them as moving particles and as propagated waves. Questioning of both these types have, I would say, contributed to gradual but hardly yet definitive modifications in the theories after the great breakthroughs in atomic research in the beginning of the century. But no such modifications of theory or rejection of "pictures" amount to a denial of the "reality" of the phenomena under study.

What are, then, these "phenomena under study"? In the first place they are macrophenomena which are the results of experiments – for example light-points on a screen or traces in a cloud chamber. To deny the reality of these phenomena would be tantamount to labelling as sense-illusions certain perceptions, universally treated as veracious. Has anybody seriously attempted this? Not to my knowledge.

We interpret the macrophenomena in question as the manifest effects of "underlying" microphenomena. This we do by constructing a theory. The language of the theory has traditionally been strongly pictorial. But the success ("truth") of the theory depends on the macrophenomena which it makes us anticipate and predict. The "pictorial microworld" is just our imagination. The *reality* is that of the macrophenomena. And this reality nobody has the slightest reason to doubt.

So what is it that "is there" when we "think away" all consciousness leaving the world "physically intact"? The answer we already gave. It is the same physical world which is there now – with its lakes and woods, the blue of the sky, and the whisper of the wind. The fact that there is nobody to enjoy the sights or hear the sounds does not detract from their "reality". To think otherwise is to let some pictures confuse us.

NOTES

[1] These "Notes" were a preliminary study for the following chapter "On Mind and Matter".

[2] This conceptualization is conditioned, it seems, by the degree of *resemblance* which there is in size, anatomical structure, and characteristic movements between animal and human bodies. In a classic study on the amoeba we read: "The writer is thoroughly convinced, after long study of the behaviour of this organism, that if Amoeba were a large animal, so as to come within the everyday experience of human beings, its behaviour would at once call forth the attribution to it of states of pleasure and pain, of hunger, desire, and the like, on precisely the same basis as we attribute these things to the dog." H.S. Jennings, *Behaviour of Lower Organisms*, 1906, p. 306.

[3] There is a danger connected with this way of speaking. The danger is one of "hypostasizing" the signification or meaning of the sign as something "thing-like", a thing signified or meant, – like the bearer of a label. The idea of the mental as something thing-like and the analogous idea, deep embedded in our thinking, of the "soul" or mind as a body-like phantom or ethereal spirit, is one of the greatest obstacles in the way of clarity when we philosophize about these matters.

[4] Cf. below p. 147 on the causal priority of the neural in relation to the behavioural.

[5] One is here reminded of Leibniz's famous mill simile. (*Monadologie*, 17.) Let us imagine a brain enlarged so much that we could enter and walk around in it and see the working of the "neural clockwork" in detail. We could then study the neural links *N* which causally connect *S* and *M* as two events in the material, physical world. Here everything happens "*comme si la mauvaise doctrine de ceux qui croient que l'âme est materielle suivant Epicure et Hobbes, estoit veritable*". (The quotation is from Leibniz's reply to Bayle.) But at the same time we would see, like a mirror image, the way in which *sensing* the input *S* by the subject connects with his *producing* the output *M*.

[6] The "real shape" of the coin is round. But it is not the roundness *we see* ("as seen") which is its "real shape". The roundness of the coin is a property which it has by virtue of the fact that all points on the edge of the coin have the same distance from one and the same point in its middle. The seen roundness is only a "symptom" of this. If we distrust the symptom we should have to carry out measurements. This too involves perception – in the form of observations of the coincidence of points on the measuring rod with a point on the periphery and a point in the centre of the coin. But the equidistance as thus perceived (observed, seen) is no more the "real shape" of the coin than is its roundness.

[7] Talk of action is not "theorizing" about the mind; but it could, with caution, be called "theorizing" about *bodily phenomena*. I know that many would resent this analogy and condemn it as "scientistic". One can do this – and yet *also* see a similarity here.

ON MIND AND MATTER

1.

The meeting between philosophy and neurophysiology has an exactly localizable historical origin. It is with Descartes, who is often called the father of modern philosophy but who is also recognized as one of the great pioneers of Western mathematics and natural science. His vision of the human and animal body as a physiological machine, however, was almost completely speculative and it was only by slow degrees in the course of the centuries that it acquired the status of a scientific view of the living soma. The work of Galvani and Claude Bernard are milestones on this journey. But it was hardly before our century that serious dialogue between the findings of scientists and the reflections of philosophers in this area became possible and urgent. Before then the field belonged to philosophy. The concepts which philosophers had devised for talking about the subject are still so deeply entrenched in our language and thinking – also in that of scientists – that I find it appropriate here briefly to recall the philosophic load which we still carry with us.

It can be said to be part of our intellectual inheritance to acknowledge a *sharp* distinction between mind and matter. The first to make it was Descartes. Material things are extended in 3-dimensional space. Mental things Descartes called "thoughts". They have no extension.

Descartes used the term "thought" in a broader sense than we ordinarily do. In addition to what we would call "thoughts" or thought-like phenomena such as beliefs, recollections, and expectations, he also included among thoughts sensations and acts of will ("volitions"). The best modern word for Descartes's "thought" or "thinking" is perhaps *consciousness*.

Matter and mind, although sharply distinct, are also, somehow, correlated. (This is related to the fact that material and mental phenomena both occur in time.) At least in the case of sensations and volitions it is inviting to think of this correlation as *causal*. Sensations are the effects, we think, of affectations of the sense-organs of a living being by material things or events. Acts of the will again cause movements of parts of the bodies of such beings.

This is what we say in common speech. But we easily feel that there may be a conflict here with what we regard as "scientific thinking". The effects of material processes, one is inclined to say, can only be material processes; and movement, being a material phenomenon, can only be effected by some other

G. H. von Wright, In the Shadow of Descartes. Essays in the Philosophy of Mind, 125–148.

A slightly abbreviated version of this paper appeared in *Journal of Theoretical Biology* **171**, 1994, pp. 101–110.

material phenomenon. To think otherwise is "animism" and irreconcilable with a "scientific view of the world".

Descartes himself subscribed to an *interactionist* view of the body-mind relationship – although he may be said to have had some qualms about it. For his successors this soon became a major headache – and they looked for different solutions.

The question of causal interaction should be distinguished from what I propose to call the question of *conceptual* (or metaphysical) *relationship* between matter and mind. Several positions on the latter question are known from the history of thought. One is the *materialist* position. According to it, the mental is, somehow, "reducible" to the material. Another position is known as *idealism*. On this view, matter is reducible to mind. A third position is known as psycho-physical *identity theory*. According to it, the material and the mental are two aspects of the same ultimate reality.

All the three views mentioned are *monistic* in that they object to the cartesian dualistic "slicing up" of reality into two different "stuffs". Everything real is ultimately the same substance. The three principal views, moreover, exist in several variants. A form of materialism which has gained wide currency in recent time is known as eliminative materialism. A form of idealism again which played a rôle in the debates stirred by logical positivists earlier in this century is called phenomenalism or sensationalism. It again has affinities with the form of identity theory known as "neutral stuff monism", once professed by Ernst Mach and later, for a time, by Bertrand Russell. A difficulty for any form of identity theory has been to steer a "middle course" between materialism and idealism without collapsing into either.

Where on this map does the position belong for which I shall try to argue? I find the question difficult to answer. I am not an "interactionist" and I hope to be able to show convincingly why. But nor am I a "reductionist". I do not know whether my view is more like a monism or like a dualism. It has a certain affinity also with the view known as epiphenomenalism.

2.

I shall present my view at the hand of an example from everyday life. The example is not so simple as to be trivial – nor so complicated as to be unperspicuous.

Dramatis personae are two people. The one I shall call "*P*" from "person". The other I shall call "*O*" from "observer". *O* is observing *P*. We can imagine that he has done so for a long time; perhaps he is *P*'s senior and has been watching *P* even from early childhood.

Now a sound is heard – moderately loud and not too far from the place where *P* (his body) is. Following the appearance of the sound *P*'s head turns so as to face the direction from where the sound emanates.

By being (observable) events in 3-dimensional physical space, the sound and the head-turn can be said to be "material (things)", or at least to belong to the "material world". There is nothing mental (psychic) about them.

Now *O* asks *P*: "Why did you turn your head?" *P* answers: "There was a sound." (Perhaps *O* did not hear it.) This would be a satisfactory answer. But perhaps *P* "embroiders" it a bit and adds: "The sound frightened me; I wanted to know what it was". Or: "I thought it was *Q* entering the room; I am expecting him".

In addition to the two events belonging to the material world, we have now introduced on our stage certain mental things: an acoustic sensation, *P*'s hearing the sound, and perhaps also some other things which *motivated*, were *reasons* for, *P*'s turning his head. These other things are mental too: *P*'s being frightened; his wanting something; his believing and expecting something.

In the case, as I have described it, *O* took for granted that the *movement* of a part of *P*'s *body, viz.* the head, was (the result of) an *action* of an *agent, viz. P*. An action which is performed by moving some limbs or some other bodily parts, we would not call a "mental thing". But it has a mental component, so to say. This consists of the reasons for which the action was undertaken. The reasons make the movements *intentional* or "*willed*". But these last two terms should be used with caution. To say that *P*'s turning his head required or resulted from an "act of will" would in most cases be misleading, if not straightforwardly false.

3.

Some comments on the concept of a reason are called for. As noted, the sound itself, the physical phenomenon, can be given as a reason for *P*'s action. But this presupposes that *P heard* it. Also an "unheard" sound can affect a body and call forth some effects in the nervous system. But it cannot be a *reason* for an action. In order to be a reason, the sound must be something of which the agent is aware (conscious) and to which he can refer in reply to a question why he did a certain thing. Therefore it is better to say that the reason is the acoustic sensation, and not the acoustic phenomenon (the sound "itself").

But how can a *sensation* be a *reason* for an action? The answer is that it can be this only by having a relation to many other things in the agent's "mental life", such as his beliefs, curiosity about things, expectations, wants, *etc.* He must have *learnt*, for example, that a sound can be a signal of something which is of interest to him – perhaps a warning of impending danger. And if *O* is satisfied with *P*'s answer "There was a sound" to *O*'s question "Why did you

turn your head?", *he* must know from long experience something about the various connections with other things which can make a sound be the reason for doing something.

Are not the reasons for an action the *causes* of the action? This is what we commonly say. The locution is unobjectionable – as long as we call the reason causes of the *action*. "Reasons cause actions" – all right. But suppose that somebody said that the reasons are the causes of the bodily *movements*. Then he is stepping into a conceptual morass. This, I am afraid, is what we too shall have to do in order to become clear about things. But before doing this we must sidestep a little in different directions.

4.

Things which count as reasons (for an action) are those, reference to which makes intelligible a reply to a question "Why did you do this?" or "Why did you not do this?" If asked, why I turned my head on hearing a sound, I answer that I thought *Q*, whom I am expecting, was knocking on the door, we understand this. But if I say that I turned my head because I felt hungry, one would not – at least not without "further explanations" – understand me. The things which count as reasons must be familiar from or integrated into a common "form of life" of the people who make and answer the "Why"-questions. Not everything counts – and certain things more obviously count than others. Sometimes we do not understand, how something which is referred to as a reason for doing something *can be* this; but when we are being told more about the case under consideration we perhaps understand it.

(We are all familiar with the case when we retort to the agent's answer "I did it because of *R*" with a sneer "That's no reason!". This often means, *not* that we do not understand the agent's answer, but that we do not *accept* (approve of) his reason. We think that he ought not to have acted for *that* reason. We then take a moralizing attitude to the case.)

Being a reason must be distinguished from *having* a reason. To be expecting somebody's arrival *is* a reason for being alerted by a knock on the door. One *has* that reason if, on a particular occasion, one has the expectation in question.

From the fact that something *is* a reason for an action it does not follow that the agent, when acting, *has* that reason for his action – and from the facts that something both is a reason and that the agent *has* that reason it does not follow that he does the action *for that reason*. Perhaps I did not turn my head in the direction of the sound because I actually was expecting somebody, but because the sound frightened me. (I thought "Was it an explosion?".) This observation raises a difficult problem:

What are the criteria of *truth* of a reason-explanation of an action? There are two *necessary* criteria, as just noted. The given reason must *be* ("count as") a

reason *and* the agent must *have* (had) the reason on the occasion in question. But the satisfaction of the necessary conditions is not yet *sufficient* for the truth of the explanation which is offered for the action. The agent may have had several reasons for a particular action of his, but it is not certain that they all influenced his acting. Perhaps in reply to a question "Why did you do this?" he refers to one or other of the reasons which he had but which did not in fact influence him – and omits mention of that reason on which he actually acted. Perhaps he wanted to lie to the questioner or to hide something about his motives. Or perhaps he was himself confused about them. One can lie also to oneself – and one may even have a *reason* for one's dishonesty.

Explanation of action in terms of an agent's motivation (his reasons) challenges interesting and intricate problems of psychology. They need not occupy us here. But it is good to be aware of their existence.

5.

Suppose that *O* himself hears the sound – or knows that it has been produced – but notices no reaction to it on *P*'s part. "Why did you not react to the sound?" he asks. Perhaps *P* answers: "I did not hear it". This is a satisfying reply. But it is not the only possible one. *P* might also have said: "I know what it is, it is nothing to pay attention to". Or: "I was making a multiplication in my head and did not let the sound distract me". Or perhaps he says: "I simply was too tired; I had not the energy to be curious". In these latter cases, *P*'s passivity is also a reaction to the sound. His remaining passive had a *reason*. This was a reason *against* turning his head, looking in the direction of the sound. (But had *P* been deaf and not heard the sound, his deafness would have been a *cause* of his remaining passive, not a reason for it.)

Omitting an action because of some reason or reasons for not doing it, is itself an "action" in its own right. I think it is an essential (conceptual) feature of (the notion of) an action that there should exist reasons for doing it on some occasions, and also reasons against doing it on some occasions. On one and the same occasion, moreover, there may be reasons both for and against performing the action. The agent will then have to "weigh" the reasons for and against; his reaction shows which ones weighed heavier. On other occasions again the agent may neither have reasons for nor against the action. If he nevertheless does it and we ask him "Why?", he would have to answer "For no particular reason". Such cases occur – but they are, it seems, not very common. If again the agent then omits to act, he would presumably say that he did this because there was no reason, for him, for doing it. This, I think, is not at all an uncommon case.

