Grammars of the Will

Lars Hertzberg

On the Use of Words: A Prologue on Method

1. In my discussion, I shall make frequent appeals to actual uses of linguistic expressions in specific contexts. A concern with the use of words, which has come to be associated with the name of Ludwig Wittgenstein, is clearly not in vogue among English-speaking philosophers today. Though the matter is rarely argued, there seems to be a consensus that attention to the actual use of words is a superseded stage in philosophy. This may be, in part at least, because attention to use has come to be linked to the idea that something called 'ordinary language' is normative for whatever is to be said in philosophical investigation, i.e. that certain contexts of use are inherently more important than others; that we must look at everyday language in order to determine the real meanings of words. This idea has come to be regarded with suspicion, since there are hardly any compelling grounds for accepting such a doctrine.

To my understanding, however, the importance of attending to the use of words is not dependent on a philosophical ideology or doctrine of any sort. Attending to use can be compared to adjusting one's lenses. It is by trying to recall actual contexts of use that one can get an issue into focus, and thus make sure, for instance, that all the parties to a discussion are speaking about the same thing. Obviously, this does not entail that any particular use (such as 'everyday use' -- though as we shall presently see this notion is problematic in other ways, too) is held to be privileged. On the other hand, it should be obvious that if the parties to a discussion are to have a problem in common, this problem must be capable of articulation in a language they share. And besides, the need for 'adjusting one's lenses' is likely to be the greatest with respect to words that are used widely throughout the language: words like 'desire' or 'intention', or locutions like 'there is' or 'I know' are more likely to be variously understood than words like '? rays', 'desalination', 'polyglot', or 'absolute majority', since the latter, as it were, carry their context on their sleeve.

Now I should like to suggest that in very many cases this adjusting of lenses is all that is needed when we are faced with a problem: once it has been determined what the question was about, the question itself may turn out to have dissolved. When this happens, there is, of course, no reason to regret it (even though some may feel this entails 'an mpoversihed view of philosophy'): this enables us to move on to other, perhaps more substantial issues. However, I do not wish to be taken to advance a general thesis to the effect that this is what one should always aim at in philosophy, or that all philosophical problems will dissolve in this way. The reason I do not is that the notion of a philosophical problem does not have sharp limits, and hence it would not even be clear what it would mean to maintain that they can all be approached in one and the ame way.

2. There is a different objection to the idea that philosophy should concern itself with the use of words. Students are often disappointed that there is so much emphasis on words when philosophical problems are explained to them. They feel that this is a tedious detour on the way to the actual object of inquiry, the question of what really exists, or what we can know, or how the mind is constituted, or how we should live, or whether our actions are free, etc. This is a feeling that a number of professional philosophers seem to share with them. The feeling is evidently connected with the idea that there is a way of picking out the object of inquiry directly, without having to start by identifying uses of words. However, the idea of a separation between attending to the use of words and attending to the things themselves is more problematic than it may seem at first.

Let us first consider a case in which such a separation seems easily made. Suppose I am told to find out more about the anatomy of a butterfly or about the way a lock works. In such a case, I can and must set aside questions of the use of the relevant words and turn to an investigation of the thing itself. Of course I must be clear about what 'anatomy' means and what it is for a lock to work, but I cannot carry out the task simply by discussing the use of those words. Here there is a clear distinction between investigating the use of words and investigating the thing the words refer to. (The fact that there is, one might say, is a point about the way words like 'lock' and 'anatomy' are used.)

However, a problem arises if we treat cases like these as models of most of the matters discussed in philosophy. When questions are raised, say, about the nature of

time and space, physical objects, causal connection, existence, personal identity, knowledge and belief, intention or desire, etc, it is not clear what it would be to turn one's attention on the thing itself, as opposed to the role of those words. We have no clear idea of what it would be to pry these objects of inquiry loose from the situations in which people speak about them, to focus on the things in isolation from what may be said about them. Here, then, the model of investigating the functioning of a lock or the anatomy of a butterfly leads us astray.

In fact, this model may be what underlies the idea that a certain area of discourse, called 'ordinary' or 'everyday' discourse, might be considered normative in relation to other areas of language. This idea presupposes that different areas of use, for instance, the 'scientific' and the 'everyday' use of certain words, are in competition with one another in some sense, and that one might, accordingly, hold different opinions as to which way of using the terms is more adequate, some arguing that the everyday use of these terms represent their genuine sense, while others claim that the scientific use is truer to the actual facts. This whole way of talking, however, presupposes that these different uses of words are different ways of saying the same thing, and this, in turn, presupposes that the object of discourse can be identified independently of the discourse itself, an idea which models all of language on the terms we use in describing physical objects and processes.

The idea underlying this line of thought might also be expressed by saying that the sense of our words is displayed by what those words refer to, and that, accordingly, their sense can be identified independently of their use. But this would be putting the cart before the horse: it is only because I understand how a word is being used that I can understand what someone is referring to by it. If we are unclear about questions of sense, then, we should turn our attention from the objects of discourse to the nature of the discourse itself: to the role of the word in human interaction and conversation.

It turns out, then, that the idea that everyday uses of language should be considered normative for language as a whole (just like the idea that scientific uses of language should be considered normative), is incompatible with the point that meaning is constituted by use, when carefully thought through.

The idea that attention to use involves a gratuitous favouring of some areas of discourse at the expense of others is based on a misunderstanding of a different contrast: that between words in their actual use (whatever the context) and words considered in separation from their use. Philosophers who assume that we can investigate the sense of a word while ignoring any use that might conceivably be made of it, end up inventing uses of their own. What is then left open, however, is what bearing their claims are supposed to have on things said outside the philosophy classroom.

We might conclude, then, that it is an illusion to think that speaking about the uses of words, or, as it is sometimes put, speaking in the 'formal mode', as it were, is a philosophical detour, and that we can get at matters more directly by speaking about the things themselves, in the 'material mode'. In fact, when the philosophical fashion shuns appeals to use, the result is a curious loss of reflectiveness: a failure of philosophical method to know itself.

Being confronted with the questions of analytic philosophers (such as, 'do physical objects have independent reality?'), without allowance being made for asking, 'what particular context of use do you have in mind?' will sometimes make one feel as if one were faced with one of those questionnaires in which the alternatives provided never seem to fit the actual case (if you're on your way to a funeral, is that business or pleasure?), and told: do not ask what the questions mean, just answer them!

3. The point concerning the method to be used here could also, perhaps, be expressed as follows: while in most traditional forms of philosophy, there has been a tendency in trying to get clear about the nature of some type of object, to focus on clearcut instances of the type in question (the way one will, in the theory of knowledge, tend to focus on unquestionable instances of something being known), in this study attention will be given instead to the kinds of conversations that will be carried out in cases in which the application of a word will itself be subject to controversy, in the conviction that these controversies, and the forms of argument used in connection with them, will reveal more clearly than the uncontroversial cases the actual role of the word we are investigating.

(to be continued)

Introduction: the Problem

1.1 What makes my actions mine? What, to be more precise, is the basis for the distinction by which we attribute responsibility to a person for some aspects of her behaviour (such as carrying off goods from a store without paying for them), while we would not think her responsible for other aspects of it (such as giving a violent jerk on being frightened)? These questions are likely to beset someone who starts to reflect on the nature of human action. Unless a basis for the distinction can be found, it would seem to be completely arbitrary; thus it seems gratuitous or unjust, for instance, that people should be punished for some of the things they do, but not for other things.

A common response to these questions is to say that the relevant relation of a person to her actions is one that involves her will: a person is not responsible for her behaviour unless that behaviour is willed by her. On closer reflection, however, this response simply renames the problem, or raises it to a different level. A number of puzzles remain. What is it for an action to be willed by the agent? Is the will a particular kind of phenomenon, and if so, what kind of phenomenon is it? Is it an experience (i.e. something we are introspectively aware of), a dispositional state (i.e. something to be detected by observing a person's behaviour), or something else? Why should the fact that an action of mine is preceded or accompanied by such a phenomenon be thought to make me responsible for it? Is that phenomenon, in turn, under the control of my will, or does it come about independently of me? If the former, an infinite regress seems to ensue, if the latter, it seems unintelligible that I should be held responsible for what I will. Yet on the other hand, if we try to evade these difficulties by denying that the will is some kind of phenomenon, we are faced with the question what kind of thing it could be. Or is it nothing at all, an illusion?

