

Free Will and Events in the Brain

Grant R. Gillett

Bioethics Centre, University of Otago.

Abstract:

Free will seems to be part of the romantic echo of a world view which predates scientific psychology and, in particular, cognitive neuroscience. Findings in cognitive neuroscience seem to indicate that some form of physicalist determinism about human behaviour is correct. However, when we look more closely we find that physical determinism based on the view that brain events cause mental events is problematic and that the data which are taken to support that view, do nothing of the kind. In fact that view meets some substantial objections which turn on the role of the contents of our thoughts and experiences in explaining our behaviour. When we look at mental or psychological content we find it is governed by rules and not just causal laws. Rule-following is an activity which invokes the role of the thinker as rational agent and this is not a causal type of explanation. The fact that the thinker as agent is important means that when we invoke the way that a subject thinks about the world, we conceive of that subject as located in a socio-cultural context. This in turn requires that we recognize an explanatory circle according to which neither brain science nor social science can claim priority in explaining human behaviour and it follows that, for the purposes of psychological explanation and the understanding of the brain as an organ subserving a system of representation, we are bound to regard human beings as reasoning beings who exert some control over their own behaviour and not just as physical systems. This defeats the causal determinist position.

Requests for reprints should be sent to Grant R. Gillett, D.Phil, M.D., Bioethics Centre, University of Otago, Box 913, Dunedin, New Zealand.

It is commonly held that a thoroughgoing acceptance of what we might call the Aristotelian claim – that mind or *psyche* is a set of functions of the human body – is incompatible with any robust doctrine of freedom of the will. The argument usually goes through the twin premises that all mental events, including actions or intentions, are physical events and that physical events are a more basic level of explanation than mental events (Feigl, 1970). The latter claim is thought to imply that physical or brain events cause their mental counterparts. I think that this claim is false and the argument leading to it fallacious. To defend my position, I will argue that the findings normally used to support the claim that brain events cause mental events are mistakenly interpreted. I will further argue that psychological explanation -- the explanation of behaviour in terms of thoughts, attitudes, beliefs, desires, and so on -- is a type of explanation that is not conceptualizable in terms of strict deterministic causality but rather embeds an assumption that the psychological subject is a free agent. If I can make a case for that view then it entails that in so far as I am a being who has thoughts, intentions, and beliefs and who does certain actions, I am not a deterministic being. Seen in this light we find that the study of human behaviour retains a place for psychological or mental explanation based on the idea of the deliberating individual as an emergent phenomenon. On this basis we can say that the human mind is a useful part of our conceptual apparatus and has a positive explanatory role in human affairs.

Philosophical Distinctions in the Problem of Freedom of the Will

In discussing freedom of the will, philosophers usually distinguish two different conceptions: the freedom of indifference and the freedom of spontaneity, both of which derive from Hume (Kenny, 1973). The *freedom of indifference* can be thought of in a strong and weak version. The stronger claim is that antecedent conditions have no effect on a human action. Sartre (1958), for instance, claims that one is always free to make up one's mind and that no antecedent psychological states control or cause one's behaviour.

No factual state whatever it may be (the political and economic structure of society, the psychological "state," etc.) is capable by itself of motivating any act whatsoever. Human reality therefore appears as a free power besieged by an ensemble of

determined processes. (pp. 435, 441)

A more modest claim which admits certain effects of antecedent events is that the action is, in some sense, *up to me* to finally enact -- antecedent conditions do not causally determine my action.

By contrast, the *freedom of spontaneity* thesis claims that my actions flow from my mental life and are explained by antecedent mental states. As such, the thesis can be interpreted causally, as, for instance, by Mill (1, 843/1973)

... a volition is not an efficient, but simply a physical cause The volition or state of mind, a state of our mind, is the antecedent; the motion of our limbs in conformity to the volition, is the consequent. (I p.355)

Mill, so as to provide an account compatible with natural science, here contrasts an efficient cause (usually thought of as the origin of some effect but which we might think of as a teleological or final cause) -- the purpose which gives origin to the process leading to the effect -- with a physical, material or directly antecedent cause which produces the effect in accordance with a law of nature. He restricts his discussion to the latter. This allows the freedom of spontaneity to be compatible with the idea that my actions are caused by their antecedents according to scientifically discoverable laws. In order to make the freedom of spontaneity compatible with a psychological version of causal determinism we need to add two further premises to the argument.

C1: *The causal thesis* states that all genuine explanation is causal explanation (Kim, 1993).

C2: *The completeness claim* is that causal explanations form a complete system which, when fully laid out, yields a set of jointly sufficient conditions for all events including human actions. This claim entails that an action is an inevitable result of its causal antecedents.

Mill explicitly espouses this position:

given the motives which are present to an individual's mind, and given likewise the character and disposition of the individual, the manner in which he will act might be unerringly inferred. ... If we knew the person thoroughly and knew all the

inducements which are acting upon him, we could foretell his conduct with as much certainty as we can predict any physical event. (1843/1974, II, p.837)

The *causal thesis* and the *completeness claim*, taken together with the freedom of spontaneity, give us *strict determinism*. Strict determinism is incompatible with any kind of indifference thesis and, most of its adherents would argue, incompatible with a robust conception of freedom of the will of any type. Most people make good these claims by an appeal to the fact that mental or psychological states are fundamentally states of the brain which are, as physical states, covered by the causal thesis and the completeness claim. So much for an initial statement of the view which tends to lead to determinism and against human freedom.

A Sceptical Response

Faced with this set of alternatives one might feel inclined to make, on behalf of human psychology, a sceptical claim supported by introspective or phenomenological data. We could argue as follows:

It seems to us that we are free and therefore, barring good evidence to the contrary, we are entitled to believe we are. One might then observe that psychology has not come up with any robust psychological laws so there is not any good evidence to the contrary. This implies that either *the causal thesis* or *the completeness claim* is flawed, but it is unclear which.