6.

Suppose next that *P* heard the sound and that *O* saw *P*'s head turning. In reply to *O*'s question "Why did you turn your head?" *P* now says: "I do not know, it was quite 'automatic'". Or perhaps he answers "*I* did not turn my head; *my head* just turned." Let us imagine that the same happens regularly – at least when the sound is moderately strong and the agent not too busily engaged in some other activity. Under normal circumstances his head thus turns when there is a sound. He, the agent, cannot account for this; nor can he, as we say, "help it". If, in reply to *O*'s question, *P* had said "There was a sound", he would be pointing to the *cause* of a *movement*, and not to the *reason* for an *action*.

In the case we are now imagining, *P*'s head-turn would be called a *reflex*-reaction in response to a stimulus, the sound. Perhaps there is such a reflex, for example in small children who do not yet "act for reasons", or in animals. This would be a "meaningful" reaction in view of the fact that sudden sounds in our immediate neighbourhood often are signals of something which is important for us, such as, *e.g.*, approaching danger. Maybe traces of such a reflex still remain in grown-up people. So that our head-turning reaction to a sound sometimes can rightly be said to be "automatic", "mechanical", "involuntary". I do not know whether this is so, or not. But I can imagine that it is so. And if what I imagine holds true this would, also for our present purposes, be interesting. Because it would show that the border between reflex reaction to a stimulus and action for a reason is not sharp. There are "borderline cases".

The relation between stimulus and response in the case of a reflex is *causal*. The sound, the physical phenomenon, in our fancied example, is the *cause* which *effects* another physical phenomenon, the turning of a head.

An effect need not infallibly follow in the presence of the cause. "Counteracting causes" may interfere. But if we are to uphold the statement that two phenomena are "causally related", there will have to be a fair amount of regularity in their successive occurence under what we would characterize as "normal circumstances". The best test of a causal relationship is an *experiment*, when under controlled circumstances the supposed cause is reproduced and we observe whether the supposed effect follows, or not.

7.

To establish that there is a causal relation between a sound and a turning of a head does not require knowledge of (neuro-)physiology. Reflexes such as the *patellar*- and *pupillar*-reflexes were probably known long before one had any knowledge at all of the function of the nervous system – or even of the very existence of a nervous *system*. The causal relation here is between two macroscopic events in the material world, a stimulus affecting a living body and the body's response to it.

A new chapter in the study of these phenomena was opened by Descartes. Not only was he the first to create a theory of the rôle of the brain and the nervous system. The theory was a fantasy or a vision rather than what we would call science. But an ingredient of Descartes's theory which eventually found a place in science too was the notion of the reflex-arch. A stimulus which from without affects the body – usually one of the so called sense organs – causes an ingoing process in the nervous system, followed by an outgoing process which effects an "outer" reaction – usually in the form of movements of the limbs or other parts of the body. The stimulus and the response are macroscopic or *molar* phenomena; the ingoing and outgoing nervous processes I shall call "microscopic". The latter, so to speak, "connect" or "bridge the gap" between the former. Knowledge of the connecting link(s) can be said to *explain* or to make us *understand* the causal relation between the stimulus and the response. But this relation is not *conceptually dependent* on its neural explanation. Even if the details of the explanation turned out to be completely false – as with Descartes who thought of the nerves as a kind of bloodvessels – the reflex is there and is as good an example as one could wish for of a causal connection in nature.

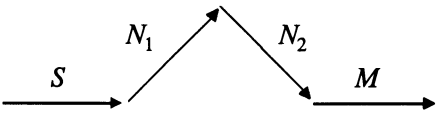
8.

We have entertained as an hypothesis that the turning of *P*'s head was a reflex on the sound. But this hypothesis is, in most cases, false. At least in grown up human beings of our culture this kind of reaction is *not* a reflex. There is no regular, next to infallible, correlation between the occurence of a sound near a person's body and a turn of his head. But there is another way of explaining and understanding the reaction, *viz.* to conceive of it as a result of an *action* for which the sound was, or gave, a *reason*. Like the cause-effect relation between a stimulus and a response, also the reason-action relation between, *e.g.*, a person's hearing a sound and his turning his head, is conceptually independent of anything that happens to or in his nervous system. One could say that to the case now under consideration – whether reflex or action – it is *essential* that there should be a sound and a head-turn, but *accidental* that there is a nervous system "linking" the two events in the material world.

(On occasion it may happen that there is no sound but that the person "thinks he hears a sound" or "seems to hear a sound" and reacts to it by turning his head. He perhaps hallucinates. But this reaction would not occur unless the agent had already learnt on many occasions, when there *is* a sound and he *hears* it, that there is a reason for reacting to it. The cases of "mis-hearing" must be marginal. This is a *conceptual* observation.)

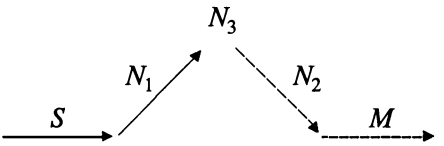
9.

Let the following schematic picture illustrate the case when the head-turn is a reflex reaction to the sound:



There is an ingoing nervous process N_1 following upon a sensorial stimulation by a sound S, and an outgoing nervous process N_2 terminating in a motoric reaction M, *viz.* turning the head. The two nervous processes are the ingoing and the outgoing branch of a *reflex arch*.

The following picture shall illustrate the case when an agent turns his head in response to hearing a sound:



The ingoing and the outgoing branches of the arch are here separated at the top. The gap between them in our picture, N_3, shall represent the fact that the "switch" from sensorial to motoric is not automatic (a reflex) but has, somehow, to be "sieved" or pass through a "screen" of controlling factors. The "screening" sometimes results in an outgoing branch, like in the case of a reflex. But sometimes it does not result in this. The "switch" fails to take place, the passage or transition from sensorial to motoric being "blocked" or "inhibited". This is why I have indicated the outgoing part, which is sometimes missing, by a dotted line.

The terms of my description of what happens in the brain are highly metaphoric. On the level for acting for reasons, however, they have a clear literal meaning. The controlling factors which determine whether there will or will not be a motoric response to the sensorial input are the reasons which the agent has for or against the action. But what do the metaphors mean in strict neurological terms? To which neural states or processes, *if any*, do they allude?

I must here warn of some misunderstandings in my picture. That there is a gap in the second picture separating the end of the sensorial line from the beginning of the motoric one should not be taken to mean that there is a definite *region* in the brain, where the "screening" of the ingoing material or the "switch" to an outgoing branch takes place. Perhaps there is such a locality, perhaps not. Perhaps whether a "switch" takes place or not depends upon

something which should better be called "the total state" of the brain at the moment of the sensory affectation. The gap in the picture only signifies that there is *something* in the brain which is responsible for whether the motoric reaction will, or will not, follow. Here "responsible" means "causally responsible". That the responsibility is causal again means that one can, in principle, specify a set of neural states or processes which are there when the "switch" takes place, and another set for the case when it does not take place, thus "blocking the passage". (The two sets may also have members which can alternate for each other, so that – for example – on some occasion when the "switch" occurs, the set contains m and on some other occasion m'. One can think of this as a reflection of the fact that different reasons may be "operative" on different occasions. And similarly for the cases when there is a "blocking".)

10.

Is the fantasy behind our second picture – when the bodily movement is the result of an action and not a reflex – realistic? If not, this would mean that there is no *neurological* explanation at all for the fact that sometimes an agent's body reacts, sometimes not, to a certain sensory stimulus. The explanation in terms of reasons for an action which we gave, and usually consider satisfactory, would then be an "invention" of ours which makes comprehensible the successive occurrence of two events in nature between which, however, there is no natural connexion. Could this be the whole truth in the matter? Something, surely, speaks strongly against this. If there is a connection in terms of reasons between two mental events of hearing a sound and intentionally turning (or failing to turn) the head, must there not also be a connection in terms of neural states and processes between the two "parallel" material states of a sound affecting the hearing organ and a movement (or failure of movement) of a head?

Let us consider what an affirmative answer to this question would amount to. Reasons why the agent reacted to the sound by turning his head might have been, we said, that he got frightened (startled) or that he was curious or that he was expecting somebody or something. Reasons why the agent, on another occasion, did *not* react could have been that he was indifferent to the sound or was tired or was concentrating intensely on something else. Could not all these different "states of mind" be reflected in characteristically different neural states? The idea seems (to me) very natural, empirically plausible – maybe even logically compelling.

Let us suppose that these different neural states had been so well specified that a physiologist, by observing P's brain could *predict* his reaction to the sound. He says, for example, "What I observe shows that P is concentrating so hevily on something that he will *not* react to the sound". Or "My observations

indicate that he is tensely expecting something, and so he *will* be alerted by the sound".

What we are supposing is thus that the things which count as reasons for or against an agent's action are, somehow, "reflected in" or "corresponding to" identifiable states and processes in his nervous system (brain). What the agent himself describes as reasons, the neurophysiologist describes as facts about the agent's neural system. The neurophysiologist gives a *causal* explanation of how it happens that a sound affecting a person's hearing organ resulted in his head turning. The agent himself gives what is often called a *rational* explanation of why his hearing a sound resulted in his turning his head in the direction from where the sound came.

11.

This correspondence between a causal and a rational explanation of a chain of events beginning with a sound and ending with a bodily movement touches the core of what is known as *psycho-physical parallelism.*

Is psycho-physical parallelism true or not? Before trying to answer this question one should ask whether psycho-physical parallelism is *possible*, or not. If the answer is that it is not possible, it follows that it is not true either. But from the fact that it is possible it does not follow that it is also true.

The question whether psycho-physical parallelism is possible I understand to be a conceptual (logical), and not an empirical, question. And I do not know of any argument to the effect that the parallelism in question were a logical impossibility. Nor can I think of any myself.

If the idea is possible, it is either true or false. If it is false, it must be *contingently* false. This follows from it's being *logically* possible. But if true, it can be this either contingently or necessarily.

I find the idea of psycho-physical parallelism plausible. If it were not this it would hardly have had so many occurrences and revivals in the history of thought – up to present times.

But the question whether the idea, if true, is a contingent or a necessary truth is intriguing. *In some sense* its truth must be a contingency. Man was familiar with mental phenomena – perceptions, sensations, thoughts, emotions of all kind, *etc.* – long before anything was known about what happens in the brain, and before one knew that there was such a thing as a nervous *system.* This is empirical (scientific) knowledge of relatively late date. I think that we can imagine that things had been quite different from what we now know they are in fact. Is it not thinkable that everything in the world of the mind, including intentionality of behaviour, went on in the way it does even if the brain and nervous system did not even *exist*? Some would perhaps say that this is logically impossible. But I am not sure. And all would agree, I think, that

what we *know* about brain-mind-correspondence – sensoric and motoric nervous processes, localization in various centres in the brain, *etc.* – is *contingent* knowledge.

To argue for the necessary truth of psycho-physical parallelism seems to me a hopeless undertaking. But there is a way to argue for its *a priori* nature. This would make it resemble another well known idea from the history of philosophy and science, *viz.* the idea of Universal Causation (Determinism). The belief that every natural phenomenon has a cause or sufficient condition may be viewed as a demand of the reason, urging us always to search for the cause or causes of whatever happens in the material world. (Some may wish to extend this demand also to the world of mental phenomena.) As such a "research programme" the idea has been one of the most powerful and profitable guides to progress in science – notwithstanding latter day doubts about its veracity in the micro-cosmos of the atom. In a similar manner, it seems to me, can the idea of psycho-physical correspondence or parallelism be regarded as an urge continuously to search for neuro-physical counterparts in the brain and nervous system to phenomena we call mental or psychological. Not least in recent decades has this project turned out rewarding – even if not as philosophically revolutionizing as some enthusiasts seem to think.

12.

Whether the brain-mind correspondence is necessarily or contingently true or false, the *details* of the correspondence can only be found by scientific, neuro-physiological, research. How this research is (or would have to be) conducted I, needless to say, know next to nothing about. But the following observation on some of the "philosophical" complications connected with it seems to me of interest to present here.

In trying to establish a correspondence between N and the reasons R, or rather N_3 and R, our neurophysiologist must rely on a careful study of numerous past cases. Which reasons for or against an action an agent has he would learn principally from the agent's verbal reports. These he would then have to correlate with his own observations on the agent's brain. When, for example, P attributes his turning of the head to his being frightened, the psychologist would notice some feature or features F of P's nervous system; when P says it was because he felt curious, O would note another thing F', and so forth. When the observations have become well established and the agent says that he turned his head because he was frightened, our neurophysiologist might retort: "You are lying! I can see that you are not frightened – you are curious". And P may have to agree. Perhaps his motivation had in fact been mixed – and this was the reason why the observer's question to him "fired"

the inaccurate verbal response. The neurophysiologist's knowledge can thus, on occasion, serve as a "lie-detector"!

There has been much discussion of the question whether the correspondence or parallelism, if it exists, is between *types* of psychological states and neural processes – or between *tokens* of such states and processes. The discussion has seemed to me confused. Practically every psychological state – say of anxiety or belief – has a great many "shades" which we can distinguish and sometimes also name on the basis of the subject's verbal and other behavioural reactions. It is plausible to think that these shades are reflected in characteristic differences between corresponding neural processes. On the other hand, in order to establish empirically the correspondences we must be able to specify a set of neural processes all or some of which are there, when we – on the behavioural observations – attribute to a subject a certain psychological state, and which are characteristically absent when the subject is *not* in that state. In the brain thousands of things go on all the time, and in order to know "where to look" to discover a "counterpart" to a given mental phenomenon we must have a sufficiently general characterization of some neural phenomena to be able to find out, after repeated experiments, whether the latter are there when the former occur – and missing when the former are not there. This does not exclude the set of neural phenomena from containing "alternative members" – answering perhaps to different shades in the psychological state. So, although the body-mind correspondence is not – can hardly be expected to be – between *types* of bodily and mental phenomena, the phenomena must exhibit enough "typical features" to be identifyable as phenomena falling under such and such types of mental and neural phenomena. A pure token-token correspondence either is a triviality – when something happens in the mind something also happens in the brain – or has to be dismissed as nonsense.

13.

In the phantasy in which we are now indulging, there is also a second psychophysical parallelism in addition to the one, the beginning and end-points of which are the sound and the movement on the one hand, and the sensation and the action on the other. This second parallelism is constituted by *O*'s questions and *P*'s verbal reactions on the one hand and *P*'s hearing the questions and *O*'s understanding the replies on the other hand.

