Talk given at the
"Mind and Society" conference
Cambridge, 2 Sept., 2005

THE PSYCHOLOGY OF VOLITION: "PROBLEM AND METHOD PASS ONE ANOTHER BY"

Lars Hertzberg

"Who is behind all this?"

When several people are involved in a shameful, controversial or forbidden activity, we sometimes consider it important to decide on whose initiative it all began. This may be a matter of allocating *responsibility*: the agent, in the truest sense, was the one who instigated the action, and brought the others along with him. When two children get into a fight, we may try to settle which of them provoked the other. In a criminal case, it makes a difference if one of the accused put the others up to it. The person inciting a riot is held more responsible than those who merely followed.

On the other hand, the question of initiative may concern the *authority* of an expression of will. Politicians may try to downplay the significance of protests by attributing them to the influence of foreign *provocateurs*, thus arguing that the unrest is not a sign of genuine popular dissatisfaction. Family members may contest a person's last will cutting them out of their inheritance in favour of his nurse, by arguing that the nurse put him up to it: it was not really his will. In a child custody case, the court may attempt to take into account the child's own choice of which parent to live with. In such a case, one party may argue that the child's expressed preference was due to interference by the other. The fact that an expression of will was influenced by someone else is thought to diminish its authority or to abolish it altogether.

In the latter cases, what is at issue is the responsibility of *the persons being addressed* rather than the subject whose will is in question. They may be reluctant, for some reason or other, to accede to the will that she is explicitly expressing, and they may defend their reluctance by questioning whether she *really* means or is behind what she says. They may claim that in ignoring or defying the other's request, they are not showing lack of respect, but rather they show respect for her *real* will. In both kinds of case, one may raise the question whether what her words or actions expressed was "fully her will" or whether

and to what extent it was shared or owned by someone else. (This problem is sometimes discussed under the title "autonomy".)

Volition and the readiness-potential

Evidently, in many cases, such issues are not to be resolved by simply asking the subject whether she really meant what she said. The problem of sincerity aside, we may be convinced that, as the saying goes, she does not know her own mind. We might call it a *psychological* problem, and suggest that one would have to be a good psychologist to be able to decide what she really wants. The question I wish to consider is: could we take this suggestion literally, and turn to experimental psychology to resolve the problems that arise in our lives concerning the human will? Could psychology supply us with the know-how required for allocating responsibility?

Let me approach the issue by discussing a famous investigation that may seem to have some bearing on it. In 1982, a research team headed by Benjamin Libet published a report on a series of experiments, in which they had set out to investigate neuronal activities connected with self-initiated hand movements. The team claimed to have shown that these spontaneous actions are preceded by a characteristic change in the brain, a so-called readiness-potential, which can be recorded by EEG. 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". When attaching an electrode to the scalp over the motor/premotor area of the cortex that is taken to control the hand, the recording shows a rise in activity culminating just before the action. This discovery, they claim, "appeared to provide an electrophysiological indicator of neuronal activity that specifically precedes and may initiate a freely voluntary movement."[1] What had apparently 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, and even the decision to perform the action, is preceded 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

[1] B. Libet, E. W. Wright, Jr. and C. A. Gleason, "Readiness-Potentials Preceding Unrestricted 'spontaneous' vs. Pre-Planned Voluntary Acts", *Electroencephalography and Clinical Neurophysiology* 54 (1982), pp. 322-335 (henceforth "RPP"). The quotation is from p. 322.

least recallable) subjective awareness that a 'decision' to act has already been initiated cerebrally". (However, in their study the time interval shrunk from 1.5 second 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"[2]. 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 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 the 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. On this finding, then, the distinction between having initiated an action oneself and having been put up to it by others has no basis in reality: *people* never actually initiate 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.

Here, I should like to discuss the outlook on the will and human agency that underlies this investigation. The study was partly motivated by the need to eliminate what the researchers felt were weaknesses in 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. They were critical of the earlier studies because they thought that the practical requirements of the experiment 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 imposed a limit on the time in which to perform the act — in other words, the test subjects may have felt

[2] Benjamin Libet, Curtis A. Gleason, Elwood W. Wright and Dennis K. Pearl, "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential)", *Brain* 106 (1983), pp. 623-642. (henceforth "TCI"). The quotation is from p. 623.

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.) To exclude this possibility, Libet's team thought they had to ensure that the movements of the test subjects were genuinely self-initiated, endogenous, or "freely voluntary" as the writers sometimes expressed it. In order to do so,

> [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. (RPP, p. 324; also TCI, p. 625.)

Volition as experience

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. Now, for this line of argument to get off the ground, 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. *The volition is concentrated in this experience.* Unless that assumption is made, the experiment shows nothing surprising: nobody has questioned that there might be distinctive occurrences in the central nervous system just before we perform a movement.