1.2 Come to think of it, even if answers to these questions were found, that would not settle all the difficulties connected with the ascription of responsibility. We will sometimes hold a person responsible for another's will, not only in the indirect way in which, by threatening someone, I may force him to do something he does not otherwise want to do, but also by directly affecting his will in such a way that I may come to be considered, to a greater or lesser extent, the author of his actions. What I have in mind here are not sinister cases like that of Dr Caligari, who put his victims into a trance in which he was able to make them carry out his wishes (in these cases it might be thought that the victims' will had no part in the proceedings, though the matter is not, perhaps, quite as straightforward as that), but rather everyday cases like those in which we account for a person's actions by saying, for instance, that someone else had put her up to it, or that he had provoked, or seduced, her into doing what she did, and so on for a whole range of cases. I would argue that the importance of such cases - what we might, for short, call cases of undue influence - has not been fully realized in philosophy.

The way responsibility will be distributed in cases of undue influence depends on the details of the case: sometimes we blame only the party exerting the influence, sometimes the party who 'permitted himself to be influenced' (an interesting locution in itself), sometimes both parties. What is peculiar about some of these cases, at any rate, is that there seems to be a complete or at least a partial separation of responsibility (authorship) and will: though someone acted voluntarily, we may (it appears) assign some or all of the responsibility for her action to some other person. (This separation will be the most evident in cases in which the person subject to the influence is a child and the person doing the influencing is an adult or an older child, but it will take place in other cases too, as will be seen later.)

In such cases, we think of one person's will as somehow moulded, produced, by somebody else; or we may say that the agent's will, though she did what she did voluntarily, was not truly her own. To allow for cases of undue influence, then, we have to add one more question to those raised above: what must be my relation to my will (whatever the will is found to be) in order for it to truly to constitute my will?

1.3 I have here been linking the concept of the will to that of responsibility for action. In fact, philosophical discussions of the will have been almost exclusively focused on the dimension of responsibility, on the question of the link between will and action (or, as it can also be put, the role of the will in the link between the agent and her actions). In these discussions, the concept of will has been closely linked with that of intention. There is, however, another important dimension to attributions of will that has been very widely neglected. What I have in mind is the notion of a person's will as something that may be seen as demanding respect. Whereas the dimension of responsibility concerns the relation of a person to her own actions, the dimension of respect concerns the consequences of an expression of will for the actions of others.

I hope an example will help bring out the nature of this dimension. Sup-pose you are casually glancing through the morning paper when one of your fellow passengers on the commuter train asks to have the paper after you are through with it. The stranger in the railway compartment is nothing to you, you do not expect to see him again, nor do you have reason to think that he has anything other than a casual interest in the newspaper; even so, we may not find it in the least odd or surprising if you were to com-ply- with his request, even in a case in which doing so meant going out of your way.

In fact, the request might have a fundamental, though not necessarily visible, impact on your way of reading of the paper: it has now come to involve a choice between options: you may try to oblige the stranger by finishing the paper as quickly as possible, or you may be irritated by his request and drag out your reading just to spite him. To go on reading as if no request had been made has now become a balanc-ing act: while what characterized your reading up until then was that the way you read the paper was largely a matter of indif-ference -- maybe you let your eye roam more or less aim-less-ly over the pages, stopping at whatever item happened to catch your fancy -- it has now been transformed into a project with respect to which a question of your 'being through with it' can be raised: with regard to each item, the question whether you actually need to read it has become a consideration. In this way, the will of another person may encroach on what might be called the sphere of indif-ference in your life.

The fact that no great stakes are involved whether on the side of reading the paper or on the side of complying with the request, should help us see the kind of significance a request by itself may have: where the alternatives are evenly balanced, it may only take a request to tip the balance one way or the other.

One possible response to this is to say that what we have noted here is merely a psychological phenomenon, one that has no direct relevance for philosophical attempts to clarify the concept of willing: it is a well-known fact that people are, in many cases, moved by other people's requests, as well as by other people's expressions of pain, discomfort, anger, fear, etc. Or again, someone might argue that while it may be an admirable moral attitude to wish to comply with other people's requests, this is no part of understanding what a request is.

It is my contention, however, that it would be a misunderstanding to dismiss the example on some such ground. While the psychological phenomenon involved has its own kind of importance, the point I want to emphasize here is one concerning grammar: the stranger's request has consequences for the description of the situation, and these consequences, as I hope will become clear later, are partly constitutive of our understanding of the will.

If it were asked, 'Why should we respect a person's will simply because it is his will?' (i.e. what justifies our giving it any consideration at all, in the absence of any independent reasons?) the answer, in brief, is that what we consider a person's will is what we are prepared to respect. If there were nothing we respected as such, without external grounds, then there would be nothing we acknowledge as that person's will.

There are two sides to the grammatical point: first, the stranger's words, provided the other party understands them as the expression of a request, makes a difference to the ways of describing his responses that are available to us: he may be said to comply with the request, to ignore it or defy it; and, in doing what he does, he may come to be seen as overly submissive, obliging, indifferent, obtuse, etc. In other words, even if the request makes no visible difference to his behaviour, it does make a difference to the way that behaviour can be understood (by others, as well as by the person himself). In fact, I would claim that there is an internal connection here: to think of something as a request is to think of it as laying such a conceptual net over the situation in which it is uttered.

The second point is a closely connected with the first. The different ways of responding to a request do not stand symmetrically with respect to intelligibility. If I tell someone, 'I handed him the paper because he had asked for it', then by those words I have stated a reason for doing what I did. If she does not understand that explanation, she does not understand what a request is. If she wishes to question the explanation, the burden is on her to give grounds why, in the circumstances, it is not satisfactory. Or, turning the matter around: it is the failure to honour a request that requires a special explanation.

Here, it might objected that something's being a request cannot be a sufficient explanation by itself; rather, whether it explains anything has to depend on the

details of the case, for instance, on what the stranger asked for: if instead of the newspaper he had asked you to hand him your clothes, just like that, out of the blue, then that by itself would hardly have explained your compliance.

This counter-example, however, is only apparent. Merely uttering some sequence of words, whether 'Could I have your paper after you're through with it?', or 'Could I have your clothes, please?' does not constitute a request by itself; an utterance can only be understood, taken seriously, as a request, given the right sort of circumstances. In the case of the latter utterance, those circumstances would obviously have to be of rather a special kind (say, their occurring as part of a conversation between two secret agents). To imagine a situation in which those words can be understood in that way, however, is precisely to imagine a situation in which complying with the request would require no special explanation.

In sum, then, I would claim that understanding something as the expression of a person 's will is internally connected with its having certain consequences, on the one hand, for the description of the subsequent behaviour of those concerned, and, on the other hand, for the distribution of the burden of explanation of that behaviour.

1.4 The dimension of respect is connected with the notion of undue influence in two ways. On the one hand, the fact that people, as a matter of psychology, tend to be moved by a respect for other people's requests is one of the most important sources of the influence, whether legitimate or illegitimate, that one person may exert over another. On the other hand, if I come to see someone's expression of will as reflecting the undue influence of someone else, then to that extent the respect I owe his will is going to be undermined. Ther eis, we might say, an internal connection between the notions of undue influence and the reduction of authority of a person's expression of will.

We see the character of our respect for another's will in its purest form in the respect we normally feel for the will of the deceased. Someone may feel that the deathbed request, say, of a parent or spouse has put her under an obligation that she must honour at all costs. However, if a woman came to think that a request made by her father on his deathbed was not a genuine expression of his will, that could make her feel no longer bound by it. This might be because she had become convinced that his will had been unduly influenced by some other person. Let us suppose, for instance, that he had asked his daughter to make over a large sum of money to a nurse who had looked after him during his last illness. The daughter might come to suspect the nurse's motives, believing that she had never really had her father's best interests at heart, but had only tried to ingratiate herself with him in the hope of financial gain. In that case, whether or not her suspicions were unfounded (and of course, the whole issue of the purity of a person's motives may have no uniquely correct solution), she might possibly come to feel that the father's request was no longer binding on her. It had come to be something she could perhaps set aside without guilt.

Our respect for the will of the deceased is formalized in laws regulating the implementation of a person's last will, which, if duly recorded, is held to create legal obligations as far as the disposal of her property is concerned. By laying down terms for a bequest, a testator may indeed create obligations ramifying into an indeterminate future. And just as an individual may not feel bound by a deathbed request if she is convinced it came about as the result of undue influence, there are various grounds for contesting a person's last will.