How is this second parallelism established?

O asks *P* "Why did you turn your head?". *P* replies "I wanted to know what the sound meant". "Materially speaking", *O*'s question is an acoustic stimulus of *P*'s hearing organ – and *P*'s reply an acoustic response produced by movements in *P*'s speaking organ (tongue, chest, mouth, *etc.*). "Mentally speaking", they are a sensation and an action respectively. *P*'s reason for

replying can be simply that he heard the question – just as hearing a sound can be a reason for turning one's head in direction to its source. But this presupposes that *P* is familiar with the "game" of asking and replying to questions. This he would also have shown if he had replied to *O*'s question by saying "This is none of your business" – or by simply keeping silent (intentionally) and then, perhaps in reply to a question "Why do you not answer?" have added "It is none of your business".

One could also say that what makes the acoustic phenomenon a reason for *P*'s reaction to it (either by answering or not answering *O*'s question) is his *understanding* the question. Understanding, to be sure, is something mental. Understanding the question presupposes, not only familiarity with the practice of asking and answering questions, but also mastery of the particular language in which the question is framed.

In this case too, as in the one when the sound "called forth" a turning of the head, there is a "region" (not necessarily a specific "location") in the brain where the "switch" from an ingoing sensory stimulus (*O*'s question) to an outgoing motoric response (*P*'s answer) takes place. In the structural and functional features of this region is "encoded" the neurophysiological counterpart of that which, speaking in mentalistic terms, makes *P*'s *understanding* of the stimulus a *reason* for his reaction to it.

At the sub-human level, we cannot test a hypothesis about the neural basis of an animal's reaction to a sound by taking note of verbal responses to questions. Here we would have to rely on more "primitive" behavioural observations. Did the animal react to the sound because it got frightened or because it was curious? The question makes sense in all those cases in which we can distinguish between characteristic forms of animal behaviour expressive of fright (agressive and defensive reactions, *etc.*) and of curiosity (exploratory behaviour, *etc.*) respectively. When the question can be answered on the basis of observations on the molar or overt behavioural reactions of the animal, we can also "in principle" correlate the findings with observations on the microscopic level of the animal's nervous system.

The fact that in the case of humans we can rely on verbal reports makes the human case more complex than the sub-human ones. But also with humans we rely on data which overt behaviour affords. This fact is of crucial importance when we are in doubt and wish to test the *veracity* of the verbal responses.

14.

Before proceeding, let us once again restate the idea of psycho-physical parallelism in our example:

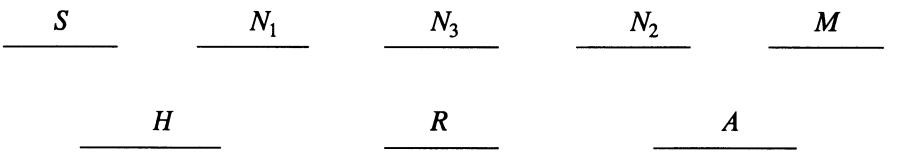
There is a temporal sequence of events and states in the material (physical) world. It begins with an occurence of a sound (*S*) and terminates in a movement

(*M*) of the head of a human body. "Mediating" between them there are states and processes in a neural system (*N*). Some of the states came to obtain already prior to the occurence of *S*. The relations between the temporally successive links in this chain (sequence) are causal (cause-effect) relations.

Secondly, there is a sequence of mental (psychic) events and states. It begins with an acoustic sensation, *viz.* a subject's (*P*) hearing (*H*) the sound *S* and terminates in the subject's performance of an action *A*, *viz.* the movement *M*. "Mediating" between them there are the reasons (*R*) for which the subject acts, some of which he may have had prior to hearing the sound. The relations between the links in the chain of mental things are rational (ground-consequence) relations. Be it observed that the reasons *R* need not be anything of which the subject is, as we say, "conscious" when reacting to the sound.

The two chains are correlated in their respective extreme links: the sound *S* with the acoustic sensation *H*, and the bodily movement with the action *A*. We have also assumed that to a description of *R* linking *H* with *A* will correspond something in *N* or the link between *S* and *M*.

Pictured in the form of two lines, the sequences are



The upper line pictures a chain of material events, from a sensory stimulation by a sound *S* to a motoric reaction *M*, connected by an ingoing nervous process N_1, a screen of controlling factors N_3, and an outgoing nervous process N_2. The lower line pictures a chain of mental events, from an acoustic sensation of hearing *H* to an action *A* mediated by reason *R*.

15.

Our task is now to investigate whether, in addition to the "parallelism" of the two chains (and the relations between the links in each of them) there is also a relation of some kind between the pairwise correlated links of the two chains of material and mental events (processes) and states.

We begin with the pair *S-H*.

It is natural to think of the acoustic sensation as an *effect* of the sound (as cause). The sound when "hitting" or "entering" the ear (hearing organ) "calls forth" an acoustic sensation. This establishes an asymmetric relation between *S* and *H*. This relation also involves temporal asymmetry. *First* the sound, *then* the acoustic sensation. This is so, we think, because what the sound immediate-

ly calls forth is a neural process which, when "propagated" from the ear to the brain (cortex) results in the subject's hearing the sound. This nervous process is a part of (belongs to) the link in the chain "mediating" between the sound and the movement. So what a sound, a physical phenomenon, in the first instance causes is something neural, *i.e.* another physical (material) phenomenon. Shall we say that it is this second physical phenomenon which causes the sensation – and that the sound is only – as philosophers would say – "remote cause" of it?

We already introduced the symbol N_1 for the nervous impact of S. Should one now say that S causes N_1 and N_1 causes H and that, therefore, by transivity S causes H?

Before answering, it should be noted that, whereas the impact of the sound on the nervous system is uncontroversially *causal* – one physical phenomenon S causing another N_1 – the relation between the nervous process(es) caused by S and the "resulting" sensation is not at all clear. A supporter of psychophysical *identity theory* would say that the neural effect of the sound and the hearing of the sound, the acoustic sensation, are *the same* (identical). This is a strange thing to say. But it is also symptomatic of a feeling we have and which I think is basically sound, namely that the relation between the neural process and the sensation is *not* a causal relation – or at most a causal relation "of a very special kind".

A "conceptual gain" of identifying the sensation with something neural would be that then the sound uncontroversially *is* a *cause* of the sensation. And this, surely, agrees with a common way of thinking about these matters.

So, what speaks *against* the identification of the acoustic sensation with the neural processes called forth by the sound? I shall mention two counterarguments, one rather trivial and another deeply "philosophical":

The "propagation" of the impact of the sound(-waves) along the hearing nerve to the brain centre cannot very well "be" the sensation. The process in the nerve may, for all I know, be interrupted – and then there is no sensation. The sensation is, somehow, the "reception" of the process by the brain. What is this? Presumably the process results in something which can be called a "brain-state" in the hearing centre. If the sensation is identical with anything neural it must be with this central state and *not* with the whole effect called forth in the neural system by the sound. In other words, the sensation cannot be identical with the whole of that part of N which we have called N_1, but at most with the "tip" of N_1 in the centre.

The second counterargument is this: A brain-state exists in *space*. It is, "in principle", open to inspection from "outside". A physiologist can "register" it on instruments. But the sensation, we think, is not "in space". So what the physiologist registers cannot be the sensation "itself", only its accompaniment or correspondence in the brain. The *sensation* is registered only by the hearing

subject. He, and only he, *has* it; it is his "private property", hidden in his "mind" like in a safe which is inaccessible to inspection by an outsider. Therefore it is absurd to identify the sensation with a bodily, "material" state.

This counterargument must surely be taken seriously. We feel that there must be truth in it. But it is also obscure and mystifying.

We are facing a dilemma. To regard the sensation as an effect of the sound would have a clear meaning, if we could identify the sensation with that which uncontroversially *is* an effect of the sound, *viz.* certain nervous processes and states. But this identification, on the other hand, seems an absurdity, confusing with each other things of conceptually utterly different nature.

16.

Before returning to this unresolved dilemma, consider next the second pair of links in the parallel chains: the neural processes and states N_3 which correspond to the agent's (subject's) having reasons R for and/or against the action A. What is the relation between R and N_3, other than their simultaneity? Does the subject have the reasons because his brain is in such and such a state – or is his brain in this state because he has such and such reasons for an action?

If we ask, *why* the subject has the reasons he has, a partial answer is that he has *learnt* to recognize a sound as a reason for reacting to it. He has learnt the "meaning" of a knock on the door and the ringing of a bell, and also learnt that a sound may signal something which it is in his interest to attend to. But these affectations on him from the outer world would not – or so we think – have resulted in that he learnt to appreciate sounds as reasons for doing something, unless they had also made a modifying impact on his neural system. They must have left some traces in it, if not of a structural so at least of a functional kind, so that, when the subject is alerted by a sound, these neural correlates of his (having) reasons for doing, or not-doing, an action are, somehow, "activated".

On this view, the neural processes effected by the sound on the hearing nerve are the causes of the activation of that "part" of the neural system, call it N_3, which "mediates" between the sensory input caused by the sound and the motoric output effecting the turn of the head. But the structure itself of N_3 is a product of factors working on P's nervous system since times *anterior* to the occurence of the sound S. The learning process which made P understand the meaning of a sound as a possible reason for action probably goes back to P's childhood. But also such reasons for the action as believing something or expecting somebody or, against the action, as being tired or concentrating on some task were, for the most part, there already *before* the sound was heard. This is why I said that the sound (or hearing the sound) "activates" the reasons for or against the action which were already implanted (encoded) in P's nervous system (brain).

But is it right to say that the neural state N_3, thus having been "built up" and now activated *causes* the subject to have certain reasons R for doing something? N_3 is physical, R we say is mental. Can something physical cause something mental? The difficulty is the same as the one we were already facing when talking of the relation between the processes in the hearing nerve and centre *and* the acoustic sensation (the hearing "itself"). We noted a temptation to say that the two are, somehow, the same, identical. There is a similar and it seems even stronger temptation to say that the subject's having certain reasons for action and his brain being in a certain state are "the same thing", – the same "ultimate reality" seen from two different points of view. And we feel, just as in the case of N_1 and H, that there is *some* truth in this – but also that we are in danger of confusing things of fundamentally different categories, *viz.* something spatial which is accessible to "outside" observation, and something non-spatial which is strictly "private".

17.

Finally let us take a look at the third and last pair of corresponding links in the two chains, M and A, the bodily movement and the action. How do they differ? It may be difficult to perceive any difference between them at all. Is not the agent's turning his head and his head's turning exactly the same event "in the world"? In some sense this is so. What makes the head's turning be an action, we said, is its intentionality. But what is this? Is intentionality some *additional* feature of the movement – *quasi* a "mental accompaniment" of it – which distinguishes intentional from non-intentional movement?

To try to construe the intentionality of the movement in this way would be seriously misleading. The intentionality resides in the fact that the subject can, if challenged, account for the movement in the terms of some reason or reasons why he performed it. For example, by saying that he was curious or frightened. If his reaction to the challenge had been "I do not know; my head just turned", the movement would not have been an action, intentional. The intentionality is thus something which, so to speak, is "embedded" already in the middle (penultimate) link R of the chain and therefore is antecedent in time to M. This fact, I think, is responsible for the idea that in action something mental (traditionally referred to as the "will") is a *cause* of some movements of the body. But this is a mistake. One can call the reasons for an action the causes of the action. (Above Sect. 3.) This is innocuous talk. But it does not make the reasons cause the bodily movement. Its cause is *neural*. It is (in) that outgoing part of the tripartite chain of neural states or processes which I have called N_2 – and which in its turn is causally connected, through the central N_3, with the ingoing part N_1 and ultimately with the sound S which put the entire chain of physical events "in motion". Thus, by transitivity, the sound S becomes the

remote physical cause of the movement of the head *M*. This is how we naturally express ourselves when talking about such things – and it is in perfect (conceptual) order.

The following additional remark on intentionality may be called for. In order to call a movement intentional it is not necessary that the subject should have formed an intention to perform it before actually performing it. *Sometimes* this happens. The agent deliberates about the reasons for and against doing something, and then reaches a decision. He perhaps thinks or says to himself: "OK, I'll do it". Then the intention can be called a genuine addition to (the previous) *R*. And to it will correspond on the neural side an addition to N_2 or N_3 (depending upon how we do the division).

But this is not how things *have* to be when we act. In many, perhaps most, cases the agent just *has* some reasons and then acts – and only in retrospect, if at all, does he reflect ("think") on them and makes it clear to himself or to others *why* he acted. This self-reflective process can be called "rationalization". *Sometimes* it consists in the agent just inventing, *post hoc*, reasons which in fact he did not have. He may do so in order to deceive others about his "real" motives – but also, on occasion, to deceive himself. These phenomena are of great interest to the psychology of action – but they must not be allowed to obscure the picture which we have drawn here of the *M-A*-relation.

To sum it up: *A, the action, is M, the bodily movement, viewed* (conceived, understood) *under the aspect of intentionality*. Viewing *M* under this aspect means relating it to the mental things *R* we call reasons for an action. This relation is not causal – although the fact that the reasons antedate the movement may create an appearance to the contrary.

18.

We now return to the dilemma we were facing when discussing the relation between the acoustic sensation (*H*) and the neural effect (N_1) of a sound *S* and the relation between the reasons (*R*) which a subject has for an action (*A*) and the "encoding" of the reasons in that "part" of his brain (N_3) through which the neural effect (N_1) of the sound has to be "sieved" if it is to result in the neural cause (N_2) of a bodily movement (*M*). The dilemma was this: Is the relation between N_1 and *H* a causal relation ("material cause – psychic effect") or is it an identity (the sensation, somehow, "being" the neural process or state)? And simlarly for the relation between N_3 and *R*. Both conceptions, the causalistic and the identity-theoretic one, have a certain intuitive plausibility. Both, on the other hand, seem guilty of some conceptual ("metaphysical", "ontological") confusion.

In order to find our way out, I think an *epistemic* move is needed. Let us ask: How does one establish, come to know, that the subject has a sensation and that it has certain reasons for or against acting in a certain way?

19.

Take sensations first. Let the example again be the hearing of a sound. How can one establish the (*objective*) *truth* of the statement that a subject hears (heard) a certain sound?