We might even turn this scepticism into a philosophical argument in the style of Davidson's (1980) anomalous monism. Davidson accepts the causal thesis but attacks the completeness claim arguing that the mental realm could not possibly form a complete or closed domain of lawlike causal explanations because all sorts of non-mental happenings like historical and physical events must be incorporated into mental explanations for them to make sense. Therefore, mental explanation is anomalous or non-lawlike explanation in the sense that we will not discover complete scientific laws governing mental events even though we may be able to say that some event explains another causally because of the realisation of mental events by brain states. For instance, one might say that Mary's father's death caused her to be apprehensive about a long term relationship with any man even though there is no general rule covering this explanation.

Many determinists find this response unconvincing, and assert that mental events (including actions or behaviour) can be nothing other than brain events and therefore are within the general domain of physical events with its general physical laws. Tightening up the determinist or physicalist case we might then add the following claims:

All explanations of physical events are causal explanations (as Kim [1993] would argue).

Actions are physical events like any other.

This yields the physicalist claim about the causal explanation of actions:

Any action is a physical event and therefore must be causally explained.

Taking this together with the completeness claim, we then conclude that a full explanation of any and all of my actions can be derived from a complete physical description of the causal system of which I and my action are parts. This conjunction of claims is certainly a serious challenge to any believer in free will. It also, however, gives believers in free will a clearcut target at which to aim. First we might look at *the mental content objection* which aims to prove that an action is not just a physical event.

The Mental Content Objection

The mental content objection is that human actions are not just bodily movements but count as actions only because of the mental content which informs them, for example, my act of going to the door is that action (rather than walking towards the west or pacing the floor) only if I have the door in mind; similarly, trying to pour wine only happens when I have wine in mind and not just on any occasion when I am handling a bottle which happens to contain some wine. This is a point about the metaphysics of action (or what makes an action an action) which can be made vivid by exploring the case of a wine glass, a tic, and a party.

Imagine that I have a tic or involuntary movement arising from an impulse which periodically emanates from my brain. This causes my right hand to move upwards

and away from my body stopping at about chest height two feet to the right of me. On the occasion in question I and another man, Stephen, are talking to a very attractive woman who is flirting with both of us. At a key point in the conversation when it seems that the flirting has a certain intent, my hand jerks up and to the right hitting Stephen's wine glass and causing Stephen to spill his wine all over his shirt. After apologies and a reference to my troublesome tic (therefore casting me in the role of the innocent sufferer of a mysterious disease), Stephen goes off to change his shirt and I invite our charming interlocutor to come for a walk in the hotel gardens.

The question is: Did I act so as to take Stephen out of contention or did my tic fortuitously but involuntarily do the job for me (along with eliciting a certain amount of concern and sympathy, no bad thing in the circumstances)?

This crucial question can only be answered by determining the psychological or mental content that informed my action; was it intentional or merely automatic? Notice that in either case it was produced directly by my brain activity in accordance with causal laws forming a complete system of explanation -- but whether or not it was an action (the action of spilling wine on Stephen and thereby eliminating my competitor) or a tic depends totally on the answer to the crucial question. We might therefore conclude that actions, even if they are physical events, can only be correctly identified and described by appeal to their psychological content and that their nature as actions depends on the psychological content concerned having an explanatory role. If they are explained by psychological content (the intention to spill wine on Stephen) then they are actions, if not (i.e., if they are explained only by a pattern of brain events causing a certain movement), then they are not. These moves tell us that the essential or fundamental nature of the act depends on a fact about psychological content rather than physical causes and effects. That blocks the following argument:

- i. All explanations of physical events are causal explanations.
- ii. Actions are physical events.
- iii. Therefore, a full explanation of any action can be derived from a description of the physical world.

Thus we are required to find another way to establish the claim that actions are causally explained by their physical antecedents. One way of doing that would be to prove, first of all, that psychological events were determined by their antecedents despite their having psychological content and not just a physical nature. This would require some form of psychological determinism: *psychologically contentful events are causally explained in a deterministic way.*

Psychological Determinism

There are several versions of psychological determinism but any one of them will do as a target for the argument that follows. Psychological determinism is the thesis that the causal antecedents of any psychological state are to be found in the history of the psychological subject experiencing that state. In any of its forms, an historical thesis such as this is open to a version of Wittgenstein's attack on Freud's theory: "[W]hen we think of causal laws in relation to physical things we think of *experiments*. We have nothing like this in connexion with feelings and motivation" (1966, p. 42). Wittgenstein observes, as would any philosopher of science, that Freud did not prove any causal story because he did not demonstrate the following twin counterfactual claims:

Cpositive: If C (in conditions O) then E; and,

Cnegative: If not C (conditions O being equal) then not E.

These are jointly and sufficiently necessary to establish the counterfactual claim required for causality: *if C had not happened E would not have happened.*

In the face of this objection, a psychologist can point to the same problem that afflicts all those who seek to explain human behaviour: the counterfactual requirement is an impossible demand because one cannot, in general, manipulate the condition C for events in the past. We cannot switch on and switch off the causal antecedents of historical human events which have led to current psychological states and therefore we must proceed by other means to establish the required counterfactual claim. We can infer the truth of the required claim because the mechanisms that we have observed make sense of

human behaviour and have explanatory value in trying to understand human thoughts, attitudes, and actions. We therefore have what suffices in many areas of science: an inference to the best explanation.

When one reflects on this it is clear that the most that inferences can establish about the causal effects of life history on present psychology still falls short of psychological determinism. Determinism requires (in relation to psychological explanation) a certain rigidity and mechanistic conditioning of events by their antecedents. This, in turn, requires either the physicalist claim about the causal explanation of human action or some other version of the causal thesis and completeness claim. However, the present argument has only done that if it is conjoined to the idea that psychological explanation is causal explanation and that such explanations are complete; which is exactly what is in question. Because this begs the question at issue, the psychological determinist therefore falls prone to *the prior assumption objection* – he or she has assumed there was an antecedent causal explanation and has interpreted the life story to fit, rather than to obtain evidence concerning just which psychological causation of the required form is the best explanation.

Therefore we must find another way to get to the claim that a full explanation of any and all of my actions can be derived from a complete physical description of the causal system producing them. The most direct way is the *causal brain mechanism claim* which aims to establish that a brain mechanism causes human actions.

