

Of Human Freedom

GEORG HENRIK VON WRIGHT

THE TANNER LECTURES ON HUMAN VALUES

Delivered at
The University of Helsinki

May 16 and 17, 1984

GEORG HENRIK VON WRIGHT was born in Helsinki (Helsingfors), Finland, in 1916. He has been professor of philosophy in the universities of Helsinki and of Cambridge, and was for twelve years Andrew D. White Professor-at-Large at Cornell University. Since 1961 he has held a permanent research appointment in the Academy of Finland. Professor von Wright is Corresponding Fellow of the British Academy, Honorary Foreign Member of the American Academy of Arts and Sciences, and Honorary Fellow of the Finnish Society of Sciences and of Trinity College, Cambridge. His published works include *A Treatise on Induction and Probability* (1951), *Norm and Action* (1963), *The Varieties of Goodness* (1963), *Explanation and Understanding* (1971), and *Wittgenstein* (1982).

“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

I

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

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*.

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 with Moore’s argument. At most Moore succeeded in cutting the discussion short for some 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 which is the topic of these lectures 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 volition or acts of the will. Human freedom, 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 no control or power. This was a way of reconciling freedom with determinism (cf. below, 152). 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 obsolescence.

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 more 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 in science. It is as follows:

Every human action has 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 responsible 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.

² In his paper “Philosophy of Action: Davidson, von Wright, and the Debate over Causation,” in *Contemporary Philosophy: A New Survey*, G. Fløistad, ed., vol. 3 (The Hague: Martinus Nijhoff, 1982), pp. 45-72.

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 only I 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 generically? 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 successful 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 tried or 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 (omit) 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” (below, p. 118). 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? Someone 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 acknowledged as duty conforms to the customs and traditions or is prescribed by the legal order of a society (cf. below, p. 118f).

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 *potential* freedom. To be free (able) to do or omit an individual action can, by con-

trast, be called *actual* freedom. It follows from what has already been said that actual nonfreedom 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 (p. 114) 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 some 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” (cf. above, p. 114). 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 acquired 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 (cf. above, p. 114).

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 perhaps not 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 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 the 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 in these lectures.

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 can 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* that we usually mean and understand by the assertion that we have Free Will."⁵ Let us therefore lay aside the problem of choice 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 action "contingent"?

No one would say that an action which I perform is logically necessary. So every action is, *ispo 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.

³ G. E. Moore, *Ethics* (London: Oxford University Press, 1912), P.131.

⁴ *Ibid.*, p. 134

⁵ *Ibid.*, p. 135

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 necessarily “perform,” provided that he has the required ability and that the occasion provides an opportunity for exercising it. Given these prerequisites, 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 in the performance (cf. above, p. 113). 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 action 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 as 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. 139).

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 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 maybe 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, p. 113). 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 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 the other. But it may also be that a factual basis for a well-grounded 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 (above, p. 123), 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 actions of the agent under consideration himself. Things have been done to him; he has for example been given a certain education and training or, on the contrary, been excluded from education and 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 independently 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 *grund*. 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, not supposed 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 challenge, 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 present 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 the 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 reasons and therefore 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 opinions 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 on the distinction 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 passion 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 (cf. above, p. 130). “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 and 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 try to 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 the 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 judged 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 (cf. above, p. 124).

When, in deliberation or in retrospect, reasons are rated for strength they are often called good or bad, better and 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 abstentions.

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 of 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 absence 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 belong 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 declining 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 no 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, and correspondingly for an omission.

In its general form the formulation above 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 with 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 reasons 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 (cf. below, p. 147).

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, we have said (p. 135).

So our problem is: how do we identify the 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 misunderstand-

ing, 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 the two 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 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 is solved [cf. above, p. 136]). 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 some other reason 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. 132). 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 shall 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 also 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 own future actions, ask him to watch himself better.