The perhaps most common method would be to ask him. "Did you hear?" His answer, "Yes" or "No", is a behavioural reaction to the question, *i.e.* to another sound. The reaction is a further acoustic phenomenon, *viz.* a sound produced by the answerer. Since his reaction presumably is intentional, the sound produced by him is the result of an action.

The case when we attribute a sensation to a subject on the basis of a verbal response of his to a question has conceptual peculiarities which complicate the understanding of its nature. (See above Sect. *13.*) We can simplify the situation by imagining an animal subject other than human. How do we know that *it* heard a certain sound? In the normal case, by observing its behavioural reactions. We note, for example, that the animal turned its head in the direction of the sound or otherwise took a listening posture, or that it leapt away as though being frightened or that it ran towards the source of the sound as if anxious to explore what it was. (Perhaps it recognized in the sound the call of another animal.)

Similar non-verbal tests apply also to the human case. The verbal reply may be a lie – the subject perhaps wants to simulate deafness. Then we must look for another reaction which "belie" his words. But also non-verbal reactions can, on occasion, be deceptive – and the deceit can be intentional or non-intentional on the part of the "deceiver".

But could one not also rely on a *neural* test for finding out whether the subject heard the sound, *viz.* by making observations on his hearing organ and/or hearing centre in the brain? Of course, one could. But this *presupposes* that one has already established a correlation between acoustic sensations and specific neural processes and states. And in doing this, one would have to rely on already established facts ("objective truths") about subjects' hearing, or not hearing, sounds in their surroundings. These facts (truths) have their criteria in other macroscopic behavioural reactions, including verbal responses to questions. The neural criteria will therefore necessarily be *secondary* in relation to the criteria provided by the subjects' overt behaviour. The latter are the *primary* criteria. This is a fact of fundamental importance to our argument.

20.

But is there not also another way of establishing whether a person hears a sound or not? A so to speak "subjective" way? It relies, *not* on the fact that the person *says* he hears it (because he may be lying) – but on the fact that he *hears* it, *has* the acoustic sensation in question. Whether *this* fact is there, or not, is, in the last resort, known only to the subject itself. He knows for certain whether he hears the sound. *How* does he know this? Just by having, or not having, the sensation!

One sometimes says that the person knows this by *introspection*. He, as it were, "looks into his mind" and notices a fact, accessible only to *his* observation, not to that of an external observer. It is as though he could identify an object "inside himself" and say of it "yes, this is an acoustic sensation, I have it". But this is, of course, only a metaphor.

Introspection is a queer idea. What is true about it, is this: One can observe, watch one's sensations. The observation may result in what I shall call a *phenomenological description* of them. A sound, *as heard by me*, may be loud or weak, shrill or dull, piercing, melodious or not melodious; it may be like a whistle or a drum-beat, or rolling like thunder, *etc., etc.* Another person may hear physically the same sound, but give a different description of his impression of it. For such differences in the phenomenological descriptions it may be tempting and sometimes also possible to give an account in "objective" neurological terms relating to the hearing capacity of the subject.

What "introspection" can show is *how* a subject hears a sound. It does not establish *that* he hears it. Because this latter is already *presupposed* in the introspective act.

That a subject hears a sound is a fact about him. *Its* criteria are behavioural, "objective". The hearing subject itself does not rely on them for coming to know that he has the sensation. It is *having the sensation* which "establishes" it for him. In this asymmetry between the sensing subject and the external observer lies the peculiar "subjectivity" of sensations – and other mental phenomena.

21.

We now proceed to the question, how one establishes, comes to know, that a person has certain reasons (R) for or against an action. The reason for reacting to a sound, we have said, can be curiosity or fright or the fact that one is expecting something or somebody. The reason for not reacting can be indifference or tiredness. Examples can easily be multiplied.

When human subjects are concerned, the perhaps most commonly used criteria on which the attribution of reasons takes place are the subject's own verbal reports. We ask, for example, "Why did you turn your head?" and

expect a reply. Here too the possibility of a lie has to be taken into account. If we suspect a lie, we would rely on other behavioural reactions of a non-verbal nature. They would often have to extend beyond the present situation. If, for example, the subject says he was "too tired" to react and we would find his answer doubtful, we would ascertain what, if anything, he did just before which could have thus exhausted him. In the case of non-human subjects we would have to rely exclusively on such non-verbal criteria.

As in the case of sensations, the subject itself does not – except perhaps in some marginal cases – rely on criteria for determining what reasons he has for or against an action. He simply *has* them. He need not watch himself walking back and forth, looking repeatedly at his watch or making faces signifying annoyance or impatience in order to determine that he is expecting somebody. Shall we say that what he thus *has* (or "is in") is a mental state (of expectation)? We can do this, although the locution is much less inviting here than in the case of having a sensation. And it must not make us forget that the criteria we have for establishing this fact, *viz.* that he is expecting something, are behaviour of the kind we just mentioned.

Could the behavioural criteria include also neural states and processes? Obviously they could. Being frightened or being tired or impatient are surely states with internal somatic "correlates" – whether we call them "behavioural" or not. But *these* bodily (behavioural, neural) criteria are secondary to the mental phenomena (of fright or tiredness or impatience) themselves. For example: that certain hormonal reactions are characteristic (accompaniments) of fright or tiredness, we have found out from observations of frightened and tired subjects – men and animals – and in order to be sure that the subjects in question *were* frightened or tired we would have to rely on *other* overt behaviour, and *not* on neural and other "hidden" criteria. We must already know, both *what it is* to be tired or frightened, and *whether* a being is, or is not, in this state, before we can establish that these states also have such and such internal "equivalents". This basic truth, however, is compatible with the possibility – albeit a marginal one – that, having found out these internal correlates, we may appeal to them as decisive grounds for a verdict in controversial cases. "The animal *must* be frightened considering the amount of adrenalin the test shows – although, strangely enough, it exhibits none of the typical fright reactions at all." This may be true. But it does not make the intracorporeal state a defining criterion of the state we call being frightened.

22.

The relation between a criterion and that of which it is a criterion is a *semantic* relation. The criteria of *X* determine the meaning of the term "*X*". Thus, for example, the presence of the behavioural criteria of hearing do not only

"indicate" but *mean* that the subject in question hears something. Similarly, the presence of the behavioural criteria of expectation or of curiosity *mean* that the being expects or is curious about something.

The criteria of sensations and other mental phenomena have the following peculiarity: No finite combination of them is (logically) *sufficient* to establish the presence of the phenomenon. On this is founded the possibility of feigning or simulating mental states – for example by an actor on the stage. Moreover, no finite combination of behavioural criteria is (logically) *necessary* for the presence of the sensation. On this again is founded the possibility of suppressing (at least up to a point) the reactions to a stimulus. I can react as though I heard a sound without hearing it – and also not react to it in any way although in fact I hear it. For this reason I shall say that the attribution of a sensation or another mental phenomenon to a being has a *residue of meaning* which is not captured by any enumeration of the behavioural reactions which are constitutive of its meaning. It is this fact about the criteria which accounts for the idea that mental states are, somehow, "hidden" or "private" and therefore not *identifiable* with the "overt" or "public" phenomena (in the physical world) on the basis of which we attribute these states to living beings.

23.

The criteria of the presence of the three links in the chain of mental phenomena *H-R-A* are thus overt behavioural features and reactions of the being of whom those phenomena hold good. This behaviour is like a "penumbra" surrounding the chain itself. It is partly anterior in time to the beginning of the chain – as when it relates to things which the being in question has learnt or to which it has reacted in the past. But it is partly also posterior in time to the end of the chain – as when it makes conjectural reference to the being's subsequent behaviour (to show, say, that the sound frightened him).

Are these criteria of mental phenomena to be called themselves "physical phenomena"? In their being overt behaviour they are events in *space* and *time*, open to intersubjective inspection from "outside". Events (in space and time) are not exactly what we call "material". But they have, normally, what may be called a "material substrate" in the form of matter in motion (the turning of a head, for example). In this sense they can be said to "belong to the material world" and to be "physical phenomena". But the aspects of them, in which we are interested when taking them as signs of something mental, are not those which a physicist or even a neurologist studies. The behavioural "penumbra" of mental phenomena has primary interest only to the psychologist.

The criteria of the three links in the chain of material (physical) phenomena *S-N-M* are wholly "inside" ("interior to") the chain itself. The two extreme terms, *S* and *M* (the sound and the bodily movement) are *macroscopic*

phenomena which we register with our sense organs, in the example: hearing and seeing. *N* (the neural processes) are "microscopic" in the sense that their registration presupposes sophisticated techniques of "indirect" observation. But they have none of the "privacy" characteristic of mental phenomena. They are public and intersubjectively accessible to inspection.

24.

That mental phenomena have epistemic priority in relation to their neural equivalents means that we possess criteria of the former which are independent of the latter. We must first know whether a person hears a sound or not before we can investigate what corresponds in his brain to his acoustic sensation. The criteria on the basis of which we establish the mental fact about him are behavioural. They belong to the same material or physical world in space and time as do the neural phenomena which we correlate with the mental ones. These two types of phenomena are causally related. Moreover: the behavioural reactions which are the criteria of the mental are caused by things which happen in the neural system of the bodies under investigation. This means that the neural phenomena have *causal priority* in relation to the behavioural phenomena on the basis of which we attribute mental states to a subject. There is thus a triple relation of priority involved in the picture we are drawing: the epistemic priority of the mental in relation to the neural, the causal priority of the neural in relation to the behavioural and the semantic priority of the behavioural in relation to the mental. Since the relation of the behavioural to the mental is semantic and not causal, we cannot by transitivity conclude that the neural is cause of the mental – nor, needless to say, the mental of the neural. With this observation one of the main difficulties in the traditional body-mind debate is overcome, *viz.* the apparent irreconcilability of body-mind interactionism with what we think of as a scientific understanding of natural phenomena.

Simplifying a little, one might say: Science investigates facts about the material (physical) world and their causal interrelations. In the nexus of these facts and their connections the mental has no place. No "ghost in the machine" must be allowed to disturbe the scientist's intellectual peace of mind.

25.

We are now at our journey's end. We have the two sets of events in the world of space and time, overt behaviour and neural processes, and between them causal relationships. We also have, like a "shadow" accompanying the first set of events, a sequence of mental states and processes, semantically connected

with it. In the picture of natural connections this set of mental phenomena plays no rôle.

From the history of philosophy is known a position called *epiphenomenalism*. It regards consciousness (the mental) as a kind of "by-product" of underlying neural phenomena – as a world standing apart from the physical world and neither having an influence on nor being influenced by it. My position has *some* resemblance with this.

I agree with behaviourism, epiphenomenalism, and various forms of materialism in rejecting the idea of mind-body *interactionism* as contrary to a scientific picture of the world. I differ from behaviourism and materialism in that I cannot accept an *identification* of the mental with states of affairs in the material world. But at the same time I have a certain sympathy with this monistic view of the stuff that the world is made of. The question, for me, is not in the first place whether this view is *true* or not but whether one can make it *intelligible*. To this end I shall make the following suggestion:

The mental phenomena or "Cartesian thoughts", *i.e.* our sensations, beliefs, desires, and volitions, the reasons we have for our actions, are how we, *as subjects*, experience that which happens in our neural system (brain). An observer can "see" this from the "outside" as a web of neural connections and register it as electro-chemical reactions, *etc.* We too can, in principle if not in practice, take part in this observational activity on ourselves. (There was a time when philosophers entertained fantasies about a "brain-mirror" in which we could watch what is going on inside our skull.) But our own sensations, thoughts, *etc.* are not the findings of observations, but experiences we have and describe in a language which makes reference, not to what happens in our brain and into which we normally have no direct insight, but to our macroscopic reactions to and dealings with things in the (material) world around us. One could therefore say that *we experience what goes on in our brain as that which its behavioural effects mean* – for example that the subject suffers pain or is in a state of expectation.

When seen against this background, the body-mind distinction appears in a new light. What seemed like a difference between two kinds of "stuff" (matter and mind) of which reality is constituted, becomes a distinction between two ways of looking at living beings. One consists in relating overt behavioural reactions to intra-bodily causes and effects. The other consists in understanding what these reactions *mean*.

A NOTE ON CAUSAL EXPLANATION OF BODILY MOVEMENT AND RATIONAL EXPLANATION OF ACTION

1.

From what has been said about the epistemic priority of the mental in relation to the neural it follows that the same priority also holds for rational explanation of action in relation to neurophysiological explanation of the movements involved in the acting. The causal hypothesis which explains the bodily movement *rests* on acceptance of the rational explanation of this movement when seen as the performance of an action. Therefore one cannot make the (assumed) causal connection between neural processes and muscular movement the *warrant of truth* of the rational explanation of an action (of the movement as action). This is what a "causal theory of action", as I understand it, tries to do. And this is what is erroneous about it.

One could say: the truth of the rational explanation is basic to the truth of a corresponding causal explanation. The existence of a causal explanation can be *postulated* for all cases where there is a rational explanation. The postulate has a heuristic value, encouraging "brain research". When one sticks to the postulate, the existence of a causal explanation parallel to a given rational one becomes (like a) logical necessity or an *a priori* feature of reality. This is similar to a view which some people have entertained about the Universal Law or Principle of Causation.

2.

It is probably right to say that the vast majority of actions which call for an explanation are cases when the agent can give just *one* reason why he did it. By studying numerous similar cases we may for a certain type of reason, *e.g.* fear, find corresponding types of characteristic neural traits which we then naturally come to think of as "releasers" of the type of motoric reaction which is characteristic of that type of action.

The case is logically more complicated, and from a psychological point of view more interesting, when an agent has several reasons for, and maybe also against, a certain individual action of his. In such cases both the agent himself and outside observers may be unsure about the rational explanation, or they may disagree or even altogether suspend judgement. Then there may exist

G. H. von Wright, In the Shadow of Descartes. Essays in the Philosophy of Mind, 149–150.

similar uncertainty and confusion also about the "parallel" causal explanation of the neural causes of observed bodily movements. But sometimes *neural* findings in other, simpler cases may help us to frame a rational explanation in terms of *reasons*. Such cases may then serve as a "lie-detector", if the agent is trying to deceive us. But this is a marginal possibility and it does not change the basic relation between the two types of explanation.

3.