The Causal Brain Mechanism Claim

One might argue that, whatever else you say about them, actions are bodily movements arising from impulses generated by the brain and its complex causal or mechanistic processes. On this basis, an explanation of human thoughts and actions of the type required by physical determinism is always, in principle, available. We can support this claim first, by refuting an objection based on unpredictability, and second, by offering a positive evidential reason to believe that the causal brain mechanism produces behaviour.

The unpredictability objection states that actions are not mechanistically predictable like the effects of a causal mechanism and therefore are not the result of a

lawlike mechanism. The reply is in the form of an argument from complexity; I will call this *the Boston airport machine defense*.¹ There is a machine in Boston airport in which balls are moved around through complex paths. You can watch for hours (through its transparent but sealed cover) to see where the balls will go and how they get there. As an observer you might try and pick where the balls will emerge, but after a while you give up, stymied. The unpredictability depends upon a complex set of interacting mechanistic processes with variable time relations to each other. That this is so prompts the following thought apropos of the present debate.

Imagine a Boston airport machine of the following type.

1. Every process is mechanistic.
2. You can put the ball in whenever you like.
3. None of the mechanism is hidden.
4. You are allowed to place a 40 cent bet on any of the four possible chutes the marble will emerge from and if you are right you win one dollar.
5. The machine is of such complexity that you cannot predict and therefore you win one in four times (chance odds) so that for every dollar you collect you pay \$1.60.

This thought experiment (not too removed from reality in these cases) is meant to make vivid the possibility that there might be totally mechanistic processes that are so complex that they produce unpredictable results for all actual observers under all practical observation and calculation conditions but which are nevertheless deterministic. Any neuroscientist will tell you that the brain is much more complex than the Boston Airport machine and therefore the thought experiment undermines the unpredictability objection.

Alongside the causal brain mechanism claim, some would identify a positive evidential reason to think that our actions are products of our brain states. Firstly, we might turn to Libet's experiments which are thought to show that certain physical events precede (and therefore cause) the mental events to which they are most plausibly related (Libet, 1985; Spence, 1996; Young 1986). Second, we might follow McGuigan (1997) who claims that mind is a by-product of a complex set of cybernetic operations in the physical body and therefore mental concepts like free will ought to be replaced by explanations in terms of physical processes (like feedback loops). We shall attend to these two positions in turn to assess their significance for the current debate.

Libet's Physicalism

Libet attempted to show that the mental cause of an action was itself only a reflection of a physical cause constituted by a preceding brain event. J.Z. Young reports on the Libet experiments as follows: "Libet and his collaborators have now been able to show that this readiness potential change in the brain occurs up to half a second *before a subject mentally decides that he intends to make a movement*" (1986, p.73).

Libet (1985) attempted to relate mental events to their physical concomitants. More specifically, he tried to time the subject's first awareness of an intention to make a movement in relation to the first detectable indication of this intention in the brain, i.e., the "readiness potential" and other physiological events preceding the eventual movement (see also Brunia and Haagh, 1986). To do this Libet employed a method for ascertaining the time of mental events. He used as a reference point the perception of a moving spot in a particular position on an oscilloscope screen. The subject had to report the position of the spot when she became aware of her intention. The time of the intention was then assumed to be the same as the time the spot was actually in the reported position. This gave the "time of the mental event" preceding voluntary movement and could be related to the EEG record which detected the "readiness potential" -- an electrical event in the brain which precedes voluntary movement. Libet timed the initial conscious awareness, as reported by the subject, as occurring about 300 - 500 ms later than the readiness potential and argued as follows.

- i. The mental event of intending to act causes the act.
- ii. The physical event precedes the mental event of intending, and is unconscious; therefore
- iii. The physical event precedes and *causes* the mental event of intending.
- iv. Therefore conscious events are determined by their physical counterparts.

Let us call this *the argument to physical priority* as it seems to establish that the readiness potential is the real cause of the conscious intention to act and, as it is a brain event, one has reason to believe that the causal antecedent of action is indeed a material or physical cause.

However, three philosophical issues need to be raised about Libet's experiments, each of which is relevant to the relation between mental and physical events assumed in *the*

argument to physical priority. They concern:

- (1) the relationship between conscious experience and human action;
- (2) the concept of simultaneity at play here, i.e., between two mental experiences and between a mental experience and a physical event;
- (3) Mill's causal antecedent model of intention and action (assumed in the experiment).

Recall that Mill assumed that an action is a bodily movement produced by a mental event called an intention. This latter event is then timed according to a report by the subject and correlated with a brain event preceding the same bodily movement. However it is by no means clear that every action is preceded by some conscious event properly called "an intention to act." There are a number of things I just do in the course of everyday activity and some of these are actions which are extended in time (adjusting the brightness of the television screen for instance). These are better thought of as a pattern of activity guided by a conception rather than as discrete acts preceded by formed proto-acts called intentions. On the discrete act view, it is, in fact, very difficult to sort out just how many intentions are involved in most of what I do. For instance, did I have an intention to type F, and then o, and then r, and then space, or did I only have an intention to type "For" when I typed the first word of this sentence? Does my intention to drive off in the car necessarily reduce to a series of discrete intentions to get the key out of my pocket, open the door, get in, put the key in the ignition, turn the key, and so on or is it not quite so episodic as such a chain of events might suggest?

In fact when we consider the ways that actions happen, the action theory of Vallacher and Wegner, according to which an action is best thought of as behaviour structured by a guiding conception, seems much more appropriate.