In a more recent article, Libet makes his commitment to this assumption explicit in spelling out the operational definition of free will used in the experiments:

> First, there should be no external control or cues to affect the occurrence or emergence of the voluntary act under study; i.e. it should be endogenous. Secondly, the subject should feel that he/she wanted to do it, when to do it or not to do it.[3]

[3] "Do We have Free Will?", in Benjamin Libet, Anthony Freeman and Keith Sutherland (eds.), *The Volitional Brain: Towards a Neuroscience of Free Will* (Thorverton: Imprint Academic, 1999), pp. 47-57. Henceforth referred to as "FW". The quotation is from p. 47.

This definition, he claims, accords with common views. We shall get back to the first condition later.

However, the idea of identifying voluntary action with a specific experience is problematic. This is brought out in the instruction to the test subjects to *"let the urge to act appear"*. There are two ways of understanding this instruction. "Letting the urge appear" might be taken to mean that he should *make* it appear: so, rather than simply flex his fingers, he should produce in himself an urge to flex them, and then watch as the urge produces the flexing – or it might mean that he should wait for an urge to appear, and then do nothing to stop it. On both readings, the agent is made out as having a peculiarly divided relation to his voluntary action: in both cases, there is something he is in a position to do on his own, as it were (*bringing about* the urge or *letting* it appear), and something for which he is dependent on the occurrence of some event (*the urge causing the movement* or *the urge appearing*).

In as far as I have to wait for something to happen, however, I am not in control. Actually, any account that equated the notion of being in control with having this or that particular experience would face a similar intractable problem.

According to Libet, the volition experience is what tells me that I am in control of my movements. In support of this, he points out that

> [m]any actions [or better, movements] lack this second attribute. For example, when the primary motor area of the cortex is stimulated, muscle contractions can be produced in certain sites in the body. However, the subject … reports that these actions were imposed by the stimulator, i.e. that he did not will these acts. (Ibid.)

The existence of such cases, however, hardly shows that some experience is distinctive of voluntary action; all it really shows is that an agent can normally tell whether his movements are voluntary or not. There are two sides to being able to tell. On the subjective side, there is the agent's inclination to *say* he was or was not in control. On the objective side, there is the fact that what he claims usually *fits into the context of life* in which he is acting[4]: e.g., his voluntary actions usually make sense or he can give reasons for

[4] Of course, there may be exceptions to this.

them, whereas his involuntary movements do not; on the other hand, they can often be given a causal explanation.

Now, the assumption seems to be that if I can tell, there must be something *from which* I tell. I must have a feeling or experience of being in control. What can be questioned, however, is not whether people do have such a feeling or not (after all, how would one go about deciding that?), but rather whether any experience can have the role attributed to it by Libet. I would have to have discovered that whenever I have such and such a feeling I am in control of what happens.

Now there *is* such a thing as discovering one is in control. In moving through an empty building, I may notice a recurrent noise and try to find out where it comes from. After a while I may discover that I am producing the noise myself: I cause it by stepping on a certain board. Here I discover that I control the noise. But this I can only do because I discover that the noise is produced by something I do. On the other hand, is there some way I could find out that *stepping on the board* is something I do? Not if my body functions normally. Evidently, it is only because I do not need to discover that I am in control of certain things that I can find out that I am in control of other things. Learning to tell what I can control could not get started from my learning to recognize a *feeling* of being in control.[5]

"Being capricious"

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.

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

[5] We sometimes use the locution "I feel as if p" or "It feels as if p" simply as a way of reporting or expressing the state I am in, without laying any claim to the truth of p. In a peculiar state of mind, or if I am drugged or drunk, I might, say, have the feeling that I am controlling the traffic lights. (Having this feeling need not entail that I believe I am in control.) But this kind of feeling could not be what gave people the idea that they could control things. On the contrary, it is because we are normally able to tell what we control that we may sometimes have this peculiar feeling of controling things.

constraining factors of any kind. They take this to exclude, not only physical or psychological constraints, *but anything that would give the subject a reason to flex his fingers* at one moment rather than another. This explains the idea that what the authors call a "capricious" action is the purest conceivable form of voluntary behaviour. The need to eliminate anything that could be a reason for performing the movement at any one moment is perhaps thought about along similar lines as the need to remove any external disturbances (such as heat or a draught) that might interfere with the measuring of some subtle physical process. 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* [my italics] timings, appear to exclude *external psychological or other factors* as controlling agents... (TCI, pp. 640 f.)