Nothing which has been said is meant to deny or belittle the scientific interest in that which happens in our brain when we perceive, recognize patterns, remember, learn various feats, think, believe, desire or shun things, *etc., etc.* But in establishing neural correlates we rely on the psychological phenomena as given to us in experience. From the psychological point of view the very existence of a nervous system, "mediating" between the world of the body and the world of the soul, is, logically speaking, an *accident*.

This is not in conflict with what has been said earlier about the *a priori* character of psycho-physical parallelism. (For example, about the existence in the brain of "counterparts" to sensations, reasons, beliefs, and other introspectively recognizable mental states and processes.) But I have wanted to emphasize the *postulational* nature of this correspondence. Like the Law (Principle) of Causation it should be regarded as a "demand of the reason" or as a "heuristic device" which we urge ourselves to employ. But is it not also conceivable that we one day shall think of it as less urgent or stringent – and begin to take interest in cases where "parallelism" seems to fail? Did not something similar happen to the law of causation under the impact of physical theory earlier in this century? If one wholeheartedly accepts *pure chance* in nature one cannot at the same time postulate *strict determinism*.[1]

NOTE

[1] If the brain and what happens in it can be conceived of as a dissipative system in the sense of Ilya Prigogine, then neural relationships constitute what has been called a "deterministic chaos".

A NOTE ON TIMING CONSCIOUSNESS

1.

Mental things are not extended in space. But they may last over a period of time. This means that they come into being and pass out of existence *in time*. This is commonplace – but yet not unproblematic.

A stone hits my foot and thus "calls forth" a sensation of pain. *When* does the sensation occur? Let us ignore the fact that it may *last* some time.

2.

Also mental phenomena such as emotional attitudes and beliefs have a duration, they come to be and pass away, but it does not, normally, make sense to ask exactly when they began and ceased to exist. (Sometimes the question may have an answer.) Sensations, however, are different.

I am hit or pinched, and I feel it (feel pain). A sound affects my hearing organ (nerve), and I hear it. (Our previous *S* and *H*.) The subject is instructed to press a button *as soon as* he has the sensation. (Or to cry "Ow", or to nod his head.) A motoric reaction is needed if we are to measure the time when the sensation occurs.

Let us assume that the timing of *S* is unproblematic. If *S* is a sound, the assumption is that we can time the moment when the sound affects the acoustic nerve. We shall also assume that the pressing of the button can be timed with precision. (These assumptions are not unproblematic – for one thing because timing involves perception, *i.e.* the reading of clocks. But this we now ignore.)

S and *H* are not simultaneous. Between them is N_1, the nervous processes which "propagate" the impact of the stimulus from the periphery to the centre (pain-centre, hearing-centre). The propagation takes time. Therefore *S* and *H* cannot be strictly simultaneous. *H* comes later. If *S* occurs, *i.e.* affects the neural system, at *T*, the sensation *H* occurs, *i.e.* "enters consciousness", at $T+x$. Is this the time when the subject presses the button?

The button pressing results from a motoric, outgoing reaction. It originates from an outgoing impulse from the brain, *quasi* an "order" of the mind to the hand when *H* is there. The execution of the order is the result of the outgoing processes which we previously called N_2. They too take time. Let us call this stretch of time *y*. We shall assume that when this time has passed, the order

G. H. von Wright, In the Shadow of Descartes. Essays in the Philosophy of Mind, 151–154.

(the pressing of the button) has been effected. Thus it seems that the *registration* of H happens at time $T+x+y$, although H actually occured ("in the subject's mind") at $T+x$.

3.

The button pressing is the M of our previous schema *S-N-M*. Between S and M there are the ingoing neural processes N_1 and the outgoing ones N_2. But there is also a connection between them, our previous N_3. In the case under discussion, N_3 is not "innately" there – as when the motoric response to the stimulus is a reflex. N_3 has to be "built in" *i.e.* formed by a learning process apparently involving a modification of neural connections. The subject must understand the instruction to press the button when he becomes aware of S. The acquisition of capacities which are needed for the "switch" N_3 to function may be a long process in time.

That the ingoing and outgoing processes take some time is obvious and presumably this is true also of the "switch" from the one to the other. Call this last "time-bit" z. The time-interval between S and M, *i.e.* between the affectation of the neural system by S and the completion of the motoric reaction M, is thus of length $x+z+y$.

Where on this time-stretch does H fall?

A naive answer is that H *occurs* at the end of time x, when the sensoric process is consummated, or during time z when the transition from ingoing to outgoing, from sensoric to motoric takes place. From the point of view of *timing* the occurrence of H, however, the timing of M is crucial. The subject was to press the button as soon as he has the sensation.

If therefore looks as though the time of H must be *between* the time of S and that of M.

It "takes time" before the subject becomes aware of S and it also "takes time" before he reacts to it. Even if it were true that the time of H is somewhere between that of S and M there is no possibility of *locating it* exactly in this interval.

4.

The intuitively most plausible timing of H is when N_1 "switches" to N_2. The "switch" is what we called N_3. Whether N_3 "takes time" we need not try to decide. But surely N_3 occurs "at a certain time".

Suppose someone said: "When H occurred, the processes N_2 were already 'on their way'". But is this (logically) possible, considering that the subject had been instructed to inaugurate N_2 only when (after) he had registered H? I think it is possible. For consider the following: the instruction which the subject

receives (to press the button) is another stimulus affecting his neural system. Can we not imagine that it creates a "central state" which secures that the processes N_1 "automatically", *i.e.* independently of the subject's conscious awareness of what produced them, *i.e.* S, continue in the process N_2? I think we can imagine this. The central state is there before the experiment starts and it ensures that as soon as N_1, produced by S, is completed, N_2 leading to the reaction M commences.

Is it also thinkable that H actually occurs *after* the motoric reaction M is inaugurated (by the neural processes N_2)? This would mean that the subject "antedates" H to *before* the commencement of N_2, whereas in fact H occurs *after* the commencement of N_2.

This, too, seems to me thinkable, and possibly there even is experimental evidence for it. Because why could not the *instruction* which was given to the subject create a preparedness in him which is "released" in a motor reaction as soon as the sound S "put in motion" N_1? It should also be observed that the separation which I have made between the three successive phases N_1, N_2 and N_3 is purely hypothetical and only serves the purpose of a first "sorting out" of what happens in the brain when a sensoric input results in a motoric output. The three phases may also be temporally overlapping. Moreover, the second one, N_3, may in fact be "timeless", once a link has been established thanks to the instruction between the ingoing process inaugurated by the sound S and the outgoing reaction terminating in M. So that, instead of three separate processes answering to three "bits of time", x and z and y, we have only one total process within which three partly overlapping parts may be distinguished.

The sensation H can thus occur at any moment in the interval between the occurrence of S and the completion of M. *Perhaps*, it could even coincide with the terminating point of this stretch of time. But it cannot occur outside the interval, *i.e.* either before S or after M. The first would mean that the sensation H is not of *that* sound S which is supposed to start the whole process. The second alternative again implies that for the timing of H some *other* behavioural sign than M, the pressing of the button, is needed. But since M by stipulation, was to be the signal that the subject heard the sound, nothing which happens after M can possibly also have this rôle.

5.

The pressing of a button by the subject in response to a heard sound is an *action*. The action is performed for a *reason, viz.* the fact that the agent hears the sound. The sensation must be antecedent to or, at most, simultaneous with the action. This does not mean that the *sensation* is the *cause* of the *action*. If it could be established, which is at least possible, that the button-pressing actually commences before the subject hears the sound, this may be regarded as

a "proof" that the sensation (consciousness, something mental) is causally inefficacious, *i.e.* it has no causal rôle to play at all. This is, I believe, how some neuro-physiologists have interpreted certain experimental findings of theirs. In fact the truth that has been thus "confirmed" is – in my opinion – *conceptual*.

6.

We can compare the case of the button-pressing with the one discussed earlier, when the subject reacted to a sound by turning his head. He did this, let us assume, because he thought there might be a danger lurking in the direction from where the sound came. Here it *may* actually happen that the subject hears the sound *after* he has actually turned his head. This would mean that the head-turn which originally was not automatic, a reflex reaction to the sound, has with time come to be this. The explanation of this might be in the fact that it has repeatedly happened in the past that the subject reacted to the type of sound in question by turning his head because this was a "sensible" thing for him to do. Originally the sound functioned as a reason. Later it became a cause. But if the subject were to *notice* that, in fact, he heard the sound only after having turned his head, then he would not easily think of the sound as a reason for the head-turn. Perhaps he would now describe what happened by saing that his head turned reflexively or for some unknown cause and that this happened to be "a good thing", since now he could take some precautions which he would have been unable to take had he not heard the sound.

ACKNOWLEDGEMENTS

I am grateful for discussions with Professor Tom Reuter. He also directed my attention to the paper by Benjamin Libet and others, "Time of conscious intention to act in relation to onset of cerebral activity" in *Brain* **106** (1983), 623–642, and to Max Velman's paper "Is human information processing conscious?", with the discussion and the comments on it by several people in *Behaviour and Brain* **14** (1991), 651–726.

ON SOUND

1.

Sounds are physical phenomena. An individual sound occupies an area in space and time. It comes into existence and fades away.

Humans and other sentient beings are acquainted with sound in hearing. An individual sound may be heard by somebody but not by somebody else. And many sounds exist without anybody ever hearing them.

To hear a sound is to have an *acoustic sensation*. And *vice versa*.

Acoustic sensations normally result from the stimulation of the appropriate sense-organ by a sound external to it. But this is not always the case. Sometimes we have acoustic sensations which do not have their origin in a sound "from outside" but in processes in the ear, the cochlear nerve, and the brain. We often speak of them as acoustic illusions.

It also happens that a sound affects the hearing organ but is not followed by an acoustic sensation. This can be due to some defect in the organ (deafness). Or it may be due to a "blocking" in central parts of the nervous system. The subject is perhaps too intensely engaged in some other activity to "pay attention", as we say, to the sound. In order to hear, it is thus not enough that there be an acoustic stimulation of a functioning hearing organ; the "total state" of the nervous system (the brain) must also satisfy some conditions which are normally but not always fulfilled. One could call them audibility conditions.

2.

Whenever a subject has an acoustic sensation it can be said to hear a sound or to hear sounds. But normally when reporting what we hear we say something more specific than just that we hear sound. What sort of things does one hear then? For example: footsteps, the whisper of the wind, the roaring of the ocean, thunder, people talking, birds singing, a drum beating, the sound of a musical instrument, a concert, an opera.

(There is also an analogical use of "hear" which means, roughly, "has come to my knowledge" and does not necessarily involve acoustic sensations or sound. "'Did you hear the rumours about *X*?' 'Yes, I read about them in today's paper'".)

In most cases, what we say we hear is not one individual sound but a structured totality of sounds. This challenges the question what constitutes the

G. H. von Wright, In the Shadow of Descartes. Essays in the Philosophy of Mind, 155–164.

unity (identity) of the thing we hear when it consists of several individual sounds – *e.g.* a tune. And also the question what constitutes the identity of *one* individual sound.

The ocean kept roaring the whole night, but the sound was intermittent and sometimes it was louder and sometimes weaker; at intervals it abated completely. Was what we heard one sound or many? A person strikes the same key of a piano twice over. One sound or two? Questions like these can be answered the one way or the other, and there is no way telling which answer is *the* right one. But sometimes there is a reason why one would give the one answer rather than the other.

The same sound may not only be intermittent in time. It can also be shifting location in space. A person walks around whistling the same tune. We may speak of the sound he produces as *one*. But in his walking round, the location of the sound in space varies. Now you hear it where *you* are – I no longer hear it where *I* am. So although an individual sound, as we said, occupies an area in space and time, this area need not be rigidly bounded and stable.

I say, confidently, that I hear footsteps on the porch. But it turns out to have been the sound caused by an object which fell from a table and from there bumped onto a chair and finally onto the floor. The object of my acoustic *perception* was what, in fact, I heard: the bumps of the falling object. What I *thought* (believed, "meant") I heard were footsteps on the porch, someone stepping onto the porch. I shall call this the *intentional object* of my acoustic *sensation*.

When I report *what* I hear, I report the intentional object of an acoustic sensation. If subsequently I find that the intentional object was not also the object of a perception – was not what I actually heard – I correct my report and say that I *seemed* to hear such and such, *e.g.* someone stepping on the porch.

There are also other uses of seeming to be noted. If my report is hesitant only, I may say that *it seems to me* that I hear such and such, or that *there seems to be* such and such a thing going on to judge from the noise I hear.

Thus reports of intentional objects of acoustic sensations are often framed in the language of seeming to hear, also when the objects *are* the things actually heard, *i.e.* are objects of acoustic perceptions.

The intentional object of an acoustic sensation is something which *could be* (have been) also the object of an acoustic perception. In other words, for any intentional object of an acoustic sensation a sound could exist which satisfies its description.

I think this holds true also of acoustic sensations which are illusory in the sense of not being caused by any sound from a source external to my ears. I hear, say, something buzzing. But it is only "in my ear" and there is no

buzzing sound anywhere near me to be heard by a perceiving and sensing subject.

In the limiting case, the intentional object of an acoustic sensation is simply that I hear "something", a sound.

3.

When we have an acoustic sensation we normally "project" the sound we hear onto some thing in the external world which caused it. This is a fact about sensations generally. It is the biological function of the various sense organs to inform us of what is going on in our surrounding. The individual learns to orient itself in its life-world thanks to the information which it receives through the senses. Thus it soon learns, for example, to trace sounds which it hears to their sources in sounding things: animals roaring, people talking, instruments being played, *etc.* Thereby the individual learns the *meaning* of what it hears and this can be of vital importance to its survival and success in the "struggle for life".

In humans this learning leads to a successive evolution of the intentional objects of their acoustic sensations. Animals too understand the meaning of acoustic signals but it is questionable to what extent one can say that their sensations have "intentional objects". This is so because of their limited linguistic abilities. It is not clear, for example, what it means to say that an animal "reports" its sensations. But there surely are *analogues* to such reporting in the animal kingdom. (For example, the warning signals of birds.)

4.

Reporting the intentional objects of (acoustic) sensations must be distinguished from reporting their intrinsic character or qualities.

An individual sound can be strong (loud) or weak (soft), high- or low-pitched, buzzing, humming, piercing, roaring, shrill, whispering, *etc.* These attributions are usually not descriptions of the intentional object of an acoustic sensation, of *what* (we think) we hear. They are rather specifications of *how* we hear what we hear. The qualities are intrinsic to the acoustic sensation. They are (sound-)*universals* under which individual sounds fall.