... people have in mind a certain idea of what they are doing or want to do and use this prepotent identity as a frame of reference for implementing the action, monitoring its occurrence, and reflecting on its attainment. (1987, p.4)

The "identity structure" theory advanced by Vallacher and Wegner implies that when we explain an action we do not seek to identify the events which preceded it but to locate it as a

movement in a complex structured pattern of interaction with the environment whereby we can see what the subject is doing (Gillett, 1992a). This pattern may be extended in time and may only make sense as we catch on to the guiding conception being pursued by the subject. The guiding conception is itself a conceptual construction fabricated out of the cognitive resources of the subject to meet that subject's predicament at that time. Thus the origin of the action is not just an event which we and the subject can observe and report on but an ongoing cognitive move which is intelligible or explicable in the context of the life project being lived by the subject. Given that the idea of action gets us into this kind of framework of thought we are no longer dealing with a causal chain where one event can be considered (in itself) to be the necessary and sufficient condition for bringing about the next in the way required by material causation (Gillett, 1992b). If this is so, the model used by Libet is not the best to use in considering the voluntary control of behaviour that is in question in freedom of the will. We shall return to this concern and I will adduce certain other points about the nature of the concepts that are an inextricable part of the individuation and explanation of action once I have examined Libet's claim on its own merits. In fact, apart from these worries about the model implied in (i), there are reasons to suspect that Libet's experiments do support premises (ii) and (iii) of *the argument to physical priority*, viz that the physical event precedes and causes the intention and is outside conscious awareness.

Experience and action. Any attempt to establish experimentally a precise relationship between mental events and an associated brain event depends on the following three assumptions: (a) that any conscious act is reportable on the basis of introspection; (b) that such reports are the only means of ascertaining the occurrence of conscious acts; and (c) that it is the reportable conscious act immediately prior to the action which is its mental "cause." Libet explicitly accepts (b), in that he "begins with the premise that [the subjective event of becoming aware of wanting or deciding to act] is only accessible introspectively to the subject himself" (Libet, 1985, p. 532). He also worries about (a) referring to the possibility, which cannot be excluded because it is not empirically testable, that "a nonrecallable phase of a conscious urge exists, so that the reported time would apply only to a later, recallable phase of awareness" (p. 535). This seems to imply that he thinks the failure to report could arise only from a failure of recall or access, rather than, as I have suggested, from a misconception of the model of action that informs the argument to

physical priority. It also paves the way for the idea that there are subconscious events which precede the conscious intention to act.

However, contra both the introspectability and reportability claims I would argue that there is no particular reason to think that conscious acts are the kinds of things that can be dated and are reportable as events. It is not difficult to perceive introspectively that there are shades and nuances in our mental acts which are not fully represented in what we can report about them to others. One might, for instance, come to think that a hero we have held in high regard has feet of clay and find oneself hard put to date the moment of coming to that belief. Or one might decide to take a more relaxed attitude to life and work and yet be unable to say when exactly one formed that resolve. We are content to report what we can about our conscious acts but it is a further step to claim that this is all there is to them. The biological function of reporting is presumably to transmit information about the environment which has been acquired by sensory means (Millikan, 1993, p.13ff). If this is true then a secondary function might be to let others know how it is with us in terms of what thoughts we are entertaining but it would seem to serve no adaptive function to report on the moment of forming an intention. We seem to beg the question as to what the function of consciousness is -- i.e., to provide reportable material on our inner lives -- if we assume that whenever there is a conscious experience, this must correspond to brain processes of such a kind that reporting on them would be biologically useful.

Let us concede, for the sake of argument, that conscious acts are, in fact, essentially discursive and therefore reportable. Libet would still have to show that the timing of the report (in whatever form) gave an accurate "fix" of the time of an event which was the (disputed) intention to act. To do this he needs some criterion for determining the time of the intention. Although Libet's subjects were asked to report the *earliest appearance* of the specified "awareness" of the intention, it is possible that preparatory mental activity occurs immediately before the subject "decides" or forms the intention to A. Indeed, it seems incredible that there would be no prelude to the realisation that one intends to A. Therefore, even if there are cerebral events going on before the subject is aware of forming the intention to A, that does not secure the conclusion that the mental preliminaries to A are preceded by cerebral preliminaries -- it merely implies that the reportable intention to A does not arise *de novo* without any mental/cerebral preparation. It may be that these

preliminaries are what Libet is picking up.

In fact, the relationship between mental events and the reporting of them seems considerably more tenuous than Libet assumes. Therefore it is questionable whether the reports of “intention” made by Libet's subjects should be regarded as having a key significance in the explanation of action. In an experiment involving compression of the upper arm, subjects were able to move their anaesthetised hand as instructed by the experimenter, while firmly denying that they were doing so (Laszlo, 1966). This suggests that what is interpreted as “intention” by subjects, at least in their verbal reports, may be sensory feedback from the relevant muscular movements, as a result of the normal functioning of the intact sensory-motor loop, rather than the events initiating such movements at a cerebral level. These worries in and of themselves make Libet's claims look slightly naive but there are further damaging philosophical considerations.

Simultaneity. Suppose that we have a reportable conscious act. Then the only public data to which we have access is the report of that event, which must be assumed to occur some time after the event itself. Hence the timing of mental occurrences is problematic. Libet attempts to overcome this problem by asking his subjects to relate the “intention” to the time of another mental event. The subject is instructed to look at a moving spot on a screen -- so that he or she experiences what Libet calls “a stream of visual sensations” from this spot -- and to pick the sensation which appears to coincide with the awareness of the intention. Libet assumes that the time of the visual sensation is readily determinable by reference to the time the spot was in the position selected by the subject, and that this pinpoints the time of the intention.

There are two weaknesses in this method. First, it is questionable whether the time at which a subject experiences a spot in a particular position is identifiable with the time of the spot actually being in that position. To investigate this Libet carried out a comparison experiment to ascertain the accuracy with which subjects were able to time external events using the moving spot method. According to Libet, "a skin stimulus was delivered at an irregular, randomized time after the start of each trial and the subject reported the time of his awareness of that stimulus [... thus] the discrepancy between the subject's reported timing and the actual stimulus time could be objectively determined" (p.534). Libet found that three of the five subjects reported the spot position which indicated a stimulus time 100 ms

earlier than it actually occurred. Thus subjects reported the mental event -- awareness of the skin stimulus -- before the stimulating event. Understandably, this conclusion is unacceptable to most theorists, including Libet himself, who describes the observation as an "error." Libet appears not to have regarded this error as shedding doubt on the validity of his method for determining the time of mental events. However, there are explanations which spring to mind which do shed doubt on his assumptions.