The question of undue influence, however, may be problematic even in cases in which the purported victim is around to give her own side of the story. We may wonder, for instance, whether a man had succumbed to a woman of his own free will, or whether she had seduced him; and here his own words are not decisive, for even if he sincerely protests that his love is genuine, we may remain convinced that this conviction is itself part of the result of the seduction. On the other hand, he may come to raise the question himself in reflecting on his feelings. To decide that he had been the victim of a seduction will normally (though not, perhaps, necessarily) mean that his view of his own love for her will change; he may come to feel that it was illusory. His feelings might, as it were, lose their authority for him.

In fact, the discussion of whether someone's will was truly his or her own may be central in a great many different contexts of life: as when parents justify forcefully abducting a child from a religious sect and having her 'de-programmed' by claiming that the child had been brainwashed by the sect members and was no longer herself; or when a court has to decide whether a child's preference for his father is genuine or is the result of more or less subtle manipulation; or when someone accused of

providing drugs claims that an undercover police agent had coaxed him into doing something he would never have done of his own accord, and that hence he was not fully responsible; or when a regime declares that political protests in the country are not a genuine expression of popular sentiment, but are inspired by outside provocateurs. We might also think of the problem facing someone trying to deal with diseases like anorexia, which could be seen as that of getting the patient to regard her wish to starve herself as an alien will and her wish to eat as her own, and not the other way round.

The aim of this study is to try to get clear about what it is we discuss when we discuss such matters. Is there a basis for the distinctions we make in these cases that can be identified and then be used to resolve the issues, or if not, does that mean that in this connection everything is up for grabs, and that the arguments that will be produced on either side of such an issue are all taken out of the blue?

What I shall argue is that these disagreements are genuine; what gives them sense, however, is not -- could not be -- the fact that there is some distinctive phenomenon to which one can appeal in order to resolve them. The manner of their resolution can only be seen from case to case. To make up one's mind about them is to take up a certain relation to the problem.

In this study, I wish to discuss the nature of these disagreements, and at the same time I try to show how trying to identify them with phenomena of kind or another is misleading.

2 - The Will as a Faculty

2.1. In a celebrated passage, William James writes:

> We know what it is to get out of bed on a freezing morning in a room without a fire, and how the very vital principle within us protests against the ordeal. Probably most persons have lain on certain mornings for an hour at a time unable to brace themselves to the resolve. We think how late we shall be, how the duties of the day will suffer; we say, 'I must get up, this is ignominious,' etc.; but still the warm couch feels too delicious, the cold outside too cruel, and resolution faints away and postpones itself again and again just as it seemed on the verge of bursting the resistance and passing over into the decisive act. Now how do we ever get up under such circumstances? (William James, The Principles of Psychology, Volume II, New York: Dover, 1890, p. 524.)

What is the problem here? Perhaps there are several. First, the person lying in bed might have a problem. Day after day she has been late for work, only because she has this terrible problem persuading herself to get out of bed on time. Is there a solution to it? Maybe she should do something about the heating, if she can afford it, or maybe the real problem is that she hates her job. Can she get another? Or maybe she should try some psychological trick, some way of fooling herself or forcing herself to get up. Should she try placing the alarm bell further away so she cannot reach it from bed? Maybe getting a dog would do the trick?

We might think about other analogous problems. I am determined to tell my boss a thing or two, but when I am face to face with him, my courage falters, and I cannot bring myself to put things quite as boldly as I had planned. Or one might have a problem about ceasing to do something or abstaining from doing something rather than a problem about doing it, as when someone tells himself, 'This will be my last piece of chocolate; --- no just one more, that will definitely be the last, etc.' Or think of a case like writer's block, where the problem may be a combination of not being able to make oneself write and not being able to abstain from doing other things that let one postpone having to face the writing. How is one to overcome these kinds of resistance or temptation?

Second, there is, one might suppose, the philosophical (or psychological) problem of trying to understand the nature of the will - of what is involved in doing something voluntarily. (I shall address that problem in the next chapter.) Third, there is the philosophical problem of weakness of the will: how is it possible for someone to fail to do what she wants to do, or to do something she does not want to do, even though nothing external prevents her from acting or compels her to act?

Now for James, it seems, these three problems are interconnected. The person lying in bed, he seems to think, has a 'how-to' problem: how am I to engage my will? In

fact, James also claims to have a kind of solution:

> If I may generalize from my own experience, we more often than not get up without any struggle or decision at all. We suddenly find that we have got up. A fortunate lapse of consciousness occurs; we forget both the warmth and the cold; we fall into some revery connected with the day's life, in the course of which the idea flashes across us, 'Hollo! I must lie here no longer' - an idea which at that lucky instant awakens no contradictory or paralyzing suggestions, and consequently produces immediately its appropriate motor effects. It was our acute consciousness of both the warm and the cold during the period of struggle, which paralyzed our activity then and kept our idea of rising in the condition of wish and not of will. The moment these inhibitory ideas ceased, the original idea exerted its effects (op. cit., pp. 524 f).

In other words, provided we can keep from attending to those circumstances that speak in favour of remaining in bed, we may find ourselves overcoming the resistance without difficulty. Now there may be some psychological truth in this observation. For James, however, its importance lay in its providing evidence in support of his 'ideo-motor' theory of volition: on his view, the condition (the sole necessary condition) for a movement being voluntary is that it is preceded by 'a kinesthetic idea' of what the movement is to be (p. 493). This theory was James's answer to the task he had set himself: to discover the mechanism of production of voluntary movements; an answer he arrived at after having discarded a rival account, according to which what makes a movement voluntary is its being preceded by a sensation of innervation of the muscles.

The relative merits of the ideo-motor theory and the innervation sensation theory are not of concern to us here. Rather, what is of interest in this connection is the very idea of a theory of volitions and what it is supposed to achieve. For James, apparently, a theory of volitions will, among other things, be of practical use: it will suggest ways in which we may overcome an inner resistance to certain actions. Perhaps it could even be said that for him the real test of a theory of volitions is its practical usefulness for such purposes.

Now undoubtedly there is a place for such a notion of 'learning to will' things. We may develop methods of self-discipline, or think of ways of 'steeling ourselves' for a difficult task, and there will even be self-help guides for people who find they habitually have a problem maintaining their resolve in one context or another. However, there is a very simple objection to supposing that advice on how to overcome inner resistance could suggest an answer to the question what willing is: I can only have a problem about overcoming an inner resistance provided there is something I want to do in the first place. My looking for a remedy for my lack of resolve is itself an expression of the fact that there is something I want to do or abstain from doing.

2.2. What is peculiar about James's problem, as we have described it, is that it seems to presuppose that we may have an indirect relation to our own will. The will appears to be a kind of apparatus that we may try to engage, and may either succeed or fail in engaging. An obvious objection to that way of looking at things, as was just suggested, is that if I try to engage my will I must want to engage it. It is easy to see how this could generate an infinite regress. In other words, the problem is that of understanding how the will could be something external to me; how I can be passive, as it were, with respect to that in virtue of which I am active — how I could be at a distance from my will.

The idea of the will as external to me is not a notion dreamt up by philosophers. We often use locutions suggesting that the will is somehow external to the agent, in trying to express the difficulties we or others may have in trying to carry out resolves. We may do so by invoking inability, saying things like, 'I couldn't help myself', 'I couldn't bring myself to ... ', 'I couldn't make myself ... ', or we may do so by expressing an attitude towards our own actions that seems to suggest the matter is out of our control: 'I'm ashamed of the way I drink'.

Consider a case. Suppose Joe is telling his wife how he has just lost his job. He had been late a number of times lately and last Friday the boss had given him a final warning, 'Be late one more time and you're out of here.' And today it happened again. He just could not help it: he could not bring himself to get up in that cold room. He had lain awake for a whole hour before he finally braced himself, and by then it was too late. What is he expressing in saying he could not help himself? We may suppose that his wife is shattered by the news, and that he is full of shame at

his failure. Or then again he may be defensive, and tell her that it was just asking too much to demand that he should get up in conditions like those.