If an individual sound is "stripped" of all its acoustic qualities what remains of it? Only its spatio-temporal location. The characterization of physical things (phenomena) as a "togetherness of qualities in space and time" seems particularly apt in the acoustic realm.

5.

If an individual sound is a compound of acoustic qualities, sound "itself" might be called a *generic* sense quality. The same is true of light and optic qualities.

There is a branch of physical theory, acoustics, which answers the question what sound is as follows: Sound is a longitudinal travelling pressure wave of molecules of a material medium – for example air or water. Then follows a specification of the kind of movement with regard to wave length, frequency, speed of propagation in various media, *etc.* This is what sound *is*. When we hear sound we hear movement, the "dance of the molecules". Another way of saying the same is that we experience (in German *erfahren*) the movement *as*, or "in the form of", *sound*. This statement, incidentally, cannot be inverted. To say that we experience the sound *as movement* would not be right.

When the frequencies of pressure charges increase or decrease beyond certain limits we, *i.e.*, human ears, can no longer hear them. For purposes of physical theory it may yet be convenient to call these movements, too, "sound" – in which case one would have to say that not all sound is audible! Sound is not just what we can hear. One also speaks of subsonic and supersonic sound. Perhaps such sound can be heard by sentient beings whose auditory sensory apparatus is different from that of humans. (Cf. above pp. 78–79 on O. Granit's experiments with birds.)

Acoustic theory is of relatively recent origin. It was not known, for example, to the Ancient Greeks. So these people did not know what sound was!? Yet they were acquainted with acoustic qualities in much the same way as we are, could identify the presence and absence of sound at a certain time in a certain place, describe what they heard and how they heard it just as accurately as we do this. There might have been many sounds for example of musical instruments which they had never known (heard) but that would not make their basic *concept* "sound" different from ours.

It is a contingent fact that sound is, that sound should have "turned out to be" wave movement of a certain kind. But there is a non-contingent relationship between this fact and *hearing*. In establishing and testing acoustic theory one had to rely on how subjects taking part in experiments responded with acoustic sensations to certain stimulations of their hearing organs – for example that they were able to register "with their ears" the variations in the swingings of tuning forks or of strings of varying length and thickness or variations in the speed of rotating perforated discs, *etc.* If there had not been this correlation between hearing on the one hand and facts as those mentioned about the experimental situations on the other hand, one would not have been investigating *sound, i.e.* the physical phenomenon with which we are acquainted in hearing.

"Waves" suggests an optic image. Waves move back and forth, or up and down, they "undulate" as we say. These are things we have witnessed with our

eyes – primarily with the billows on the waters or with undulating strings. But the sound waves we do not see. They do not manifest themselves to the senses in the picturesque forms which we perhaps associate with them in our thoughts. To speak of them as "waves" is *analogical* talk. Its literal meaning is a web of physical theory based on experimental findings. The waves we actually *see* are not "sound-waves". They are movements of physical objects which in turn "stir up" movements in the air (the medium). These are the sound-waves. But them we do not see.

But do we not sometimes literally hear as well as see waves? For example, when we *see* the sea moving and *hear* the billows hitting the shore? The answer is No. The billows which we see undulating on the waters *produce* (cause) sound, *i.e.* stir up movements in the air which we then hear as sound. But *these* movements we do not see.

6.

The conception of sound as longitudinal waves propagated from the sound-producing source through the medium separating the source from the hearing ear has the following philosophically important consequence: it helps us to understand what I propose to call "the (logical) possibility of hearing". In order to be heard, sound must somehow reach or enter the ear. But since the contact between the source of the sound and the sense organ which is the recipient of the sound is not direct (as, for example, in the case of touch), it is a problem for philosophical reflection, how this *can* happen. The problem is known from Ancient philosophy and is related to ideas about causation and "action at a distance".

In the case of sound the separation is between the *source* of the sound and the *sense organ*, the ear. The sound itself is not separated from the hearing ear. Rather one would wish to say that it fills the space separating the ear from the source of the sound, the sounding thing. The whole of the separating space? Of this we have no immediate experience. But we have some primitive experience of sound having to "travel" from its source to the ear – for example, when seeing lightning a few seconds before hearing the thunder. The wave theory of sound makes us *understand* the time-lag of hearing in relation to seeing.

But sound waves impinging on or entering the ear or striking the eardrum is not the end of the story how sound becomes perceptible. The sound not only has to enter the ear but also, so to speak, to "enter the mind", *i.e.* produce or result in an acoustic sensation, the subject's hearing the sound.

How does that happen? How is it even possible? These are no longer questions of acoustics (physics). Ultimately they are questions of philosophy. But there is another science which can claim a "mediating" rôle between the

study of the physical phenomenon sound and the mental phenomenon hearing. This is neurophysiology.

When entering the ear the sound-waves call forth processes in the cochlear nerve. They are propagated to central parts of the brain (the "hearing centre"). Here they give rise to a sensation – provided that certain conditions of audibility are satisfied. (Cf. above p. 155.) Simplifying, the neural processes involve movements of electrical charges and may be regarded as electrochemical chain reactions along the nerves. One could say that, just as the sound waves carry the sound from its source to the ear, the neural processes carry it along the cochlear nerve to the mind. But this, of course, is a mode of speech full of conceptual, *i.e.* philosophical, pitfalls.

The sound-waves do not "carry" the sound. They *are* the sound, travelling from its source to the ear. What then, if anything, can the neural processes be said to "carry"? Certainly not anything *distinct from themselves* – except perhaps in a metaphorical sense. One can say that they carry information, the "message" of the waves to the hearing centre in the brain.

Would it perhaps be right to say of the neural processes that *they* are the sound? Or that they are the sound *transformed* from waves in the medium separating the ear from the source to movements in the nervous system? Then sound would, as a physical phenomenon, exist in two very different forms – like water and ice! This would mean that the mind "receives" the sound as something different from the waves in the medium. What the *mind* receives, however, is an acoustic sensation. We hear the sound. Shall we therefore say that we, the subjects, hear the neural processes? Can one *hear* neural processes? Well, if it is correct to say that one hears sound-waves, why should it not also be correct to say that one hears those waves when transformed into neural processes?

7.

What is it to *hear*?

One answer is: To hear is to *react* to sound. Another: To hear is to *attend* to (a) sound. Related to attending and reacting is *listening*.

Assume that a sound, for example of a saw, goes on continuously. But I hear it only intermittently. I hear it when I, as we say, "lend an ear to it", *i.e.* attend to it. Attending is then like an action, something I do, like turning a switch on and off. But it is also like forgetting and remembering. I forget about the sound and do not hear it. Then I, as it were, remember it and *listen*: there it is. But all the time the sound was there and could have been heard by me had I turned my attention to it. These are familiar phenomena.

A sound can *alert* me. Is it like this: I hear it and *then* react, attend, listen to it? So that one can hear a sound before reacting or attending to it? I think

this is not the right way to describe the case. I should rather say that being alerted by a sound *is* a form of reacting to it – for example by a jerk or a turn of the head. It is good to consider concrete cases when we say that a sound alerts us.

Reacting is passive, attending active. Reacting somehow seems more basic. In order to attend to a sound I must first have reacted to it, have been alerted by it. But all these locutions are vague and easily slide into metaphor.

How does one react to a sound? One typical reaction can be described as *orienting* oneself towards its source. I turn my head and look in the direction from where the sound seems to emanate. Or I approach the source; someone was perhaps calling me, and I recognize the sound as a call. Or I am curious – what can it be? Another reaction is that I *withdraw* from the assumed source of the sound. Perhaps the sound frightens me. Related to the withdrawing is the reaction of "shutting out" the sound, *e.g.* by holding one's hands tightly pressed against the ears – a very familiar gesture.

Note that all these reactions are *behavioural*. As such they are events in the physical world, the world of the body. This holds true also for the reaction to sound which we call listening, although it normally manifests itself, not in bodily movement, but in the suppression of movement. Or it consists in taking what we call a "listening posture".

Some of the reactions to sound are probably inborn, reflexes. Others – presumably the majority – are acquired through learning. Also when acquired they can be spontaneous and not result from conscious reflection about "what to do".

Hearing thus has, like other mental phenomena, behavioural criteria. And, as with those other phenomena, the subject can, normally, suppress them when he hears, and feign them when he does not hear. In order to suppress and feign them he must, however, be familiar with them. He must "know" what it is to hear, *i.e.* how one reacts to sound. Suppressing them is like saying to oneself: "I hear – but I am not going to show this". This too is a reaction to the sound. (Intentional omission is also a form of behaviour.)

The behavioural reactions to sound are constitutive of hearing. Their presence (occurrence) *means* that the subject hears something.

The relation between the sound and the behavioural reactions to it is *causal*. The physical phenomenon, sound, when it strikes the hearing organ of a sentient being, causes the reactions which *mean* that the subject hears the sound. The relation between the reactions and the acoustic sensation (the subject's hearing the sound) is, however, *semantic*. There is no simple relation: physical cause – mental effect. The Cartesian idea of mind-body causal interaction is a misunderstanding.

With great caution: The mental is the meaning of complex patterns of bodily reactions. Related to Aristotle's view of the soul as the "form" of the body – and of the body as the "stuff" of the soul.

8.

Sound which enters the ear (affects the hearing organ) causes behavioural reactions to it – under appropriate conditions of audibility. But this causing is not direct. Immediately the sound causes processes in the cochlear nerve which are propagated to a hearing centre in the brain. "There hearing takes place", we say naively. But hearing is reacting to the sound and the reactions are behavioural. The behavioural reactions are caused by outgoing nervous impulses from the brain to the muscles. So, evidently, it must be that the ingoing sensorial impulses caused by the sound in their turn cause outgoing motor impulses, and through them the behavioural reactions which mean that the sentient being in question hears it.

Assume that we hear something. If asked what we hear we would, usually, refer to some phenomenon in the external world which caused the sound, – for example say that we heard thunder or a drumbeat or someone crying "help". We should not ordinarily say that we hear sound waves (pressure waves). But, surely, if we hear something we (also) hear sound and sound *is* waves of a certain kind. One is tempted to say that *immediately* when hearing something we hear sound which we then "project" to its source in the external world and describe accordingly.

But this cannot be quite right. The sound-waves cause processes in the nervous system which in turn cause reactions, the meaning of which is that the recipient of the sound hears something. So what we "most immediately" experience when hearing sound are the nervous processes caused by sound (waves). Shall we say we hear *them*? We have already raised the question (p. 160). I think we may answer in the affirmative. And also say that the nervous processes, the behavioural effects of which ultimately constitute hearing, *are* the sound transformed. We are moving over a ground for whose description there is not a settled use of language. In support of my suggestions I shall offer the following argument:

As said at the beginning (p. 155), we sometimes have acoustic sensations which originate from processes in the hearing organ which are not caused by an affectation of the organ by a sound from outside. They are known as tinnitus phenomena. The processes may have some remoter cause in the inner ear, or the brain or perhaps just be "spontaneous". We speak of such cases as acoustic illusions or, perhaps, hallucinations. We say, for example, that we *seemed* to hear the doorbell or telephone ringing but soon made sure that there was no such sound in the surrounding to be heard. Was what we heard then not a

sound? Surely we *heard* it. We may even have reacted to it in exactly the same way as to a "real" sound, – *e.g.* by lifting the telephone receiver. (This sometimes happens with me.) Calling the sound "unreal" or even "nonexistent" is only to create a linguistic mythology. To say that we *heard a sound* is correct use of language here! But the sound we heard was *not* sound-waves in a medium. As physical phenomenon the sound was "reduced" to that which we most immediately experience in every case of hearing, *viz.* certain processes internal to the auditory nervous system. To say that we hear *them* is no more strange than to say that we hear wave movement in the air or some other medium as sound. (Cf. above p. 158.)

When the intentional objects of acoustic sensations refer to sources of sound in the external world, we often say that we hear the sound-emitting or -producing source itself. For example footsteps or church-bells or the rustling leaves on the trees. Sound-producing things are also said to sound, be sounding. But *sound* is not sound-producing. Sound does not make the molecules in the air dance. Their dance *is* sound.

9.

Just as the conception of sound as pressure waves in a medium helps us to understand how it is possible to hear things at a distance, similarly the conception of sound as nervous processes throws light on the *subjectivity* characteristic of (acoustic) sensations.

Two or more persons can hear the same sound. Their acoustic sensations – of the same or of different sounds – can be compared and found to be qualitatively similar, even identical. Such comparisons are based on the subject's (introspective) reports on the intrinsic qualities of the sound they hear. But: my acoustic sensations are mine and yours are yours; you cannot *have* my sensations, nor can I have yours. How shall we understand this "privacy" or "subjectivity"? I think the answer is as follows:

In order for you to have my acoustic sensation you should be able to hear *my* nervous processes. Assume that those processes were, somehow, *sound-producing*. So that sound produced by them could (through the medium separating us) enter your ear and cause in your inner ear and cochlear nerve processes which *you* then hear. What you hear, however, would not be *my* nervous processes but sound which they produce and which was propagated, first to your ear and then from there to the hearing centre in *your* brain. The acoustic sensation you have would be yours, not mine.

Normally, for all I know, the processes in my cochlear nerve which I hear when I have an acoustic sensation, are *not* themselves sound-producing. But it is known that they *can* be this.[1] It may happen that those processes are so "violent" that *they* make the ossicles and tympanic membrane vibrate and

produce sound, *i.e.* waves in the medium which can be heard by an outsider. But this is *another* sound, different from the sound which I heard directly, without the "mediation" of processes in my ear. So even when there is such a "secondary" sound-production by the nerve, your acoustic sensation is not the one which I have.

Philosophers have sometimes entertained the phantasy that several subjects might have their nervous systems (or part of their nervous systems) *in common*. If this imagination answers to a logically possible reality, would it then be right to say that they have the same sensations? I do not know whether the phantasy answers to a real possibility – nor whether, if it does, it would be right to say that such Siamese brain-twins have shared sensations.

*

Much of what has been said in this paper about sound and hearing applies also *mutatis mutandis* to other sense modalities, for example to light and seeing. But there are also important differences between the modalities which must be noted. Each of them therefore deserves a separate treatment.