One plausible explanation is that the subject reconstructs the course of events from a scanning or interpretation of the sensory stream (a position espoused by Dennett, 1991, pp.153-162; Gillett, 1999b). Whatever the true explanation, the skin stimulus experiment undermines confidence in the moving spot method of timing conscious events, and leads to considerable uncertainty about time relations between "mental events" (in terms of introspective identifications) and physical events. If the subject cannot tell when something as determinate as a skin stimulus occurs, then it seems unrealistic to expect the subject to give an accurate and precise time at which an intention is formed. Think, for instance, of the remark "I just knew I had to pull the trigger, but don't ask me the exact moment when I knew that." Libet, however, forces this impossible exactitude on his subjects as if it were normal to be able to specify such things.

Another effect may also operate. The subject's reaction time, in the sense of the time that elapses between her awareness of an intention to make a movement and her registering a certain position of the moving spot, might well be significant. The subject might think, "I think I am aware of an intention to make a movement; is that really definite enough for me to register? Yes it is, now where is the spot?" If this is plausible, then again it mitigates against Libet's position on the time order of the mental events involved.

If we adopt a more fluid conception of brain function, the very idea of "simultaneous mental events" and determinate time relations to physical events begins to appear deeply problematic. We tend to assume, by analogy with a common sense view of physical events, that each mental event M_a has a determinate temporal relation to any other mental event M_o , i.e., that M_a happens before, after, or at the same time as the M_o . However, if we consider brain processes, the before-after relationship is really only applicable when we are talking about spatially determinate and identifiable events at synaptic junctions or nerve cells. Nobody, not even the most committed physicalist, pretends to understand the exact shape of

the physical event corresponding to a given mental event (as Dennett and I have argued). Thus it is not clear when the pattern of excitation and quiescence comprising the posited “physical event” begins and ends. In fact, we are increasingly coming to see that the idea that there might be neatly packaged brain events corresponding to identifiable mental events is physiologically implausible. The brain has a complex stream of activity in which it is unclear where processes start and stop and patterns of excitation are, at best, approximations to processes with some identifiable informational content (even though highly specific patterns can sometimes be found, particularly in individual neurones). The move towards connectionist accounts and locally holistic models of brain function has, for instance, undermined the mechanistic physiology of much traditional neuroscience. Therefore the claim that a physical process in one part of the brain constitutes an event which can be timed in relation to another event elsewhere looks to be false.

This gives us further reasons to suspect not only Libet's conclusion (shared by other neurophilosophers) that physical events precede their mental counterparts but also the whole idea that causal determinism has much to say about human mental life. In summary,

- (1) We should question the naive assumption that conscious intentions are isolated “events” whose occurrence in the mind for the purpose of introspective reports can accurately be timed.
- (2) Libet disregards the difficulties of fixing accurately the time of any “mental act,” even a simple one such as detecting a sensation from the skin.
- (3) Libet relies on a naive model of brain function according to which activity proceeds from identifiable state to identifiable state and there are discrete brain events apt to correspond to the mental events as referred to in the two previous contentious claims.
- (4) The model of behaviour in which an *intention* equals a *mental event preceding and causing action* is itself suspect.

It seems that the argument from experimental data to the priority of the physical over the mental is flawed and theses concerning supervenience or even epiphenomenalism must also give up any claim for support from Libet's psychophysical observations. We can now turn to

McGuigan who argues that cognitive events should be reduced without residue to physical (neuromuscular) events.

McGuigan's Physicalism

For over thirty years McGuigan was engaged in research on covert motor activity associated with cognitive processes (1966, 1978, 1994, 1997). He typically monitored EMG patterns in the tongue and other speech muscles and finds that subvocal or covert speech behaviour occurs during much of our cognitive activity including such things as silent reading, problem solving, memory tasks, and learning. McGuigan (1997, p. 355) found that electrically measured motor activity from striated muscle was activated specifically according to the cognitive demands of the tasks being undertaken (1997, p. 355). In summarizing his extensive experimental work, McGuigan claims that:

covert muscular movements are present during cognitive activities even when efforts are made to eliminate them” and “that with progressive relaxation, cognitive effects are eliminated. ... We thus arrive at the following proposition: the human mind is the functioning only of systems of the body. As those systems selectively interact through neuromuscular circuits according to cybernetic principles, they (1) generate cognitive processes (the contents of mind, some of which we can verbally report) and (2) program behaviour, both overt and covert, to accomplish our purposes. (1997, p. 367)

Thus, all our thinking proceeds in concert with neuromuscular activity and is dependent on that activity. When we plan to act (or realize a conception in practice) we depend on bodily mechanisms to do the work of structuring the conception. For McGuigan this indicates that the brain-body complex as an interacting whole is the locus of control for action and is causally dependent on the forces operating on it. If we run with this view, it is plausible that cognitive activity is normally adapted to manipulations of a physical environment or an overt signaling system and that by suppressing the overt manifestations of that activity we allow the pace of association and computation to be accelerated beyond that allowed by its (physically acted out) evolutionary and

ontogenetic forbears. On this reading, the characteristics of our holistic activity, involving both body and mind (as a reflection of the patterns discernible in what we do in the world) would only be susceptible to physical explanation if the patterns that it evinced were able to be explained by physical variables and forces. I will defeat this last move (crucial to deterministic physicalism) by arguing that physical variables cannot explain the patterns in our activity which our brain-body complex cybernetically encodes. That is the thrust of *the rule-following objection*.

The Rule-Following Objection

The rule-following objection concedes that thoughts and actions are realized as regularities evident in our bodily movements arising from impulses generated by the brain and its complex neural processes but denies that the relevant regularities can be captured by causal descriptions or explanations. The relevant patterns are not captured in physical terms because the contents of the thoughts and actions (his desire to remove his rival from the scene; his belief that wine stains shirts; the thought that a stained shirt is a reason to leave the party) depend on rules governing the way we classify things and therefore think about them. These rules exist in and through social practices rather than as physical entities and they do not causally compel the subject to follow them. If these rules are relevant to psychological explanation and they are not captured in physical terms and, what is more, they do not causally produce the behaviour they inform or structure, then it is clear that something other than physical causes is needed to explain human action.