Whether Joe is shameful or defensive, his failure apparently makes sense against the backdrop of a certain kind of difficulty. This notion of a difficulty, I would suggest, helps us understand the way in which our will may be at a distance from us. We all know it may be hard to get out of bed in a cold room, that is why we know what he is talking about. Of course someone might retort, 'What d'you mean? It doesn't bother me one bit.' But by saying this he would show that he understands what we are talking about. He knows what the 'it' is that does not bother him at all. This is obvious if we imagine a different case. Suppose someone said: 'I just could not bring myself to open the door. I was standing there in the street and could not go in.' In that case, we should hardly know how to respond, until we are given a background which makes it intelligible why there should be a resistance to opening a door. (Does this concern all doors? Was he afraid of what he would enounter inside? Etc. Until we understand this, we could not say that this is a thing that does not bother us.)

But what is meant by saying that it is hard to get out of bed in a cold room? Suppose Joe had told his wife, 'You know what? The funniest thing happened to me this morning. Just as I had woken up and was going to get out of bed I suddenly found I wasn't able to. There I lay, all eager to get up and go, and nothing happened. It lasted for an hour. What do you suppose it could have been?' We might imagine all kinds of explanations to go with that: it might have been some kind of epileptic seizure, a passing bout of apoplexy, a muscular cramp or what have you. Either he was suddenly unable to engage his muscles, or the muscles were temporarily unequal to their task. It might be discovered that this condition somehow had to do with the coldness of the room, it was a form of paralysis brought on by a prolonged state of hypothermia. Even so, this is not what we normally mean when we say that it is hard to get up in a cold room.

Nor would the difficulty of getting up be like that involved in playing a game of pick-up-sticks, in which you have to make a strenuous effort to make exactly the right movements, the slightest error or lapse of attention resulting in failure, or like that involved in throwing darts, where no matter how hard you try, you cannot be sure of hitting the mark: all you can do is take aim, throw, and then see what happens. When we say it is hard to get up in a cold room, we do not mean that it requires a great deal of concentration if it is to succeed, or that all you can do is try, success being out of your hands.

2.3. The relation between the coldness of the room and the difficulty of getting up is of a different kind. It is not something one has to discover. Rather, I choose to stay in bed because I feel the coldness of the room. Or we might say: I am drawn to the warmth of the bed in contrast with the coldness of the room; this is how I respond to the situation. This is the sense staying in bed makes to me (which is not the same as saying that I judge that it does make sense to stay in bed).

The difficulties we described before were difficulties that were taken to explain why the man did not manage to do what he wanted to do. They are instances of some of the ways in which the discrepancy between his intention and his behaviour might be explained. They do so by showing that his behaviour was not expressive of his intention, but was due to what might be called external circumstances, i.e. he behaved the way he did because he had limited control of his bodily movements (paralysis), or of physical events, or both (pick-up-sticks, darts). But in the case of the man staying in bed, we cannot say that he failed to get up because the difficulty of getting up caused a discrepancy between his intention and his action. His staying in bed was (in one obvious sense) not unintentional. The difficulty operated by (so it seems) making him stay in bed voluntarily. In a sense this was a difficulty that his will made for itself.

In the case of the man staying in bed the role of the difficulty, it might be said, is not to explain a form of behaviour, but to take up a perspective on it that makes it possible to hold on to two apparently incompatible descriptions: the man wanted to get up, and yet he stayed in bed of his own free will. The concept of a difficulty, as it were, cancels the normal link between will and action, makes our descriptions of what a person wants or of what she does immune to normal objections of a certain form. (This view of the role of the concept of a difficulty is, I believe, clarifying to soem extent. It is also open to some objections, to which I wish to return later. But before I do I wish to take a closer look at the way we speak about difficulties in this type of context.)

It might seem natural to describe the conflict in terms of a play of forces, the warm bed pulling the agent in one direction, the thought of work in the other. The forces are not symmetrically related, they have opposite values. It seems natural to describe the attraction of the bed in negative terms, the will to go to work in positive terms, or maybe one could speak in terms of a force pulling downward and a force lifting upward. It also seems natural to identify the agent with the upward force: being strong means pulling oneself up, being weak means permitting oneself to be pulled down. Or one might say: the upward pull is dependent on the active participation of the agent, the downward pull happens of it own accord. The downward force is constantly present like gravitation, taking control as soon as the agent relaxes.

I wish to make a couple of comments on this way of conceiving of the matter. The words 'upward', 'downward', 'force', etc are obviously being used in something other than their ordinary senses here. They are used in what I should like to call a 'secondary' sense. However, I should like to claim that they are an indispensable part of our understanding of the situation, in the sense that there is no other, as it were more direct, way, of expressing the same point (I do not simply mean the problem of finding an exact paraphrase). If we were to eliminate these terms, the conflict or problem, I want to claim, would simply vanish. There would then be no weak-willed people, only persons who change their mind or who say one thing and do another. To consider something a case of weakness of will is to place the discrepancy between a person's words and actions in relation to such a play of forces, and this connection brings out the sort of discrepancy this is, i.e. what the relevant sense of 'doing something other than you want to do' is.

But since these words seem indispensable to a description of the situation it may be tempting to suppose that they must be understood literally. Pushes and pulls really are in operation here, it appears, and the problem is to establish what they consist in. It seems to me that our inclination to construe it in this way has come to shape the way in which we conceive of the whole problem, and made it seem unsoluble. I shall try to explain what I mean.

When it is asked what the forces pulling upward and downward consist in, the question, I should like to claim, is in an important way wrongly put. If we say that someone abandoned her determination we are not asserting that her will was overcome by certain external factors. Taking that literally, it would mean that she did not act of her own free will, that perhaps she had a firm and determinate will but had no control over her behaviour. But in that case she could not be said to have abandoned her determination. Consequently it has to be seen that the difficulty that makes her abandon her determination is one that her will, as it were, makes for itself. In other words, when I say that someone was late for work because she could not conquer her reluctance to get up, I am not putting forward a hypothesis about a possible cause of her being late. I simply describe what she did in a way that is intelligible against the background of our shared understanding of human action.

The idea of this play of forces, we may say, is constituted by the judgments we make about people's actions. When we say that someone stayed in bed because it is sometimes hard to get up what we say is understood because the utterance belongs to a shared way of speaking about actions. One might say, somewhat tautologously perhaps, that this form of discourse forms the moral universe in which our conversations about human action take place. (Donald Davidson evidently denies this, arguing that weaknes of will is a value-neutral notion.)
Part of what constitutes this shared form of discourse is the fact that not anything whatsoever can be invoked as a difficulty. Thus, if someone were to say, 'Yesterday I was dead tired and wanted to go to sleep rightaway, but I couldn't bear the thought of lying down in a warm bed, what with the room being so cold and all', this remark would probably be understood as a joke or a piece of irony. Only against quite a peculiar background could we take it seriously. Similarly, it would be hard to accept it as a case of weakness of will if we were told that the reason someone died in a fire, or the reason he failed to interfere when his wife was being assaulted in an adjacent room, was that he was unable to get out of bed. (We might conclude instead that we was mentally paralyzed, in a state of panic, or the like.) Since this shared understanding constitutes the background of the way we describe a person's behaviour, it cannot itself be taken to be conditioned by empirical discoveries.

2.4 Our inclination to give the force-related concepts used here a literal interpretation is connected with two other ideas which tend to distort our view of what is involved in weakness of will. The picture we have is of various inclinations to act, various 'dispositions', on a collision course. First of all, there is an assumption (perhaps mostly unspoken) that weakness of will is exceptional. It is a case of the normal course of events having somehow been made inoperative. Normally, it is

thought, there is a determinate answer to the question of what agents want, and no problem about their executing their wishes. I believe this picture to be misleading and to have given rise to difficulties. In questioning this, I do not mean to assert that the opposite is the case, i.e. that cases of weakness of will are the rule and that it is only in exceptional cases that we act in accordance with our will. The point I wish to make is that the question of a correspondence between a person's actions and her will is one that arises only in special circumstances, and that, among the cases in which it actually does arise, a failure of resolve is no more exceptional than its opposite.