Obviously, the border between rational argumentation and "brainwashing" 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 the 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 consciousness has changed, and that he has acquired a new understanding of his own 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 in return 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 temptation 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* (cf. above, p. 132). “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 (ac-

tion)?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 (cf. above, p. 138). 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

⁶For how my opinions on this question have changed, cf. my book *The Varieties of Goodness* (London: Routledge & Kegan Paul, 1963), p. 190, and the papers "Determinism and the Study of Man" (1976) and "Explanation and Understanding of Action" (1981) reprinted in *Philosophical Papers I, Practical Reason* (Oxford: Basil Blackwell, 1983).

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 (cf. above, p. 139). 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 — not least in the ominous year 1984.

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 being 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."

However distasteful these analogies in terms of computer language may sound, I think they are useful when the problem of free action is discussed in the context of our present cultural situation. In the next lecture I shall make further use of the metaphor.

II

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.

Also, actions which take place for reasons 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 elsewhere the *results* of the action.⁷ Further changes effected by the results of actions I 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) prerequisites of (the results of) an action. These overt prerequisites have in their turn a covert background in the tension and relaxation of muscles “inside” the body. 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 prerequisites 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 nevertheless has 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.

⁷ *Norm and Action* (London: Routledge & Kegan Paul, 1963), p. 39f.

⁸ *Ibid.*

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 reasons, 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 fists, 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 conformity with common usage, I call human action.

2. There was a time when we 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 acquired 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 (cf. above, p. 111).

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 (cf. above, p. 150). 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. 120). Since, at least in many cases, we do not know what our choices (of course of action) are going to be, we say it is possible that we are going to do a certain thing but also possible that we are going to omit the action.⁹ 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.”

⁹ Cf. Moore, *Ethics*, p.136.

There is a “classic” defense 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 motives and reasons for action are (comparable to physical) causes.¹⁰ 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.” 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 the 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 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 defense of compatibilism leaves another problem unsolved, however. One could call it a problem of *congruence* or

¹⁰ 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, on 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 Edward Westermarck in *The Origin and Development of the Moral Ideas*, I-II (London: Macmillan, 1906–1908).

¹¹ Cf. Westermarck, *Origin and Development*, vol. I, ch. XIII, for a good clarification of the distinction.

parallelism (cf. above, p. 112). Granting that reasons are causes, we seem to have two parallel but independent causal chains here. On 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 hands 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 things 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 innervations moving my hands really *the same*, only described in different ways? One description is in obscure “mentalist” terms (“embarking on the action,” “my want becoming active”), the other in, seemingly, clearer physical (neural) terms. I shall return to this question (below, p. 157ff).

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 guests for dinner. This I obviously did for some reason. The normal reason would be that I had 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 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” 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 speaking 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 “mentalist” 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 innervations and 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 programming 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 answer 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 innervations 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 performance 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 something 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 innervations 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 his 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 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.¹²

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 type 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 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

¹² 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.

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 (cf. above, p. 158).

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.¹³

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

¹³ Cf. Wittgenstein, *Zettel*, §608 (Oxford: Basil Blackwell, 1967).

neurological causes could produce somatic effects (movements of the 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 to 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 innervations and innervations 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 pre-determined, 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 (cf. below, p. 166). 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 “fortuitous” or “gratuitous” actions — except possibly by looking for causes of the movements which constitute their physical aspect (cf. above, p. 147). 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 (above, p. 124f), 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 (cf. above, p. 157). 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 explanation 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 “primi-

tive” 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 innervations caused by 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 a book from the shelf.” The snatching is then an action with a physical aspect of its own, e.g., closing my fingers round 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 he did something for a certain reason, and we say that the physical aspect of his action would have occurred 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 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 then say 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 innervations 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 sometimes 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 of his own making (cf. above, p. 127).

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 from 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 successfully 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 classic rôle. In other areas it may have to be modified (“relaxed”) or it will be dropped as useless (cf. above, p. 150).

11. Have I wanted to say that the 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 on 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. 140).

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 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 more complex, more developed. It is not by being exempted from the bondage 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.