Central to the objection is the thought that our classifications obey rules, and that such rules underpin our ways of dealing with the world. For instance, to classify something as a work of art is to draw attention to important features of it that are realized in but not described by its physical characteristics, so that the claim that I paid a certain amount of money for something because I believed that it was a work of art, would give such a non-physical feature an important role in explaining my behaviour.

In a standard causal explanation, or at least in that form of it espoused by Mill and required for causal determinism, the causal antecedents are themselves sufficient (given the background conditions) for the effect. For example, if there is fire then, given certain

other conditions, there will be smoke. Rule-following is not like this in that the existence of the rule plus the conditions in which it could be applied do not necessarily cause you to apply it in constructing your action. When following a rule you act in a certain way because applying the rule as you ought will advance your project; if there is a rule *then you ought to follow it* if it is relevant to what you are doing but it is not necessitated that *you do follow it*. The rule does not cause you to behave in accordance with it; rather you take the rule as a reason to act in a certain way and bring about the result you want. Therefore if the rule is an important part of the explanation of the behaviour that is exhibited, then there is a non-causal element in action. It follows that if I could vindicate the claim that explanations in terms of mental content were conceptually grounded in the idea of rule-following then mental content would rest on a non-causal feature of human activity. The argument is as follows.

- (1) Actions are explained by mental content.
- (2) Mental content is not specifiable apart from the rules governing it.
- (3) Rules do not cause mental content to be as it is and therefore do not causally produce behaviour.
- (4) The occurrence of rules in the specification of content makes explanation in terms of rules non-causal (although the abilities concerned may arise from the operation of causal mechanisms).

We now need to show that mental content involves rule-following and therefore that the explanation of action in terms of the mental content informing it (as I argued in the wine-spilling case) derivatively has the same dependence. Fortunately the relationship between mental content and rules is not hard to defend.

For any particular mental or psychological content (such as *there is a tree in front of me; that is red; I am playing Mozart*) there are norms governing what counts as an experience with the features justifying the relevant judgement. For instance, to be capable of believing <that frog is green> I must correctly apply the concepts <frog> and <green>. I must therefore have some proper grasp of what the concepts <frog> and <green> ought to be used for. Consider Xavier who says he believes *That frog is green* but, on

enquiring, we find that he has no idea what sort of thing a frog is or what green looks like. In that circumstance, it must be quite unclear to anybody, including Xavier, what is going on in his mind even if he says “That frog is green” as a result of brain processes going on in him. The true nature of what is going on turns on whether he is correctly using the rules for the concepts <frog> and <green> (by applying them to the right sort of things and connecting them appropriately in his thinking) which is not a matter of discovering causal connections among his brain processes.

At this point we can avail ourselves of Wittgenstein’s helpful analogy between mental activity (and the rules governing it) and chess. It is clear that the rules of chess govern my moves (as long as I am playing chess) and my knowing the rules equips me to play the game. But the rules of chess do not cause me to move in particular ways; I could always make other moves than those allowed and so just opt out of playing that game. If the rules of chess do constrain my moves it is because I elect to go on playing the game and to structure the moves I make in the way required to do so. Thus even though the rules have a formative effect on certain features of my behaviour during the game they do not causally compel me to follow them but rather tell me what I ought to do to participate in that project. But that implies that they affect my behaviour only subject to my acquiescence, which is not the way with any deterministic connection. In playing chess *it is up to me* whether I behave in a certain way even though the form of my behaviour is constrained by my wish to play chess. Thus, having got the telling phrase “up to me” into the picture we are veering close to an anti-determinist position.

To clinch the argument we need at this point to remind ourselves of the guiding conception (or identity structure) model of human action (Vallacher and Wegner, 1987) and to secure the analogical connection between chess and intentional action. The guiding conception model acknowledges the role of mental content in structuring action as follows:

An intention is a conception structuring and informing a tract of activity in the world.

I have argued that this should replace the model implicit in both Mill’s argument and the causal claim, viz., *an intention is a mental event preceding and causing an action*.

Vallacher and Wegner’s model replaces this with the idea that a guiding conception, say,

drawing an octagon, lends structure and coherence to a person's behaviour. On this model the intention is only mistakenly identified with an event immediately preceding the action because it guides the agent throughout the action and may need an adjustment in execution if something odd shows up. What is more, the guiding conception is made up of concepts (which are governed by rules). If we accept this model, then our specifying a person's action is inseparable from specifying the rules that person is following which conceptually structure the relevant behaviour. The analogy with chess is as follows: when I act I draw on certain rules to shape a particular response to a set of environmental conditions and if I make a mistake in applying those rules then I do not do what I intend. Think for instance of the attempt to hit a deep fly ball to right field in baseball. To do this I must conceive of what I want to do and then exercise the skills needed to hit the ball where I want it to go. If I do not do that correctly I will strike out or perhaps hit the ball to shortstop and be thrown out. What is more, just as a successfully executed move (in chess or baseball) may not bring about success in my overall game plan (for reasons I might not have foreseen) so an action may not serve my project of the moment in the way I hoped. Thus there are at least three different ways that a rule-governed performance can fail: (i) it can fail because of lack of skill (you fumble the piece and mess up the board); (ii) it can fail in conception ("That is not how to move a rook"; c.f. "Don't just try and hit it; hit a fly ball."); and (iii) it can fail as a strategem ("Uh, oh, moving your rook there has really opened you up to my queen"; "Uh, oh, your fly ball caused two outs not just one").

We can now sketch out the following argument about mental content and freedom of action.

1. Actions are structured and explained by mental content.
2. Mental content is based on rule-following.
3. Rules require to be followed; they do not causally produce a performance in accordance with them.
4. It is up to me whether I follow the rule or depart from it on any given occasion.
5. Rule-following seems to embed a condition which mentions me as a freely choosing agent.
6. Explanations in terms of mental content entail non-deterministic choices

explained by mental content.