According to Davidson weakness of will consists in performing an action in spite of judging that it would be better to perform a different action: 'If a man holds some course of action to be the best one, all things considered, or the right one, or the thing he ought to do, and yet does something else, he acts incontinently' (op. cit., p 21). And he asks rhetorically: 'Does it never happen that I have an unclouded, unwavering judgement that my action is not for the best, all things considered, and yet where the action I do perform has no hint of compulsion or of the compulsive?' (op. cit., p. 29) I find this characterization similarly problematic. Davidson is obviously right in pointing out that there is a conflict between doing X while agreeing at the same time that it would be better not to do X. However, it also seems strange to suggest that people generally do what they do because they judge it best. The normal thing is for the question of what would be the best thing to do not to arise.

Suppose we ask someone, say a person queuing in front of a newsstand in order to buy the evening newspaper, whether he considers what he is doing the best thing he can do in the circumstances. He would be likely to take our question as an impertinent hint that he ought to be doing something else (cp the policeman's question to a gang of teenagers loitering on the pavement, 'Are you sure you don't have anything better to do?'), but if we manage to convince him that we are neutral and curious spectators, he might well tell us he did not understand our question. 'The best thing in view of what?' he might ask. In most cases we do not think that what we are doing is the best thing we can do in the circumstances, nor on the other hand do we think for that reason that there is some better thing to do. Either judgment seems to presuppose a specific context, say a situation in which the goal to be pursued is clearly given, thus if I am in a boat caught in a storm at sea I might wonder whether trying to bail it out is the best thing to do, or whether one should concentrate instead on trying to keep the bow up against the waves. (On the other hand, it is not unusual for us to do things without stopping to consider that there may be objections to what we are doing that we could not sincerely reject if they were to be voiced. Hence conceivably the question of weakness of will might come to be raised in a great many more cases than those which it will in fact be raised. There seems to be no way of determining in advance whether some action might come to be seen weak-willed in a certain range of circumstances.)

In Davidson's argument there seems to be a play on an ambiguity in expressions like 'better to do' and 'the best thing to do'. We sometimes use them to express a comparative judgment: 'in consideration of all the relevant circumstances, this alternative is the best'. However, this presupposes that there is an understanding in advance of what the relevant considerations are to be taken to be, and this is hardly the case with respect to a great many of the actions we undertake. However, the expressions are also, and more commonly, used simply to express a decision. During a night at the pub I may get up and say, 'I'd better get home and get some sleep.' This does not mean: 'at this moment the utility of going home and getting to bed overtook that of staying on and chatting with my friends.' Nor does it mean that at this moment I find I have to revise my earlier calculation of expected utilities. It simply means, roughly, 'Now I'm going home', putting the decision in the context of the advantages of getting a good night's sleep. Saying that one had better do something and yet not doing it would either show that one did not mean what one was saying or that one had changed one's mind; it would not yield a case of weakness of will.

2.5 Another misconception connected with the inclination to understand the notion of force involved in weakness of will literally is the idea that cases of weakness form a clearly delimited category. On this view, the cases in which we lack the strength to do what we want to do are surrounded on one side by cases in which we carry out our will without difficulty, and on the other side by cases in which we lack the ability to act differently than we do. Thus, David Pears supposes that cases in which we are weak-willed can be distinguished from cases in which we are literally confronted with an irresistible impulse to desist from what we want to do. But this idea is problematic. What does it mean for an impulse to be literally irresistible? The literal / non-literal distinction does not sem to get a grip here. Could there, for

instance, be an objective method of determining whether or not in a given case resisting an impulse was within a person's power. How could this be measured? Could we design an experiment in order to decide whether or not an impulse was irresistible? What would that be like?

In fact I once heard about such an experiment from someone who had taken part in it. The test persons were to keep their arm in a bucket of cold water until the cold got unbearable. The object was to establish the amount of discomfort the test subjects were able to endure. The person who was relating the experiment kept her arm in until the experimenters felt obliged to ask her to take it out. She never did reach a the point at which keeping her arm in presented itself to her as unendurable. Evidently, she took the whole experiment too seriously. What confused her was her acceptance of the assumption on which the experiment was based, i.e. that will and ability are to be thought of as two independent variables: as if the impossibility of keeping one's arm in was something that occurred independently of one's attitude to what one was undergoing. In fact, my finding something irresistible or unbearable cannot be separated from my giving up the attempt to resist it or endure it. A spectator may disagree, accusing me of having given up too easily, but this would simply mean that he did not condone my attitude; there can be no neutral method of etablishing who is right in a matter like this.

What this suggests is that concepts like irresistibility or unbearability are not experimental categories, since the conditions for experimentally establishing their applicability can never be fulfilled. A person's ceasing to do something can only be taken to show that going on doing it would be unbearable provided she is sincerely making an effort to endure. And this means that precisely when she gives up, the test conditions are no longer fulfilled. In a sense, it could be said, what the experiment measured was the point at which the test subjects decided the experiment was not worth any more pain.

Considerations like these show some of the limitations of the applicability of comparing the will to a faculty.
(to be continued)

3 - The Will as Experience

3.1. Conceivably, there might have been a different (or additional) reason why William James wanted to focus on the case of someone trying to persuade himself to get out of bed, apart from those considered in the previous chapter. On this interpretation, it provides a case of someone launching into action from a state of passivity. Here the question about the nature of will apparently comes to a head, since, in distinction from the case, say, in which I act in immediate response to an event, there is nothing here besides the act of will itself to explain why I do what I do. If we are able to describe what happens at this moment, then, it might be thought, we shall have captured the essence of what it is to be an agent.

3.2. Now it seems natural to suppose that, if there is a distinctive event, an act of will, that occurs when we set ourselves to do something, then the event in question must be conscious. In other words, the act of will would evidently have to be something that we experience, and which puts us in a position to tell that we did what we did voluntarily. In this chapter, I wish to discuss what seems to speak in favour of such a view, and to point to some of the problems involved in it.
We can find expressions of the type of view I have outlined here in contemporary philosophy of action. For instance, Carl Ginet writes :

> ... when we voluntarily exert parts of our bodies, we experience this exertion and we experience it as voluntary. We experience the specific exertions that we make voluntarily, and it is part of this experience that the exertions seem to us to be ones that we control in a quite direct way (p. 24)
> ... this experience of voluntarily exerting my legs, even when most 'automatic', also differs radically from the occasional experience I have of exertions by parts of my body that I do not voluntarily control (ibid.).

(To understand what is being said here, it may be useful to know that Ginet uses the locution 'exert parts of our bodies' to refer to what we normally do when we perform actions; see p. 23.)

According to Ginet, what he calls 'the normal subjective experience of voluntarily exerting the body in a certain way' has two constituents. One is a kinesthetic perception of the exertion, that is, a direct perception of an exertion by some part of

the body. This would also be felt in a case in which the movement is involuntary. The second constituent is the experience of controlling this exertion, or, as Ginet also puts it, its appearing to me that the movement is voluntary. 'The voluntariness of the experience of voluntary exertion', Ginet says, 'is a further part of it, distinct from the perceptual part, an aspect that would be more conspicuous by its absence than by its presence' (pp. 28 f.).

Having this experience, according to Ginet, is a necessary condition for an action's being voluntary: an action cannot be voluntary unless it seems voluntary to the agent (p. 30). It seems safe to assume that he thinks the entailment holds in the opposite direction, too: on Ginet's view, then, I do something voluntarily if and only if my doing it is accompanied by the experience of voluntariness .

Now Ginet gives no argument to convince us that such experiences do occur, but this is only natural since he claims to be referring to an experience with which we are familiar. He admits, however, that we may overlook this experience precisely because of its great familiarity (p. 24), perhaps in order to explain why some philosophers have questioned the meaningfulness of speaking about volitional experiences.

3.3. We are not normally confused as to whether we did what we did voluntarily or not. This might make it natural to suppose that voluntariness (the initiating event) consists in a specific experience, something present to our consciousness, just like a toothache or a feeling of nausea. Just as we could not, presumably, have unconscious toothache or unconscious nausea, our movements cannot, we suppose, be unconsciously voluntary. (The issue of whether voluntariness involves a specific experience should be distinguished from the issue of whether wanting something may be an experience. I shall address that issue in the chapter on will as desire.)

Suppose that, while we are moving, my wife hands me an expensive vase. The vase slips from my hands, falls to the floor and is shattered to smithereens. One of my wife's reactions may be the angry question 'Now what did you do that for?'. The question is probably rhetorical. It is rather like asking, 'How could you be so clumsy?' She does not expect an answer to the former question any more than she would to the latter. All the same, I might have an answer to the first one. After all, maybe I did it on purpose. I may long have been waiting for an opportunity to get rid of the vase (I cannot stand it, perhaps because it was given to my wife by her former husband whom I wish to put out of my mind — and hers; or because, expensive as it is, it is in terribly bad taste, and I do not have the heart to tell my wife; or maybe I simply want to get back at her for having offended me).