Of course this would mean that, if I not only have a reason for doing what I do, but my reason for acting is bound up with some present occurrence to which I am responding, my action is even less free. As Libet puts it, "A quick reaction to an unwarned stimulus also lacks a preceding RP, *and* it is not a freely voluntary act" (FW, p. 52; my italics)[6]. Thus Libet sees no distinction between my reaching out to stop my camera from falling to the floor and my dropping the camera when startled by a strong explosion. Neither response is "freely voluntary".

In fact, Libet's discussion has some analogies with an earlier treatment of freedom of the will. I am thinking of William James's discussion of the problem of getting out of bed. In a celebrated passage in *The Principles of Psychology*, 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

[6] It is not clear whether Libet takes this as additional evidence that voluntariness and RP go together, or whether RP is here being treated as a criterion of voluntariness.

> 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?[7]

One reason why William James chose a case like this may have been that in order to get the role of the will into focus, we should consider a case of someone *launching into action from a state of passivity*. Here, it appears, the question about the role of the will 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.

In the Libet case, the agent has no reason to do one thing rather than another; in the James case there is a deadlock between the urgent need to get up and go to work and the unpleasantness of exchanging the warm bed for the cold room. What they have in common is that there is no motivating force in operation, driving the agent to do one thing rather than another. This is precisely what seems to make them suitable as paradigms for the study of volition *provided* the will is taken to be a force of its own beside our various motives etc (freedom of will = "freedom of indifference"), the operation of which is most clearly seen when the different factors motivating us to act are either passive or deadlocked.[8]

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 that having reasons for performing an action somehow constrains one's freedom of action, as though one's reasons for acting were independent circumstances *competing with my will* for control of my behaviour ("external psychological factors as controlling agents").

The authors seem to be 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

---

[7] William James, *The Principles of Psychology,* Volume II, New York: Dover, 1890, p. 524.
[8] Cp also George S. Howard and Christine G. Conway, "Can There Be an Empirical Science of Volitional Action?", *American Psychologist* 41 (1986), 1241-1251.

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 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. On the contrary: we should be baffled if we were to ask someone who appeared to be carrying out some activity in a normal fashion why he was doing what he was doing, and it turned out that he did not have a ready answer for us. In certain circumstances (though the case is not easy to imagine), this might make us conclude that he was acting under some strange compulsion.

In fact, contrary to the authors' assumption, the idea of acting *capriciously* or *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. Being capricious means acting with disregard for whatever reasons may have a bearing on one's action. Hence, in the test situation, the only way the test subjects *could* have acted capriciously would have been by not flexing their fingers at all, or by disobeying the instructions in some other way, say, by deciding in advance when to flex them.

On initiation

In running together two senses of the question "why", the authors are running together two different grammars of discourse about behaviour. This is evident, too, in 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 around 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. (In "Do we have free will?", Libet writes, *without* scare quotes, "The brain was evidently beginning the volitional process in this voluntary act well before the activation of the muscle that produced the movement." (p. 49).) Perhaps it will be thought

that the problem involved in attributing decisions to the brain is simply a matter of *style* or 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 – rather in the way that, on board a ship one has to remember to use the word "starboard" rather than "right-hand side", even though everybody knows what one is talking about anyway. But actually, the problem goes much deeper: in fact, the closer we look at what is being said in saying that the brain makes decisions, the harder we find it to understand what could be meant. The passage cannot be read literally; if it is not read literally, on the other hand, it is not clear whether anything at all is being said. To speak about the brain making decisions is to invoke the image of a little person, a *homunculus*, lodged inside the skull, registering impulses and calling the shots. Of course, no one would take that image seriously. I want to argue, however, that unless a *homunculus* is tacitly assumed, Libet's entire project collapses. [9]

Consider the idea that the occurrence of the readiness-potential constitutes the moment at which an action is originally initiated. Why is *this* particular occurrence singled out? Evidently, there are recordable processes going on in the brain all the time, as revealed by EEG. If not, that would mean that the brain is dead. But why should the *rising* curve be considered the beginning of the volitional process? Why should this occurrence rather than any other stage in the process leading up to it be called *the beginning*? After all, the specific changes in the EEG curve are hardly produced *ex nihilo*, but rather they reflect neurological occurrences each of which is connected with other occurrences in accordance with the laws of neurology. The only reason to single out this particular change as

---

[9] In fact, as far as linguistic etiquette is concerned, the shoe is on the other foot: using anthropomorphic terms in speaking about the brain or the central nervous system seems to be the accepted practice, say, in psychology textbooks or in popularized science. Consider, for instance, the following textbook passage about what it takes to brake a car:

> ... the brain must know where your foot is as well as where you want it to go. The brain must contain some sort of register of the position of the body parts relative to one another, which is used to plan directed movements. ... A specialized part of your brain receives continual *feedback* from leg and foot muscles so that you are aware of how much pressure is being exterted and can alter your movements accordingly. (R Atkinson, R Atkinson and E Hilgard, *Introduction to Psychology* (8th ed), Harcourt, Brace, Jovanovich 1983, p 31.)