7. Therefore strict mental or psychological determinism is not true.

I have argued that if you exhibit and use mental content in the production of a behaviour, as you must if that behaviour is to be structured by a conception and count as an action of yours, you are acting in accordance with certain rules. You act in accordance with the rule because you *elect* to play the relevant game or do something in a certain way (like hitting a fly ball to right field). This implies that it is up to you whether you act in the requisite way, you could always opt not to play that game or not to make exactly that move. I have also argued that the rule-performance link is metaphysically distinct from a cause-effect link in that it is not a mechanistic link because it is up to me which rules I follow in structuring my activity. Thus the rules imply a form of indifference to prior conditions (it depends what I choose to do) which is unlike the rest of the causal order as described by the natural sciences (it is not up to the billiard ball or the car which way it moves). The rule-performance link depends on the agent's electing to follow some rule in structuring her activity as the best way to serve her ends in the situation at hand.

At this point we need to stress that this indifference is central to specifying an action and we should not slip into the thought that a rule-follower is just a complex thing with multiple causal processes going on in it. A summated cause-effect link account of human behaviour considers all the causal antecedents of an action and regards them as sufficient to cause (deterministically) the action. By contrast a summated rule-performance link explanation of human action considers the concepts according to which the guiding conception is structured and, on that basis, the explanation acknowledges the role of the agent in using rules to control her behaviour. The agent chooses the conceptual rules to deploy in structuring her behaviour and uses those rules like tools to configure what she is doing. Even a complex machine does not, except by colourful metaphor, make its mind up and self-configure in this (revisable and reflective) way and therefore a summated (or complete or integrated) cause-effect link conception of brain function is a conception of a very different metaphysical stripe from a conception which takes in rules and concepts (or meanings) as important in the explanation of human behaviour. I have argued that the rule-performance model is the correct model for psychology because it is

the only way of specifying the meaning of a person's actions.

We might however, still have on our hands a determinist who wants to make the “last ditch” physicalist claim that it is all a matter of the brain and physical causal processes so that, whatever else you say, physical and causal explanation in terms of the biological and natural sciences must be able to provide a full and adequate explanation of human behaviour. This is a kind of *physicalist fundamentalism* and, quite apart from its dogmatism, there are other reasons to worry about it. We could see this if we examine an example in which psychological explanation requires the kind of richness available from an integrated rule-performance or meaning-giving picture of behaviour.

The Explanatory Circle

At this point we should note that cultural influences are of the rule-performance or meaning-involving type and not the cause-effect type. This claim could be argued for in the following way.

1. Cultural praxes involve the structuring of our actions according to things as they are signified or have meaning in my cultural group and its rule-following practices;
2. The things that I react to when I think and act may be physically identical to things encountered by a person of another culture but I may react to them in culturally specific ways;
3. Culture influences my thinking through my use of signs and symbols to construct ways of thinking about the world; these ways of thinking are formed during the process of socialization (Gillett, 1999a);
4. In the process of socialization I learn to adopt certain ways of classifying things according to the rules governing the use of signs and symbols for those things in my cultural group;
5. Nothing forces or causes our offspring to act in the ways that accord with our attempts at socialization (as is evident in teenage rebellion – a phenomenon medicalized as oppositional defiance disorder); they must play the game by our rules if they are to become like us and thereby learn our ways of adapting to the world (Gillett, 1993).

We can make these ideas vivid by looking at an example of an entrenched prejudice which affects a person's behaviour in profound ways. Imagine a person, Roger, who, in his 40s, has the following beliefs:

ULP: "*unemployed people are lazy and people are poor because they don't budget properly*"(Consedine, 1989, p. 174).

PW: *poor people should work and be more responsible.*

He therefore forms the following Monetarist intention:

MInt: *I will make poor people work.*

We might explain his intention as follows: Roger believes that poor people are lazy and irresponsible because he lived in New Zealand from when he was aged 7 in the year 1965 until 1998. At this time the socio-politico-cultural climate in New Zealand, formed as it was in the shadow of Monetarist ideology, influenced him in such a way that he came to his belief about the poor.

We now have the following adequate explanation for ULP and MInt.

E1. (The Aristotelian claim) Roger's beliefs are all configurations of his brain

From which

E2. Roger's holding PW corresponds, in some intuitive sense, to a state of his brain.

E3. Roger's belief that the poor are lazy and irresponsible is explained by the fact that he lived in New Zealand in the years 1965 to 1998 along with the facts about the socio-politico-cultural climate in New Zealand at that time.

E4. The socio-politico-cultural facts about New Zealand and their effect on Roger's rule-following practices or mental content explains his holding ULP.

E5. An adequate explanation of Roger's holding ULP incorporates facts that are not neurophysiological facts (this defeats the causal completeness claim for mental states).

E6. At least one feature of Roger's current brain configuration is explained by facts which are not physical facts. (This defeats causal completeness for brain states.)

E7. The configuration of Roger's brain is not fully explained by a purely physical description of the system of which it is part.

E8. Roger's attitudes and actions are essentially explained by his beliefs and other mental contents and cannot be fully explained by a causal description of the type espoused by deterministic physicalism no matter how compendious that is.

In order to establish that we need a rule-performance picture of what the brain is enabling us to do to get an adequate explanation in Roger's and other cases of beliefs, we can then return to a more general version of the Aristotelian claim (E1) and the reasons for believing that socio-cultural explanations are of the rule-following type. The general Aristotelian claim is that mental events are all bodily events in so far as they concern or are informed by our rational and social function. To this I have added the argument that a contentful mental act such as a belief or an intention involves rule-following and the rule-following is not like causal compulsion. Thus we derive the claim that mental acts are configurations of the brain when those configurations are taken to realise or be informed by rule-following practices and that the way they enter into psychological explanations reflects the meanings that we construct for ourselves.

We now have an explanatory circle and no principled reason to break it at any point. The fact that a certain person behaves in a certain way is explained by the contents of that person's beliefs, intentions, desires and other mental acts. These mental contents are determined by the rules that are in operation in the relevant mental act and rules are not instances of causal deterministic explanation. Therefore, determinist fundamentalism flies in the face of the need for adequate explanations of human behaviour. It can only be reinstated by a characteristic fundamentalist move, which is to pound the table and assert the consequent; in this case to make an assumption that begs the question at issue by asserting that all real explanations are causal explanations.