Even if I do not tell my wife, then, I might be clear myself about why I let the vase slip. On the other hand, if it was an accident, I would probably be clear about that too. I am not here discussing the case in which I made a mistake about what I was doing: e.g. because I thought this was a different vase.

If I did it on purpose, I might conceivably have confided my plans for letting the vase slip to my diary: in other words, I could engage in what we might call justificatory discourse with respect to the dropping of the vase. (Or I might engage in exculpatory discourse of a certain kind, i.e. I might try to make clear why I had thought it a good idea at the time.)

On the other hand, is there such a thing as 'voluntary doing' in the absence of an ability to engage in justificatory discourse? Let us consider this possibility. The idea that voluntariness consists in having a certain experience seems to entail that there may be such a thing, since my having an experience does not seem to be dependent on what I am able to do before or after having it.

Could we, in other words, imagine the following case? Right after the vase drops, I confess that I dropped it on purpose. However, I am unable to give a reason for doing it. All I can say is that I had a very distinct impression of voluntariness just as the vase was beginning to slip from my fingers. But wait a minute! that does not seem to be the right way of describing what happened. If I speak about the vase slipping from my fingers, I am not describing something I do. There is nothing there that could be either voluntary or involuntary.

There is of course the case in which the vase starts to slip and I, voluntarily, do nothing to stop it. Maybe on a sudden whim I think 'What the hell! let it drop', perhaps looking forward to a bit of drama and excitement. But that does not fit the bill for two reasons: first, letting the vase drop on purpose is not the same as dropping it on purpose, and second, in this case I could in fact (if I am honest) give an account of what possessed me to let it drop (the anticipated drama and excitement).

What we need, then, is a case like this: I have a good firm hold on the vase, then suddenly decide to let go of it. Right afterwards, however, I have no answer to give to the question why I did. All the same, I insist that I did it on purpose. Now the story may go on from here in various ways. Maybe I simply shrug my shoulders, saying 'I don't know' (or maybe even 'how should I know?') when asked why I did it, or then again I may be as baffled by what I did as my wife.

In some cases it is quite normal to respond to the question 'why did you do it?' by shrugging one's shoulders and saying 'I don't know'. If someone asked me, for instance, 'Why did you choose that particular chair?' (suppose we are in a huge lobby and there are plenty of empty chairs all over), I might say 'I don't know', meaning, roughly, 'For no particular reason', implying that I am perfectly happy to switch if you like. The answer would be given in the spirit of saying that this is the sort of thing one will do without having any particular reason. But if I gave that answer to my wife's question why I let the vase drop, she might feel I was adding insult to injury: not only do I deliberately break her favourite vase, but I pretend that this is the kind of thing one will do without any particular reason: in other words, she will probably think I am not being sincere.

Suppose, however, I am sincere. In that case, I must be puzzled that my wife should wonder why I did it (I might ask her' 'Why do you ask?'). The case is now beginning to look somewhat weird. It seems to presuppose that my wife and I have a very different understanding of what I did; that we have radically different attitudes towards vases and the breaking of vases; a difference in outlook that seems hard to imagine in view of the fact that we have shared a house for a number of years, bought things together, for one another, etc. (It is not so easy to see what my attitude is really supposed to be. Don't I even care about the broken glass all over the floor?) Since the case is weird, it is hard to know what we should say about my action.

Consider, on the other hand, the case in which I am genuinely baffled by my action. All I can say about it is that I did it on purpose. Again, it is not so easy to imagine what the case is like. In saying 'I did it on purpose', or 'I wanted it to fall' I will normally be taken to shoulder responsibility for what I did. But what am I shouldering responsibility for in this case? Do I feel guilty about what I did? If so, what precisely do I feel guilty about? I did not do it to offend my wife, to start a scene, or the like. It was not a case of failing to control my temper, nor proof of my malicious nature, nor an expression of my indifference to worldly possessions. There was just my dropping of the vase, unconnected with what went before or came after.

Nor is it clear what there was about the dropping of the vase that I did voluntarily. My dropping the vase was not part of any project which required transforming the state of the world from one in which there is the vase to one in which there is not, etc. Did I simply not want to hold it, did I mean for it to fall on the floor, to break, to make a sound, to produce splinters, all or none of the above, or what?

Under these circumstances, we might be inclined to regard my conviction that I did it on purpose as a queer phenomenon. Without the context of justificatory discourse, it would not have the normal consequences of such a confession (are we even to call it a 'confession'?). If a small child were to say 'I did it on purpose' and then could give no account of why he did it, we should simply teach him that those words are not used in this way.

(People might suspect that I was trying to cover up the fact that I had lost control. Maybe I had been noticing lately that I had a tendency to drop things, and I did not want anybody to notice that there was anything the matter with me, because I was ashamed or did not want to worry them. But in that case, if I had my wits about me, why did I not think up a justification as well?)

Suppose we try to imagine this queer phenomenon extended in time, to five minutes, say. During this time, I carefully prepare the smashing of the vase. I clear a corner of the room, then ask my wife to get the vase out of its box and hand it to me, I walk with the vase to the corner, very deliberately raise it over my head and then smash it to the floor with all my might. Afterwards I say I did it all on purpose, but have absolutely no idea of why I did it. But now suppose that during my preparations my wife asked me what I was doing. Here, again, we can develop the story in different ways. If I say at the time, 'I'm going to smash that vase, but I have no idea why', it sounds as if I were acting under a strange compulsion.

Suppose, on the other hand, I do give reasons for smashing the vase while I am preparing to do it, although afterwards I will deny that those were my reasons, and maybe even argue against their being good reasons. Here a lot would depend on what kinds of reason I give at the time. If what I say is fully consistent with things I

would normally say and do, people might conclude that I had in fact been acting purposely in breaking the vase, but had then suffered a radical loss of memory; if the reasons I give are weird and totally out of character, they would probably think that I was suffering from some strange possession. (If I fail to give her any answer at all, a lot would depend on whether or not that would be out of character for me.)

It is obviously hard to say anything definite about what we might say in cases such as these, precisely because they are so out of the ordinary. At any rate, these cases suggest that when the claim of voluntariness is made in the absence of an ability to engage in justificatory discourse, there is a kind of thinning out of the whole notion of voluntariness: the attribution of voluntariness, in such a case, seems to be cut off from its ordinary connections with what we say and do. So we might say: the problem with thinking of voluntariness as an experience is that the question whether or not someone had this experience while he was acting, in the end, is not important; what matters to us is the way what he did was bound up with other things said and done by him. From this we might conclude that it is not very helpful to think of voluntariness as an experience.

3.4. To show more in detail what would be involved in thinking about voluntariness as a matter of having an experience, I wish to discuss a well-known neurological study that seems to take such a view of the will for granted. A research team headed by Benjamin Libet proposed to investigate neuronal activities connected with self-initiated actions. They claimed to have shown that our spontaneous actions are preceded by a characteristic change in the brain, a so-called readiness-potential, which can be recorded by EEG. What makes their study interesting to us is the way the experimental design seems to embody the view of attributions of will that I have been discussing in this chapter. The researchers take it for granted that if we are to study the will we must focus on an experience occurring at the time of performing a voluntary movement.

The occurrence of readiness-potentials had been discovered in studies done by other researchers, who had detected what was described as 'a scalp-recorded slow negative potential shift that begins up to a second or more before a self-paced act'. This discovery, they claim, 'appeared to provide an electrophysiological indicator of neuronal activity that specifically precedes and may initiate a freely voluntary movement.' What seemed to have been shown was that, when a person decides to do something, then although she may herself consider her action freely initiated and spontaneous, unbeknownst to her the action is preceded, by 1-1.5 seconds, by a specific change in the brain, which can be recorded by EEG.