NOTE

[1] The remarkable fact that the cochlea is capable of generating audible sound was first established by P. Zurek in 1981. The phenomenon is also known as "spontaneous oto-acoustic emission". I am indebted to Professor Reuter for this item of information.

CONCLUDING POSTSCRIPT.
ON PAIN AND SOUND.

1.

Let there be an affectation *A* of a subject's body from outside. For example: a sound striking the ear or a stone hitting his leg. The "hit", we assume, causes (effects, puts in motion) some afferent neural processes *N* thus causing the subject, as we say, to have a sensation *S*. For example: to hear sound or to feel (suffer) pain. *N* may continue in the form of efferent neural processes causing bodily movements or behaviour *B* characteristic of having the sensation *S*. For example: reacting to the sound by taking a listening posture or reacting to the pain by a cry or a twist of the face.

The afferent and efferent (sensorial and motoric) part of *N* we have earlier distinguished as N_1 and N_2, and the neural mechanism responsible for the "switch" from sensorial to motoric we have denoted N_3. In the temporal dimension *N* may thus equal $N_1+N_3+N_2$.

A, *N*, and *B* are physical or material events and states. *S* we call mental or psychic.

2.

There is a far-reaching analogy between hearing sound and feeling pain. But there are also noteworthy conceptual differences.

I have tried to argue that in hearing, having an acoustic sensation, we immediately experience (sense) nervous processes. Similarly, I would like to say that in feeling pain we immediately experience (sense) something that goes on in our nervous system. In both cases the sensing (having the sensation) is strictly simultaneous with those processes. We hear them and feel them. The nervous and the sensational (physical and mental) coincide without being identical. It is inviting to speak of them as the "subjective" and the "objective" side (aspect) *of the same*. "The same" what? Shall we say "reality"? It is difficult to find the right word here. An observer can study the nervous processes, "measure" them in various ways. But he cannot feel or hear them. Only *I* can.

It is important – very important – *not* to think of the relation between the nervous processes and the sensations as a succession in time: first *N*, then *S*.

G. H. von Wright, In the Shadow of Descartes. Essays in the Philosophy of Mind, 165–170.

This would mean thinking that *N* causes *S*. But this, simply, is not the case. What *N* may cause are behavioural reactions *B* which to an observer *mean* that the subject has *S*. But since *S* and (at least part of) *N* are simultaneous, it is inviting and, in most cases, also innocuous to say that *S* causes *B* too – for example, that the feeling of pain (a pain sensation) is the cause of pain behaviour.

The cause of *N* normally is some bodily affectation (affliction) *A*, – in the case of pain, for example, a hit on the leg or the cut of a knife. Since *A* causes (at least part of) *N*, it is natural and, as such, also innocuous to think that *A* causes *S*, too. So what we get seems a perfect example of "Cartesian interactionism": something physical, *A*, causing something mental, *S*, – and something mental, *S*, causing something physical, *B*. This is quite in order as long as one can avoid muddling things by philosophizing about them.

Is my position to be labelled, philosophically speaking, "dualism" or "identity-theory"? One can accept both labels – and at the same time reject both.

3.

N caused by *A* may "outlast" *S*. The nervous processes which the bodily affectation causes (inaugurates) may not immediately be sensed (felt or heard), but only when they reach some central parts of the brain. Perhaps it is possible to localize, in space and/or in time, the "stretch" or "segment" of *N* which is actually sensed. (I spoke earlier of the "tip" of *N* in the brain as that segment of *N* which is sensed by the subject.) Does neurophysiology know the answer? Is my question even neurophysiologically intelligible?

The part-whole relation between *N* and *S* which I have in mind here makes it natural to think of the initial part of the neural processes as cause of the (subsequent) mental phenomenon *S*. This would give to the neural processes a kind of temporal *and* causal priority in relation to the sensational, the physical in relation to the mental. This agrees with the way we talk about these things and is, properly understood, "philosophically innocuous" and in order.

There is also another sense in which *N* may outlast *S*. Namely, by having a continuation in the motor reactions (efferent impulses) which cause the behavioural reactions *B* characteristic of having *S* – say, hearing a sound or feeling pain. These are the "parts" of *N* which previously I called N_2 and N_3, the second being responsible for what I called the "switch" from sensorial to motoric. These relationships again make it plausible to think of the sensation *S* as causally efficacious in relation to the temporally later parts of *N*. If taken at its face value, this, too, is in order.

4.

I have said that when hearing sound or feeling pain we *immediately* hear or feel what is going on in our brain (nervous processes). Saying this repels us, somehow.

When speaking about sound I also said that processes in the cochlear nerve are the sound "transformed" from sound-waves to those processes. A reason for saying this was that we do not hear the sound-waves directly, but only thanks to the "mediating rôle" of the nervous processes. Sometimes there are no sound-waves at all "behind" the acoustic sensation. (Tinnitus phenomena.) Yet we hear a sound. So, must we not say that then *what* we hear are nervous processes? Moreover, we hear them "directly" or "immediately" – or else the nerves would have to be "sound-producing", which normally they are not. (See above p. 163f.) Since we cannot hear sound without nervous processes occurring, is it then not true that, whenever we have an acoustic sensation, we *immediately* hear, without any further physical mediation, nervous processes? It seems to me that we must say this – and that it is fully reconcilable with the fact that when we report what we hear we normally, as a result of learning, *project* what immediately we hear to its source in the external world or, when speaking of pain, to its source in the human or animal body.

5.

We must next consider a difference between hearing sound and feeling pain.

Sound is a physical phenomenon. And so of course are the nervous processes which sound may "release" in the ear (cochlear nerve). But is pain too a physical phenomenon? Pain exists only in a living body. We often call pain a "bodily sensation". What is physical about it are the processes N caused by some affliction A which we feel as ("in the form of") pain. Could we then not say that pain as physical phenomenon *is* the nervous processes which (or part of which) we experience as pain?

This would be another *contingent identity*. For establishing (coming to know) it, the sensational member of the identity is crucial. Hearing sound and feeling pain provide the criteria with the aid of which we identify the *physical* phenomenon of sound as waves or pain as intra-bodily nervous processes.

6.

Pain behaviour is the bodily *expression* of pain. This expressive relation is *semantic*. If there did not exist behaviour which is (logically, conceptually) characteristic of pain, the statement that a subject suffers pain would not make sense. But a subject can have pain without exhibiting pain behaviour – and he can exhibit pain behaviour without suffering pain. Pain behaviour *means* that

the subject has pain. But to say that the subject has pain does not mean that he behaves in a way characteristic of pain.

Pain behaviour can be objectively observed. An external observer can be aware of it and on that ground attribute pain to the subject. His behaviour means that he is in pain. Can the subject himself be aware of his pain behaviour? Certainly. But this would not be the ground why he attributes pain to himself: "I scream and jump on one leg – so evidently I have pain in the other leg", is nonsense.

To the subject itself the behavioural signs of pain do not mean that he has pain. That is: he does not "read off" from the signs that he suffers pain. For him the relation between the signs and his pain is, rather, the reverse of what it is for the observer. If asked why he is groaning, tending to his foot, *etc.* he could answer that he does this *because* he has pain in his foot. Is the pain then the cause of his behaviour? Is it not rather the *reason* for it? The *cause* is in the nervous system. If I cried out in pain and being asked why I cried answered "because I am in pain", my interlocutor would probably understand that I wanted to be helped, relieved of the pain. Perhaps my cry actually was "help!". Perhaps I had deliberated before crying: "It hurts terribly – better ask for help", and then cried out. The *reason* why I cried was that I wanted assistance. This I wanted *because* of the pain. Is the pain now cause or reason? Shall we say that the pain caused me to want help, or that it was a reason for this? And what is the difference between saying the one thing or the other?

If I am in pain and scream to call for help, then my crying is not part of the pain behaviour which means that I am in pain. (Or better: it can be this too but in so far as it is intended as a call for help it is *not* pain behaviour. Pain behaviour is my *spontaneous* reactions to the pain. Some such reactions, I am sure, are "inborn". Others are learnt – for example, verbal reactions. Pain behaviour is *expressions* of pain. "It hurts" *may*, but *need not*, be a statement.

Pain thus has a twofold rôle in these contexts. "Did you cry out because you could not help it – or did you cry out because you wanted to call attention?" The question may not have a univocal answer. This is a typical, and interesting, feature of the case.

7.

There is a (short) time-lag between N and B. B follows a little later. The temporal relation between S and B is more complex. If S occupies, in time, only a part of (the duration of) N, it is conceivable that B, which has its cause in N, actually happens *before* S. Pain behaviour *may* occur before pain is felt or a head may turn before the sound, to which the turn is a reaction, is heard. An example would be touching a hot plate and removing one's hand instantly,

before feeling pain. Removing the hand is then a reflex. But it is also pain behaviour, a spontaneous reaction *B* to a bodily affliction *A*.

Generally speaking: the bodily reactions which are *criteria* of the subject's having a sensation – *e.g.* of pain – must be spontaneous ("primitive", "automatic", reflex-like). Therefore, their cause cannot be the sensation "itself". Their cause is the nervous processes of which, however, *S* may be a temporal part or segment. But *S* may play an independent rôle as a reason for some behaviour which is not criterional in relation to *S*, but undertaken for the sake of some end (*e.g.* being helped).

The upshot of this is that the cause of the criterional *B* is *N*, and that *S*, the sensation or "consciousness", is not causally efficacious. This was said before. (Cf. above p. 154.)

8.

Is having pain necessarily something the subject experiences, feels? Could he exhibit pain behaviour without experiencing (feeling, being conscious of) pain? A person is in a state of coma after an accident. We see him perform contorted movements, hear him groan silently, *etc.* Afterwards he cannot recollect having felt anything. Here it would be right to say that he had suffered pain, but *not* right to say that he felt, was conscious of, pain. So, one can have pain without feeling it. But this, surely, is a marginal case.

A different case is when we say of a person that he has pain but that he intermittently does not notice or pay attention to it. Or that he "forgets" about his pain. We are all familiar with such cases. Yet their description is (philosophically) problematic. Can a person, who does not exhibit pain behaviour really "forget" about the pain he *has*? Should we not rather say that he *has* the pain only intermittently? When he does not feel it, the pain simply is not there, does not exist.

A person can thus, in marginal cases, be said to have pain without feeling it, but only provided he exhibits pain behaviour. But a person cannot correctly be said to have pain, if he does not feel it and does not exhibit pain behaviour.

Shall we say of the person who exhibits pain behaviour but does not feel pain that he suffers *sub-conscious* pain? Some people would, I think, wish to use this locution. But it is easily misleading and therefore it is better to avoid talking about "sub-conscious pain". (The sub-conscious is a mystification.)

9.

To say that a person has pain but does not exhibit any pain behaviour is to attribute to him a state in which a person is said to be when he exhibits behaviour (including verbal reactions) of a characteristic kind, *i.e.* of that kind

which *would* make us say that he is in pain. "Pure mental states" without any physical manifestations are a kind of counterfactual construction.

It is interesting that existential statements about the "purely" mental and also about the "purely" physical have to be construed counterfactually. (Cf. above pp. 120–121.) This points to the artificiality of the *sharp* ("Cartesian") mind-body or mental-material distinction. It supports the idea that the two are *conceptually intertwined*. The "impenetrable subjectivity" of the mental (psychic) and the "totally mind-independent objectivity" of the physical (material) are, so to speak, marginal cases of the real.

INDEX OF PERSONS

SUBJECT INDEX

163. D. Mayr and G. Süssmann (eds.), *Space, Time, and Mechanics.* Basic Structures of a Physical Theory. 1983 ISBN 90-277-1525-4
164. D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic.* Vol. I: Elements of Classical Logic. 1983 ISBN 90-277-1542-4
165. D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic.* Vol. II: Extensions of Classical Logic. 1984 ISBN 90-277-1604-8
166. D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic.* Vol. III: Alternative to Classical Logic. 1986 ISBN 90-277-1605-6
167. D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic.* Vol. IV: Topics in the Philosophy of Language. 1989 ISBN 90-277-1606-4
168. A. J. I. Jones, *Communication and Meaning.* An Essay in Applied Modal Logic. 1983 ISBN 90-277-1543-2
169. M. Fitting, *Proof Methods for Modal and Intuitionistic Logics.* 1983 ISBN 90-277-1573-4
170. J. Margolis, *Culture and Cultural Entities.* Toward a New Unity of Science. 1984 ISBN 90-277-1574-2
171. R. Tuomela, *A Theory of Social Action.* 1984 ISBN 90-277-1703-6
172. J. J. E. Gracia, E. Rabossi, E. Villanueva and M. Dascal (eds.), *Philosophical Analysis in Latin America.* 1984 ISBN 90-277-1749-4
173. P. Ziff, *Epistemic Analysis.* A Coherence Theory of Knowledge. 1984 ISBN 90-277-1751-7
174. P. Ziff, *Antiaesthetics.* An Appreciation of the Cow with the Subtile Nose. 1984 ISBN 90-277-1773-7
175. W. Balzer, D. A. Pearce, and H.-J. Schmidt (eds.), *Reduction in Science.* Structure, Examples, Philosophical Problems. 1984 ISBN 90-277-1811-3
176. A. Peczenik, L. Lindahl and B. van Roermund (eds.), *Theory of Legal Science.* Proceedings of the Conference on Legal Theory and Philosophy of Science (Lund, Sweden, December 1983). 1984 ISBN 90-277-1834-2
177. I. Niiniluoto, *Is Science Progressive?* 1984 ISBN 90-277-1835-0
178. B. K. Matilal and J. L. Shaw (eds.), *Analytical Philosophy in Comparative Perspective.* Exploratory Essays in Current Theories and Classical Indian Theories of Meaning and Reference. 1985 ISBN 90-277-1870-9
179. P. Kroes, *Time: Its Structure and Role in Physical Theories.* 1985 ISBN 90-277-1894-6
180. J. H. Fetzer, *Sociobiology and Epistemology.* 1985 ISBN 90-277-2005-3; Pb 90-277-2006-1
181. L. Haaparanta and J. Hintikka (eds.), *Frege Synthesized.* Essays on the Philosophical and Foundational Work of Gottlob Frege. 1986 ISBN 90-277-2126-2
182. M. Detlefsen, *Hilbert's Program.* An Essay on Mathematical Instrumentalism. 1986 ISBN 90-277-2151-3
183. J. L. Golden and J. J. Pilotta (eds.), *Practical Reasoning in Human Affairs.* Studies in Honor of Chaim Perelman. 1986 ISBN 90-277-2255-2
184. H. Zandvoort, *Models of Scientific Development and the Case of Nuclear Magnetic Resonance.* 1986 ISBN 90-277-2351-6
185. I. Niiniluoto, *Truthlikeness.* 1987 ISBN 90-277-2354-0
186. W. Balzer, C. U. Moulines and J. D. Sneed, *An Architectonic for Science.* The Structuralist Program. 1987 ISBN 90-277-2403-2
187. D. Pearce, *Roads to Commensurability.* 1987 ISBN 90-277-2414-8
188. L. M. Vaina (ed.), *Matters of Intelligence.* Conceptual Structures in Cognitive Neuroscience. 1987 ISBN 90-277-2460-1
189. H. Siegel, *Relativism Refuted.* A Critique of Contemporary Epistemological Relativism. 1987 ISBN 90-277-2469-5