I would therefore conclude that causal or physicalist determinism in human psychology is an unreasonable position and that a philosophically informed psychology ought to make room for modest indifference and genuine freedom of the will. Modest indifference of the preferred type concedes that there are many influences on a person to act in a certain way in a given situation but preserves the insight that it is up to her (as a

whole, socially situated, individual who has formed herself in the light of her culture but not in ways compelled by it) to determine how she will actually behave. This is not strong indifference as espoused by the existentialists nor is it compatible with the view that human behaviour is adequately explained by material causes in the way Mill suggested.

Loose Ends and Loose cannons: Strong Indifference

Some philosophers who believe in strong indifference (such as Nietzsche and Sartre) would say that there are no antecedent events at all which have a causal influence on action and that all deterministic or causal explanation is merely a retreat from the radical non-givenness of human will and existence. While this assertion is consistent with the antideterminist strand of my argument, it is not consistent with my overall approach which recognizes the importance of psychological history and social factors in explaining human behaviour. The arguments in favour of modest indifference in fact allow us to embrace a realistic understanding of the relationship between subjectivity and discourse or social context in which we are (consenting) products of a process of acculturation or socialization.

It is implicit in E3 and E4 (the premises linking individual belief to social practices) that mental acts such as beliefs are explained by the facts about the socio-politico-cultural context of the believer. To deny that this is true would weaken our explanatory power in relation to mental acts: the phenomenology of our lived experience as agents is that our attitudes and beliefs are sensitive to the discursive contexts in which we find ourselves. Therefore a modest indifference is an important part of our understanding of people as subjects and agents although the mode of explanation they make possible is non-causal and therefore non-deterministic.

Conclusion

The present arguments render untenable any strong physical determinism about mental acts. Libet's experiments, which seem to show that mental events are caused by physical antecedents, are deeply flawed and show nothing of the kind. McGuigan's position demonstrates nothing more than that our bodies are the medium whereby rules

and instruction configure our behaviour. Our behaviour is best thought of as comprising intentional actions in so far as what we do is subject to identity structures which dictate its form and content. The identity structures are assembled from the concepts we have mastered and these are mastered as we learn to follow the rules for classifying things in culturally significant ways. We are left with the conclusion that, in a way unique to mental or psychological explanation, the socio-politico-cultural context of human beings helps us to understand the thoughts, attitudes, and feelings of others but that this understanding lacks the conceptual features of deterministic causal explanation and ought to be reconceptualised as relating individuals to their life contexts in quite another way.

References

Brunia, C.H.M., and Haagh, S.A.U.M. (1986). Preparation for action: slow potentials and EMG. In H. Heuer and C. Fromm (Eds.), *Generation and modulation of action pattern*. (pp. 238-287). Berlin: Springer Verlag.

Consedine, B. (1989). Inequality and the egalitarian myth. In D. Novitz and B. Willmott (Eds.), *Culture and identity in New Zealand* (pp. 172-186) Wellington, New Zealand: GP Books.

Davidson, D. (1980). *Essays on actions and events*. Oxford: Clarendon.

Dennett, D.C. (1991). *Consciousness explained*. London: Penguin Press.

Feigl, H. (1970). Mind-body, *not* a pseudo-problem. In C.V.Borst, C.V. (Ed.), *The mind/brain identity theory* (pp 33-41). London: Macmillan.

Gillett, G. (1990). Consciousness, the brain and what matters. *Bioethics*, 4, 181-198.

Gillett, G. (1992a). *Representation, meaning and thought*. Oxford: Clarendon.

Gillett, G., (1992b). Actions, causes and mental ascriptions. In H. Robinson (Ed.), *Objections to physicalism* (pp 81-100). Oxford: Clarendon.

Gillett, G. (1993). Social causation and cognitive neuroscience. *Journal for the Theory of Social Behaviour*, 23, 27-45.

Gillett, G. (1999a). Dennett, Foucault, and the selection of memes. *Inquiry*, 42, 3-23.

Gillett, G. (1999b). Consciousness and lesser states: The evolutionary foothills of the mind.

Philosophy, 74, 331-360.

Kenny, A. (1973). Freedom, spontaneity and indifference. In T.Honderich (Ed.), *Essays on freedom of action* (pp. 87-104). London: Routledge.

Kim, J. (1993). Explanatory realism, causal realism, and explanatory exclusion. In D.Hillel Ruben (Ed.), *Explanation* (pp. 228-245). Oxford: Oxford University Press.

Laszlo, J. (1966). The performance of a simple motor task with kinaesthetic sense loss. *Quarterly Journal of Experimental Psychology*, 18, 1-8.

Libet B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioural and Brain Sciences*, 8, 529-566.

McGuigan, F.J. (1966). *Thinking: Studies of covert language processes*. New York: Appleton-Century-Crofts.

McGuigan F.J. (1978). *Cognitive psychophysiology: principles of covert behaviour*. Englewood Cliffs, New Jersey: Prentice Hall.

McGuigan F.J. (1994). *Biological psychology: A cybernetic science*. Englewood Cliffs, New Jersey: Prentice Hall.

McGuigan, F.J. (1997). A neuromuscular model of mind with clinical and educational applications. *Journal of Mind and Behaviour*, 18, 351-370.

Mill, J.S. (1974). *A system of Logic: Rationcinative and inductive* (Volume 1 & 2), Toronto: University of Toronto Press. (Originally published 1843)

Millikan, R.G. (1993). *White queen psychology and other essays for Alice*. Cambridge, Massachusetts: MIT Press.

Spence, S. (1996). Free will in the light of neuropsychiatry, *Philosophy, Psychiatry, and Psychology*, 3, 75-90.

Vallacher, R.R., and Wegner, D.M. (1987). What do people think they're doing? Action identification and human behaviour *Psychological Review*, 94, 3-15.

Vanderwolf, C.H. (1985). Nineteenth century psychology and twentieth-century electrophysiology do not mix. *The Behavioural and Brain Sciences*, 8, 555.

Wittgenstein, L. (1966). *Wittgenstein: Lectures and conversations* [C.Barrett, Ed.]. Oxford: Blackwell.

Young, J.Z. (1986). *Philosophy and the brain*. Oxford: Oxford University Press.

ⁱ This is, I believe, a successful defence of the causal brain mechanism against the unpredictability objection.