Libet and his team claimed to be able to confirm the finding 'that cerebral initiation of a spontaneous, freely voluntary act can begin unconsciously, that is, before there is any (at least recallable) subjective awareness that a "decision" to act has already been initiated cerebrally' (however, in their study the time interval shrunk to between a quarter of a second and a second). They concluded that this 'introduces certain constraints on the potentiality for conscious initiation and control of voluntary acts' . I understand this to mean that, when we think of ourselves as having reached a certain decision at a given moment, the feeling that up until that moment we were free to decide to act or not to act as we chose is in fact an illusion, since the action, if we undertake it, will have been anticipated by events occurring in the brain up to a second before we are aware of making our decision. In other words, Libet and his team might be thought to have given empirical confirmation of Spinoza's claim that freedom is nothing but the illusion that we produce our own actions.

In one of the tests carried out by Libet's team, the test subject was to follow a spot of light revolving in a clocklike circle around a screen in front of him, and, 'when he felt like doing so, to perform [a] quick, abrupt flexion of the fingers and/or wrist of his right hand' (TCI, p. 625), and 'to note and later report the time of appearance of his conscious awareness of "wanting" to perform [this] self-initiated movement' (ibid, p. 627, italics in original). The overall result was that the onset of 'readiness-potential', as measured by EEG, preceded the time at which the subject reported being aware of wanting to perform the movement by between 1,055 and 240 milliseconds (thousandths of a second), i.e. between a second and a quarter of a second.

I do not wish to discuss the details of this experiment. Rather, my concern is with trying to uncover the outlook on the will and human agency that underlies this investigation.

An important part of the motivation for this study was the need to eliminate what the researchers considered to be weaknesses in the earlier studies of the phenomenon. Their idea of what those weaknesses were and how they should be eliminated throws interesting light on how they understood the notion of self-initiation. Their main criticism of the earlier studies was concerned with the fact that the practical requirements of the experiment may have imposed constraints on the subject, thus compromising the 'fully endogenous nature of the acts'. Thus, the number of acts to be performed within a given interval of time may have imposed a limit on the time in which to perform the act — in other words, the test subjects may have felt they had to hurry — and this and other factors may have acted as 'external controlling influences on the subject's initiation of the act'. (RP, p. 322.) In order to exclude this possibility, Libet's team held it to be of particular importance to make sure that the movements of the test subjects were genuinely self-initiated, endogenous, or 'freely voluntary' as the writers sometimes expressed it. In order to ensure this,

> [a]n additional instruction to encourage 'spontaneity' of the act was given ... to [one group of] subjects ... For this, the subject was instructed 'to let the urge to act appear on its own at any time without any preplanning or concentration on when to act', that is, to try to be 'spontaneous' in deciding when to perform each act; this instruction was designed to elicit voluntary acts that were freely capricious in origin. (RP, p. 324; also TCI, p. 625.)

3.5. Libet and his co-workers apparently thought that, as far as 'self-initiated' actions are concerned, there are two alternatives: either the behaviour really is initiated at the time reported by the subject, in which case it is (or at least may possibly be) brought about by his own decision, or else it is initiated at an earlier moment in time, in which case the subject's 'decision' can no longer make a difference. For this line of argument even to get off the ground, however, it must be taken for granted that voluntariness, if there is such a thing at all, is a matter of the agent's having the experience of deciding to act at a given moment in time.

Another underlying assumption is that in investigating the role of the will, what we are investigating is the mechanism by which purportedly voluntary behaviour is produced. To show that behaviour was genuinely voluntary we have to rule out its having been produced by an alternative mechanism.

In sum: either willing is a conscious initiating event, or there is no such thing as voluntary behaviour.

We should note that the authors use the terms 'self-initiated', 'endogenous', 'freely voluntary', 'spontaneous' and 'capricious' as more or less exchangeable expressions. Self-initiation, or spontaneity, for them, is apparently the absence of determining or constraining factors of any kind. Hence the need to eliminate, not only any physical constraints on the subject's behaviour, but, in addition, anything that would give the subject a reason to flex his fingers at one moment rather than another. This explains the idea that a 'capricious' action is the purest conceivable form of voluntary behaviour. The need to eliminate anything that could seem to provide a reason for performing the movement at any one moment is rather like the need to remove any external disturbances (such as heat or a draught) when weighing substances on a very finely tuned scale. Only by establishing a state of complete balance can we be sure to detect what effect the will may have on behaviour. As the authors put it, with evident approval,

> the simple voluntary motor act studied here has in fact often been regarded as an incontrovertible and ideal example of a fully endogenous and 'freely voluntary' act. The absence of any larger meaning in the simple quick flexion of hand or fingers, and the possibility of performing it with capriciously whimsical timings, appear to exclude external psychological or other factors as controlling agents... (TCI, pp. 640 f.)

This line of thought, however, is obviously based on a misapprehension of what it is to act for a reason. The authors seem to think of my having reasons for performing an action as somehow constraining my freedom of action, as though my reasons for acting were independent circumstances competing with my will for control of my behaviour ('external psychological factors as controlling agents'). To rule out any possible reason for performing an action, however, is to rule out the possibility that the agent might engage in justificatory discourse with respect to it. But as we have

seen, far from strengthening my status as an agent, the fact that I am unable sincerely to engage in justificatory discourse with respect to what I have done or am about to do would tend to diminish that status or to cancel it out.

One source of confusion here might be that the authors are running together different senses of the question why something was done. It is clear that, if they want to study voluntary behaviour, they must eliminate the possibility that the movement they are recording was actually produced by some causal factor beyond the agent's control. For instance, the experiment would have failed if it turned out that the subject's movement was in fact a spasm. But one would of course be grossly mistaken in concluding from this that a person's movements are not fully voluntary if there is any answer at all to the question why he did what he did. (By the way, contrary to the authors' assumption, the idea of acting capriciously or acting on a whim seems to get no foothold in a context in which it makes absolutely no difference what I do or when I do it.)

3.6. In running together two senses of the question 'why', the authors are running together two different grammars of discourse about behaviour. This should be evident if we consider the way they speak about the initiation of behaviour. They interpret the finding that some neuronal activity associated with performing an action takes place before the time the subject recalled initiating the action as follows:

> the brain evidently 'decides' to initiate or, at least, prepare to initiate the act at a time before there is any reportable subjective awareness that such a decision has taken place. It is concluded that cerebral initiation even of a spontaneous voluntary act ... can and usually does begin unconsciously (TCI, p. 640).

As the scare quotes surrounding the word 'decides' indicate, the authors are aware that the word is being used here in a somewhat peculiar way: of course, people make decisions, not the brain or some other bodily organ. However, they do not seem prepared to take this insight far enough. The problem involved in attributing decisions to the brain is not simply a matter of linguistic etiquette, as it were; as if it were clear to everybody what would be meant in speaking that way, although there is a slightly exasperating prohibition on saying it (the way one has to remember to use the word 'starboard' rather than 'right-hand side' in speaking of a ship, even though there is no danger of a misunderstanding). The problem here goes deeper: the closer we look at what is being said, the harder we find it to understand what could be meant. The passage cannot be read literally; but if it is not read literally, it seems to say nothing at all.

The difficulty about understanding what it would mean for the brain or the central nervous system to make a decision or to initiate something has to do with the way these notions are connected with the notion of making a beginning. Something shameful or remarkable was done; several people were involved; however, we may think it important to determine whose idea it was or who started it (one might speak about 'the moral importance of the beginning'). The question of who made the decision or took the initiative is often a question of who is to get the credit or to shoulder the blame for what was done; thus, if what was done was shameful, those who simply followed are usually less to blame than those who started it.

Suppose a husband and wife are having an argument. As will sometimes happen, the argument may come to turn on who started it, which may in turn be a matter of exactly when it started. John and Jane were discussing whether they should sell their piano or hold on to it. They would both agree that at first they were having a peaceful conversation; they would also both agree that five minutes later they were having an all-out quarrel, accusations flying back and forth. What they disagree on is precisely when they ceased to be friendly. Which was the first unfair or sullen or provocative remark, which justified the other party's offended reaction? When John said, 'So you've really decided you won't be playing any more?' was there a touch of irony in his voice? Was he needling her about never sticking to her plans? To be sure, they only bought the piano six months ago, when Jane decided she wanted to resume her playing. But he was not going to criticize her for changing her mind (he says). She was just too sensitive, maybe she was in fact ashamed because she never carried things through, and she projected her own feelings onto him. — But what did she have to be ashamed of (she replies)? Wasn't the money hers to spend as she liked? Besides, she had noticed the way he would frown every time she sat down to play. She knew all along he did not really want her to play. So she was actually

giving it up to please him. — 'But I never said a word', he retorts. — 'You didn't have to, I knew how you felt; you were jealous of my playing', she replies. Etc., etc.