190. W. Callebaut and R. Pinxten, *Evolutionary Epistemology.* A Multiparadigm Program, with a Complete Evolutionary Epistemology Bibliograph. 1987 ISBN 90-277-2582-9
191. J. Kmita, *Problems in Historical Epistemology.* 1988 ISBN 90-277-2199-8
192. J. H. Fetzer (ed.), *Probability and Causality.* Essays in Honor of Wesley C. Salmon, with an Annotated Bibliography. 1988 ISBN 90-277-2607-8; Pb 1-5560-8052-2
193. A. Donovan, L. Laudan and R. Laudan (eds.), *Scrutinizing Science.* Empirical Studies of Scientific Change. 1988 ISBN 90-277-2608-6
194. H.R. Otto and J.A. Tuedio (eds.), *Perspectives on Mind.* 1988 ISBN 90-277-2640-X
195. D. Batens and J.P. van Bendegem (eds.), *Theory and Experiment.* Recent Insights and New Perspectives on Their Relation. 1988 ISBN 90-277-2645-0
196. J. Österberg, *Self and Others.* A Study of Ethical Egoism. 1988 ISBN 90-277-2648-5
197. D.H. Helman (ed.), *Analogical Reasoning.* Perspectives of Artificial Intelligence, Cognitive Science, and Philosophy. 1988 ISBN 90-277-2711-2
198. J. Wolenski, *Logic and Philosophy in the Lvov-Warsaw School.* 1989 ISBN 90-277-2749-X
199. R. Wójcicki, *Theory of Logical Calculi.* Basic Theory of Consequence Operations. 1988 ISBN 90-277-2785-6
200. J. Hintikka and M.B. Hintikka, *The Logic of Epistemology and the Epistemology of Logic.* Selected Essays. 1989 ISBN 0-7923-0040-8; Pb 0-7923-0041-6
201. E. Agazzi (ed.), *Probability in the Sciences.* 1988 ISBN 90-277-2808-9
202. M. Meyer (ed.), *From Metaphysics to Rhetoric.* 1989 ISBN 90-277-2814-3
203. R.L. Tieszen, *Mathematical Intuition.* Phenomenology and Mathematical Knowledge. 1989 ISBN 0-7923-0131-5
204. A. Melnick, *Space, Time, and Thought in Kant.* 1989 ISBN 0-7923-0135-8
205. D.W. Smith, *The Circle of Acquaintance.* Perception, Consciousness, and Empathy. 1989 ISBN 0-7923-0252-4
206. M.H. Salmon (ed.), *The Philosophy of Logical Mechanism.* Essays in Honor of Arthur W. Burks. With his Responses, and with a Bibliography of Burk's Work. 1990 ISBN 0-7923-0325-3
207. M. Kusch, *Language as Calculus vs. Language as Universal Medium.* A Study in Husserl, Heidegger, and Gadamer. 1989 ISBN 0-7923-0333-4
208. T.C. Meyering, *Historical Roots of Cognitive Science.* The Rise of a Cognitive Theory of Perception from Antiquity to the Nineteenth Century. 1989 ISBN 0-7923-0349-0
209. P. Kosso, *Observability and Observation in Physical Science.* 1989 ISBN 0-7923-0389-X
210. J. Kmita, *Essays on the Theory of Scientific Cognition.* 1990 ISBN 0-7923-0441-1
211. W. Sieg (ed.), *Acting and Reflecting.* The Interdisciplinary Turn in Philosophy. 1990 ISBN 0-7923-0512-4
212. J. Karpiński, *Causality in Sociological Research.* 1990 ISBN 0-7923-0546-9
213. H.A. Lewis (ed.), *Peter Geach: Philosophical Encounters.* 1991 ISBN 0-7923-0823-9
214. M. Ter Hark, *Beyond the Inner and the Outer.* Wittgenstein's Philosophy of Psychology. 1990 ISBN 0-7923-0850-6
215. M. Gosselin, *Nominalism and Contemporary Nominalism.* Ontological and Epistemological Implications of the Work of W.V.O. Quine and of N. Goodman. 1990 ISBN 0-7923-0904-9
216. J.H. Fetzer, D. Shatz and G. Schlesinger (eds.), *Definitions and Definability.* Philosophical Perspectives. 1991 ISBN 0-7923-1046-2
217. E. Agazzi and A. Cordero (eds.), *Philosophy and the Origin and Evolution of the Universe.* 1991 ISBN 0-7923-1322-4
218. M. Kusch, *Foucault's Strata and Fields.* An Investigation into Archaeological and Genealogical Science Studies. 1991 ISBN 0-7923-1462-X

219. C.J. Posy, *Kant's Philosophy of Mathematics.* Modern Essays. 1992 ISBN 0-7923-1495-6
220. G. Van de Vijver, *New Perspectives on Cybernetics.* Self-Organization, Autonomy and Connectionism. 1992 ISBN 0-7923-1519-7
221. J.C. Nyíri, *Tradition and Individuality.* Essays. 1992 ISBN 0-7923-1566-9
222. R. Howell, *Kant's Transcendental Deduction.* An Analysis of Main Themes in His Critical Philosophy. 1992 ISBN 0-7923-1571-5
223. A. García de la Sienra, *The Logical Foundations of the Marxian Theory of Value.* 1992 ISBN 0-7923-1778-5
224. D.S. Shwayder, *Statement and Referent.* An Inquiry into the Foundations of Our Conceptual Order. 1992 ISBN 0-7923-1803-X
225. M. Rosen, *Problems of the Hegelian Dialectic.* Dialectic Reconstructed as a Logic of Human Reality. 1993 ISBN 0-7923-2047-6
226. P. Suppes, *Models and Methods in the Philosophy of Science: Selected Essays.* 1993 ISBN 0-7923-2211-8
227. R. M. Dancy (ed.), *Kant and Critique: New Essays in Honor of W. H. Werkmeister.* 1993 ISBN 0-7923-2244-4
228. J. Woleński (ed.), *Philosophical Logic in Poland.* 1993 ISBN 0-7923-2293-2
229. M. De Rijke (ed.), *Diamonds and Defaults.* Studies in Pure and Applied Intensional Logic. 1993 ISBN 0-7923-2342-4
230. B.K. Matilal and A. Chakrabarti (eds.), *Knowing from Words.* Western and Indian Philosophical Analysis of Understanding and Testimony. 1994 ISBN 0-7923-2345-9
231. S.A. Kleiner, *The Logic of Discovery.* A Theory of the Rationality of Scientific Research. 1993 ISBN 0-7923-2371-8
232. R. Festa, *Optimum Inductive Methods.* A Study in Inductive Probability, Bayesian Statistics, and Verisimilitude. 1993 ISBN 0-7923-2460-9
233. P. Humphreys (ed.), *Patrick Suppes: Scientific Philosopher.* Vol. 1: Probability and Probabilistic Causality. 1994 ISBN 0-7923-2552-4
234. P. Humphreys (ed.), *Patrick Suppes: Scientific Philosopher.* Vol. 2: Philosophy of Physics, Theory Structure, and Measurement Theory. 1994 ISBN 0-7923-2553-2
235. P. Humphreys (ed.), *Patrick Suppes: Scientific Philosopher.* Vol. 3: Language, Logic, and Psychology. 1994 ISBN 0-7923-2862-0
Set ISBN (Vols 233–235) 0-7923-2554-0
236. D. Prawitz and D. Westerståhl (eds.), *Logic and Philosophy of Science in Uppsala.* Papers from the 9th International Congress of Logic, Methodology, and Philosophy of Science. 1994 ISBN 0-7923-2702-0
237. L. Haaparanta (ed.), *Mind, Meaning and Mathematics.* Essays on the Philosophical Views of Husserl and Frege. 1994 ISBN 0-7923-2703-9
238. J. Hintikka (ed.), *Aspects of Metaphor.* 1994 ISBN 0-7923-2786-1
239. B. McGuinness and G. Oliveri (eds.), *The Philosophy of Michael Dummett.* With Replies from Michael Dummett. 1994 ISBN 0-7923-2804-3
240. D. Jamieson (ed.), *Language, Mind, and Art.* Essays in Appreciation and Analysis, In Honor of Paul Ziff. 1994 ISBN 0-7923-2810-8
241. G. Preyer, F. Siebelt and A. Ulfig (eds.), *Language, Mind and Epistemology.* On Donald Davidson's Philosophy. 1994 ISBN 0-7923-2811-6
242. P. Ehrlich (ed.), *Real Numbers, Generalizations of the Reals, and Theories of Continua.* 1994 ISBN 0-7923-2689-X
243. G. Debrock and M. Hulswit (eds.), *Living Doubt.* Essays concerning the epistemology of Charles Sanders Peirce. 1994 ISBN 0-7923-2898-1
244. J. Srzednicki, *To Know or Not to Know.* Beyond Realism and Anti-Realism. 1994 ISBN 0-7923-2909-0
245. R. Egidi (ed.), *Wittgenstein: Mind and Language.* 1995 ISBN 0-7923-3171-0

246. A. Hyslop, *Other Minds*. 1995 ISBN 0-7923-3245-8
247. L. Pólos and M. Masuch (eds.), *Applied Logic: How, What and Why*. Logical Approaches to Natural Language. 1995 ISBN 0-7923-3432-9
248. M. Krynicki, M. Mostowski and L.M. Szczerba (eds.), *Quantifiers: Logics, Models and Computation*. Volume One: Surveys. 1995 ISBN 0-7923-3448-5
249. M. Krynicki, M. Mostowski and L.M. Szczerba (eds.), *Quantifiers: Logics, Models and Computation*. Volume Two: Contributions. 1995 ISBN 0-7923-3449-3
Set ISBN (Vols 248 + 249) 0-7923-3450-7
250. R.A. Watson, *Representational Ideas from Plato to Patricia Churchland*. 1995 ISBN 0-7923-3453-1
251. J. Hintikka (ed.), *From Dedekind to Gödel*. Essays on the Development of the Foundations of Mathematics. 1995 ISBN 0-7923-3484-1
252. A. Wiśniewski, *The Posing of Questions*. Logical Foundations of Erotetic Inferences. 1995 ISBN 0-7923-3637-2
253. J. Peregrin, *Doing Worlds with Words*. Formal Semantics without Formal Metaphysics. 1995 ISBN 0-7923-3742-5
254. I.A. Kieseppä, *Truthlikeness for Multidimensional, Quantitative Cognitive Problems*. 1996 ISBN 0-7923-4005-1
255. P. Hugly and C. Sayward: *Intensionality and Truth*. An Essay on the Philosophy of A.N. Prior. 1996 ISBN 0-7923-4119-8
256. L. Hankinson Nelson and J. Nelson (eds.): *Feminism, Science, and the Philosophy of Science*. 1997 ISBN 0-7923-4162-7
257. P.I. Bystrov and V.N. Sadovsky (eds.): *Philosophical Logic and Logical Philosophy*. Essays in Honour of Vladimir A. Smirnov. 1996 ISBN 0-7923-4270-4
258. Å.E. Andersson and N-E. Sahlin (eds.): *The Complexity of Creativity*. 1996 ISBN 0-7923-4346-8
259. M.L. Dalla Chiara, K. Doets, D. Mundici and J. van Benthem (eds.): *Logic and Scientific Methods*. Volume One of the Tenth International Congress of Logic, Methodology and Philosophy of Science, Florence, August 1995. 1997 ISBN 0-7923-4383-2
260. M.L. Dalla Chiara, K. Doets, D. Mundici and J. van Benthem (eds.): *Structures and Norms in Science*. Volume Two of the Tenth International Congress of Logic, Methodology and Philosophy of Science, Florence, August 1995. 1997 ISBN 0-7923-4384-0
Set ISBN (Vols 259 + 260) 0-7923-4385-9
261. A. Chakrabarti: *Denying Existence*. The Logic, Epistemology and Pragmatics of Negative Existentials and Fictional Discourse. 1997 ISBN 0-7923-4388-3
262. A. Biletzki: *Talking Wolves*. Thomas Hobbes on the Language of Politics and the Politics of Language. 1997 ISBN 0-7923-4425-1
263. D. Nute (ed.): *Defeasible Deontic Logic*. 1997 ISBN 0-7923-4630-0
264. U. Meixner: *Axiomatic Formal Ontology*. 1997 ISBN 0-7923-4747-X
265. I. Brinck: *The Indexical 'I'*. The First Person in Thought and Language. 1997 ISBN 0-7923-4741-2
266. G. Hölmström-Hintikka and R. Tuomela (eds.): *Contemporary Action Theory*. Volume 1: Individual Action. 1997 ISBN 0-7923-4753-6; Set: 0-7923-4754-4
267. G. Hölmström-Hintikka and R. Tuomela (eds.): *Contemporary Action Theory*. Volume 2: Social Action. 1997 ISBN 0-7923-4752-8; Set: 0-7923-4754-4
268. B.-C. Park: *Phenomenological Aspects of Wittgenstein's Philosophy*. 1998 ISBN 0-7923-4813-3
269. J. Paśniczek: *The Logic of Intentional Objects*. A Meinongian Version of Classical Logic. 1998 ISBN 0-7923-4880-X
270. P.W. Humphreys and J.H. Fetzer (eds.): *The New Theory of Reference*. Kripke, Marcus, and Its Origins. 1998 ISBN 0-7923-4898-2