The process is evidently very complicated. There are the things *you* know, and there are the things *your brain* knows. Some of these things the brain evidently keeps to itself (the brain knows exactly where your foot is, but it obviously does not want to trouble you with the information), other things the brain lets you know as soon as it finds out about them, etc. (I wonder what else this brain knows; does it know it is a brain, for instance?) The brain does some planning of its own, but at the same time it is obedient to all your wishes (it must know where you want your foot to go). The brain, in short, is something like a very efficient slave.

interesting, it seems, is that it immediately precedes the agent's reported decision to act. *Neurologically speaking*, there seems to be no compelling reason to suppose that the process leading up to the movement might not have been initiated, say, by some event occurring 30 minutes earlier.

"But hold on!" someone will exclaim at this point. "At least the process can´t begin before the test subject has been given the instruction to flex his fingers." Well, how do we know that? We can only make this claim by drawing on our everyday understanding of human action. But it is precisely the validity of this understanding that is supposedly being tested in the experiment.[10]

Our inclination to regard the rising curve as the initiation of the act, it appears, is a reflection of the way in which we are inclined to imagine the role of the brain, rather than based on an empirical discovery. In fact, Libet more or less gives the game away, when he writes,

> the actual initiating process in the brain probably starts before our recorded RP, in an unknown area that the activates the supplementary motor area [which is thought to be the source of the recorded RP] in the cerebral cortex. (WF, p. 51.)

Once the door is opened to speculation about unobserved processes in unknown areas of the brain, we are pretty far removed from any idea of an *empirical* inquiry into the neurological initiation of human action.

Starting an argument

Talk about initiating something, being behind something, and the like, normally has a place in discourse about human affairs, in the contexts in which praise and blame are apportioned. For instance, starting an argument means transforming a conversation that is not unfriendly into a hostile one. Trying to decide who did this on a particular occasion is often a matter of trying to decide who first responded in a way that was not justified by what had gone before. This may not always be easy, or even possible. Suppose, for instance, a husband and wife are having an argument. Who started it? Which was the first unfair, sullen, impatient, provocative or snide remark, which justified the other party's offended reaction? If we ask them, they probably will not agree; rather, many quarrels

[10] Actually, if we pursue this line of thought, we end up having to acknowledge that we do not even know what we are testing, since we rely on being able to communicate with the test subjects.

sooner or later come to *turn around the very question of who started it*. A neutral bystander might have a clear idea, but she might also reach the conclusion that both were equally to blame, that the discussion gradually escalated into an argument without there being any one point at which the debate had become heated. (Or she might think it was all just due to an unlucky chain of misapprehensions. Neither, at first, had meant anything bad, but she could see why each of them may have thought the other one did.)

There seems to be no given standard for settling such a matter. Cut out this piece of dialogue from the lives surrounding it, and we can make nothing at all or anything we like 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.

In any case, in apportioning blame in such a case we attribute the change of the conversation to someone (to one party or both). It is ultimately a question of what we consider just. Obviously, there cannot be any sense in looking for a corresponding change in neurology. For one thing, with neurological occurrences there is no issue of justification. For another thing, if a neurologist fails to trace an occurrence to the stages preceding it, she would not conclude that she had discovered the *start* of something. Rather she would either believe that she had made a faulty observation, or that our knowledge of the processes in question was deficient.[11]

In making it seem as if our actions were initiated in the brain, I have tried to argue, Benjamin Libet and his team have performed a conjuring trick. His experiment rests on three assumptions, all of which I have tried to show are questionable. There is no specific experience by which we tell whether or not our movements are under our control. For an action to be fully voluntary does not mean that it is performed without reason or motive. There is no ground for singling out some neurological occurrence as constituting the initiation of our voluntary actions.

[11] Nor would it help us to be privy to the thoughts of the two parties. People who argue do not usually start out by *deciding* to argue. (It might even be said that a quarrel is not genuine if it starts according to plan.) We might even recognize, say, that the husband had no bad intent in making some remark, but yet he was being thoughtless since he should have realized that, in the circumstances, his wife was bound to be hurt by it.

If my criticism is correct, Libet has not been able to show what he claims to have shown (indeed, it might be questioned whether a project such as his made any sense in the first place). As for the problem we started out with: whether psychology can help us allocate responsibility by identifying who initiated some joint action, this outcome should lead us to suspect that we have been barking up the wrong tree all along. The idea that we might allocate responsibility by identifying who initiated the activity is misguided, since an attribution of responsibility is already embodied in the notion of taking the initiative. It would be futile to seek a factual foundations for our judgments, since *the relevant facts are constituted* by our judgments.

Åbo Akademi University
Finland