Who is right? How are we to decide? Was Jane right in being offended, or was John's remark genuinely innocent? Cut out this piece of dialogue from the lives surrounding it, and we can make nothing at all or anything we like out of it. If we know nothing about their shared history, what they had been talking about just before, the kinds of conversation they usually have, their normal ways of responding, what kinds of life they lead and what kind of relation they have to one another, we cannot really tell what is going on here. We do not even know whether this disagreement is serious or not.

Nor would it help us to be privy to the thoughts of John and Jane. Very few arguments get started by one party or the other saying to himself: 'It would be nice to have an argument, let's see if I can't make her angry.' (It might even be said that a quarrel is not genuine if it starts according to plan.) This is an area in which the line between the intentional and the non-intentional is anything but sharp. We can imagine John feeling quite innocent at the time, being surprised that Jane flared up. But later maybe he feels guilty, comes to believe that there had indeed been a touch of irony in his remark, and apologizes to Jane. But neither does that necessarily settle matters: we might suspect that he is just to submissive, taking on the blame even when he is not to blame. And we might feel that Jane was guilty, even though we know she will never admit that even to herself. Or then again, maybe they both blame themselves, and we may agree with either of them or both. Anyway, admissions of guilt are not the ultimate proof of guilt.

There evidently is no neutral standard by which to settle the matter. In John and Jane's case, their view of who started it may be internal to their argument, or to their continued relation. For someone else to have an opinion on the matter would be for that person to take a position. Someone who had no relation to them might feel it makes no sense to have an opinion on the issue. (If John and Jane go to a marriage counselor, the counselor would probably be wise to keep her thoughts on who had started it to herself.)

Obviously, the question of who started it all is not a matter of who brought things from a state of rest into activity. The one who begins an argument is not necessarily the one who starts talking, but the one who changes the character of the talk into something hostile. If an angry mob sets fire to the house of an undesirable neighbour, the one who started it all is not necessarily the first to move in the direction of the house, or the first to strike a match. Someone else may have been egging them on. And that leads on to the question what constitutes 'egging someone on', which is a question of where the blame should be laid. (The arsonists would probably claim that it was the neighbour himself who started it all.)

The causality involved here, in short, is of a moral, not of a physical nature.

3.7. How is the notion of initiation to be applied to events in the brain? It should be clear that in saying that 'cerebral initiation even of a spontaneous voluntary act ... can and usually does begin unconsciously' (TCI, p. 640), the authors do not mean to be read as saying anything about where the blame or credit should go. They are not speaking about 'initiation' in the sense in which we speak about initiating an argument; they are not saying that we are wrong in attributing the initiative to John rather than to his brain. Nor are they claiming that it would be an excuse for an arsonist to say, 'I didn't start it, the whole thing was initiated in my brain'. (Or could we imagine John telling Jane, 'I wanted to surprise you with flowers, but then I found my brain had beaten me to it'?) But if they are not saying this, what are they saying?

As long as we keep ourselves strictly to physiological processes it does not seem clear what place the notion of initiation might have. If there is an unpredicted change, say, in the EEG pattern of the brain, the conclusion neurologists are likely to draw is that it was produced by a hitherto unknown cause, and they might go on to look for the cause, in the brain or outside of it. If they were to say, 'This particular change had no cause, it was a case of the brain itself initiating a process', one would have a feeling that the neurologists were not doing their job properly; the conclusion that a brain wave was self-initiated is not the sort of thing one would expect to come out of the observation of EEG patterns. We should distrust a neurologist who drew this conclusion the way we should distrust a car mechanic who told us the rattle in the engine was self-initiated rather than caused by the condition of the engine.

One might wonder what made the researchers think that the movement was initiated by this particular brain pattern. This description conveys the picture that the brain is

in a state of passivity, and then suddenly out of nowhere a flash of activity occurs. However, the researchers say nothing about the state of the brain before the occurrence of those distinctive EEG patterns. It would in fact be rather surprising to hear that the brain was at complete rest before the pattern occurred (if anything, that would make it sound as if they had discovered a genuine case of Cartesian interaction). The fact that Libet's group took no interest in what happened before the pattern occurred evidently shows that for them, what made this particular occurrence an initiating event was in fact what came after it, i.e. the subject's decision to act. What singles out this particular event as the initiating event is nothing but the fact that it immediately precedes the subject's decision. Hence the conclusion of Libet et al. that voluntary actions are not actually initiated by the agent at the time she imagines, but that the initiating event occurs earlier on in the brain, is based on an argument implicitly accepting the idea of the action being initiated by the subject at the time reported by her. 'Brain initiation', in other words, is nothing but a shadow cast backward in time by 'agent initiation'. Had they been consistent, they should rather have argued that initiation is not to be found anywhere along this route, which would have been another way of admitting that they had chosen the wrong route for studying what it is to initiate an action.

How did they come to paint themselves into this corner? Their whole line of argument gets going because they assume that the willing of an action, i.e. that which makes an action voluntary, is some event (say, an experience) occurring at a certain moment in time. And since it turns out that the action is already presaged at the moment we suppose the willing to occur, they are inexorably led to on the conclusion that it was not produced by will: hence voluntariness turns out to be an illusion. This problem could not have arisen had they thought of voluntariness, not as a matter of some occurrence taking place at a given moment in time, but as a question of the way an action fits into the surrounding life of the agent.

Previous work on the topic by Lars Hertzberg:

[1] 'Blame and Causality', *Mind* 84 (1975), 500-515. (Reprinted in Hertzberg¸ *The Limits of Experience*, 1994.)

[2] 'On Moral Necessity.' In Raimond Gaita (ed), *Value and Understanding: Essays for Peter Winch.* London and New York: Routledge, 1990. 102-117. (Reprinted in Hertzberg¸ *The Limits of Experience*, 1994.)

[3] 'Om att urskilja människan i en människa.' In Steen Brock & Hans-Jrgen Schanz (red), *Imod forstandens forhekselse - en bog om Wittgenstein.* Aarhus: Modtryk, 1990, 63-74.

[4] 'Primitive Reactions - Logic or Anthropology?' In Peter A. French, Theodore E. Uehling, Jr och Howard K. Wettstein (eds.): *The Wittgenstein Legacy*. (Midwest Studies in Philosophy. Volume XVII.) University of Notre Dame Press, Notre Dame, Indiana 1992, ss. 24-39.

[5] 'Voices of the Will.' Forthcoming in L. Alanen, S. Heinämaa and Th. Wallgren (eds.), *Commonality and Particularity in Ethics.*

[6] 'On Being Moved by Desire.' *Philosophical Investigations,* Vol. 18, No 3 (July 1995).

[7] 'Mitä hymyssä piilee?' [Finnish translation of 'What's in a Smile?'.] In Ilkka Niiniluoto and Juha Räikkä (eds.): *Tunteet*. Helsinki: Yliopistopaino, 1996. Ss. 237-250.

[8] 'Pain, Anger and Primitive Reactions.' (Forthcoming.)

References:

David Cockburn, 'Responsibility & Necessity', *Philosophy* 70 (1995).

Donald Davidson, 'Actions, Reasons and Causes', reprinted in his *Essays on Actions and Events* (Oxford, 1980).

Harry Frankfurt, 'Freedom of the Will and the Concept of a Person', reprinted in his *The Importance of What We Care about* (Cambridge, 1988).

Herbert Fingarette, *Heavy Drinking: The Myth of Alcoholism as a Disease* (Berkeley etc, 1988).

John Heil & Alfred Mele (ed.s), *Mental Causation* (Oxford, 1993).

Anthony Kenny, *Action, Emotion, and Will* (London, 1963).

Anthony Kenny, *The Metaphysics of Mind* (Oxford, 1989).

B. Libet et al., 'Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity', *Brain* 106 (1983).

Gilbert Ryle, *The Concept of Mind* (Harmondswoth, 1963).

G. F. Schueler, *Desire* (Cambridge, Mass., 1995).

Sean A. Spence, 'Free Will in the Light of Neuropsychiatry', *Philosophy, Psychiatry and Psychology* 3 (1996).

Charles Taylor, 'What is Human Agency?', reprinted in his *Human Agency and Language* (Cambridge, 1985).