# Could Anyone Justiably Believe Epiphenomenalism?

Richard Swinburne

[Swinburne, Richard, 2011, "Could Anyone Justiably Believe Epiphenomenalism?", *Journal of Consciousness Studies*, vol 18, no 3-4, 2011, pp.196-216.]

*Abstract:* Epiphenomenalism claims that all conscious events are caused immediately by brain events, and no conscious events cause brain events. In order to have a justified belief in a theory someone needs a justified belief that it or some higher-level theory predicts certain events and those events occurred. To have either of the latter beliefs we depend ultimately on the evidence of apparent experience, memory, and testimony, which is credible in the absence of defeaters; it is an undermining defeater to a belief produced by apparent memory that it was not caused by a past belief, and to a belief produced by apparent testimony that it was not caused by an intention to say what the speaker believes. A justified belief in epiphenomenalism requires either evidence about when conscious events occurred or evidence about what some theory that brain events are caused solely by physical events predicts, but epiphenomenalism rules out the availability of the evidence of apparent memory and testimony on these matters. Hence only a rare individual scientist who could hold in her mind at one time the proof that a theory makes certain predictions could have a justified belief that epiphenomenalism is true. It follows that recent neurophysiological work in the tradition of Libet has no tendency whatever to provide a justified belief in epiphenomenalism.

## I Introduction

I understand by 'epiphenomenalism' the theory that [X] all conscious events have physical events (viz. brain events) as their only immediate causes, and [Y] no conscious events cause physical events.[1]. By one event causing another event 'immediately' I mean the first event causing the second event without doing so by causing some third event which in turn caused the second event. As a theory about what kinds of events cause other kinds of events, epiphenomenalism must be regarded as a scientific theory. Hence this theory cannot be justifiably believed by someone without their justifiably believing that it makes true predictions (checkable without presupposing the truth of that theory) not predicted by serious rival theories, or is a con-

sequence of some higher-level theory which makes true predictions not made by serious rival theories[2]. This will hold on any mildly plausible internalist or externalist account of justification. The internalist will regard successful prediction as an a priori requirement for justification, while the externalist will hold that the scientific method requiring successful prediction is a reliably truth-conducive method (or satisfies some other externalist requirement) and that that is what makes a belief acquired by that method justified[3]. I argue in this paper that (with a very small exception) no one could ever have this justification for epiphenomenalism[4].At the end of the paper I apply my general results to show that experiments of the type pioneered by Benjamin Libet could not yield that justified belief in epiphenomenalism.

## II Ontology

Epiphenomenalism assumes event dualism; that there are both physical events, including brain events, and mental events. There are different ways for a dualist to make the distinction between mental and physical events. Since all who accept the distinction think of sensations as mental events, definitions of the mental in terms of the intentional are clearly inadequate, and so I suggest that the distinction is best made in terms of the private versus the public. I define a mental event as an event of a kind to which the person involved in the event has privileged access, that is a way of access to whether an event of this kind is or is not occurring which is necessarily not available to others, by actually experiencing it or its absence. I define a physical event as an event to which no one person has privileged access, that is a public event. I define a pure mental event as one which does not entail the occurrence of a physical event. Perceptions such as my seeing a tree are mental events since I can know better than anyone else whether or not I am seeing a tree, but they are not pure mental events since seeing a tree entails that there is a tree present – and that is a physical event. But sensations such as pains and beliefs such as the belief (of which I am currently aware) that I am seeing a tree are pure mental events, since it is not entailed by the occurrence of sensations or beliefs that anything public is happening.[5]

Conscious events are a sub-class of pure mental events. They include both those pure mental events which exist only while the subject is conscious of them, and also pure mental events of which the subject is conscious but which may exist while the subject is not conscious of them. The first group includes not merely sensations such as pains, but also occurrent thoughts. If I am not in any way aware that the thought ‘today is Thursday’ is now crossing my mind, it isn’t crossing my mind. Among the pure mental events of which I can be-

come conscious but which may continue to exist while I am not conscious of them are beliefs, desires, and intentions (in the sense of intentions or purposes currently directing my actions). I have innumerable beliefs of which I am not currently thinking, and may be fulfilling some intention (e.g. to walk to the railway station) while thinking about quite other matters. Yet I may become conscious of my beliefs (e.g. when I investigate them), desires, and intentions (especially when I form a new intention, that is make a decision).

Many philosophers have held that talk about beliefs and intentions is analysable in terms of talk about public actions or the brain events which cause them. But that cannot be correct, since intentions only lead to public actions when combined with beliefs. Different combinations of beliefs and intentions may lead to the same public actions. You have a headache, and ask for an aspirin, I give you a pill; it is a poison pill and you are poisoned and die. My action may be the result of a belief that the pill was an aspirin and my intention to cure your headache; or the result of a belief that the pill was poisonous and my intention to kill you. Of course we can often make good inductive inferences from someone's past behaviour and public life experiences to the intentions and beliefs which it is most probable that he will have had; but that is compatible with him suddenly forming a new intention (e.g. to kill someone) or forgetting a belief (that the pill was poisonous), and clearly the agent has a way, while he is doing the action, of knowing about what his intention is and so whether he has some requisite belief, not available to others. As a theory about conscious events, epiphenomenalism must be regarded as concerned with all such events insofar as the subject is conscious of them.

## III Epistemology

How can anyone have a justified belief that a scientific theory predicts certain events?

Scientists in the centre of the field will have calculated that it makes these predictions. And if a scientist can hold all the calculations in her mind at one time, it will be for her a deliverance of reason, evident a priori, that the theory does make these predictions. Alas, for any scientific theory of any complexity most experts at the centre of the field will be unable to hold in their minds at one time all the relevant calculations; even as the scientist reads through the text of her calculations, she depends on her memory towards the end of the calculations for her belief that the initial calculations were correct. Later in life all that she may remember is that it did seem to her earlier that the theory made those predictions. She may have a diary in which she recorded this, which will be – as it were – her testimony

about this to herself and others. Non-scientists and scientists less central in the field will depend on the testimony of those whom they regard as experts, that they have made those calculations. So what makes a belief that the theory predicts certain events justified is experience (of 'seeing' that the calculations are current), memory, and testimony; or rather, since all of these sources may be mislead, it is apparent experience, memory, and testimony. These three sources provide our evidence that the theory makes true predictions. True, we do not normally speak of someone's apparent experience of the truth of calculations, or their apparent memory or testimony that they made certain calculations which seemed to them correct showing that the theory predicts certain events, as part of the 'evidence' for the correctness of the theory. But we do in fact rely on these sources for the justification of a belief that a theory predicts certain events, and –as we shall see shortly – any evidence undermining the belief that the apparent experience, memory, or testimony was veridical will undermine the justification of the theory[6]. So in a wide sense of 'evidence', everyone requires apparent experience, memory, or testimony of this kind as part of the 'evidence' which will give them a justified belief that the theory is true.

And how can anyone have a justified belief that the events predicted in fact occurred? Someone will normally depend on the evidence of the same three sources. Certain observers will (apparently) in a wide sense experience these events – that is if they are physical events, they will perceive them, or if they are conscious events they will experience them. (I will use the wide sense of 'experience' to include perception in future.) Later, the observers may (apparently) remember having experienced the events; and others will depend on the (apparent) testimony of observers about these (or the observers may depend on their own apparent written testimony.) Alternatively, a believer may have a justified belief that the events predicted occurred because it is a consequence (deductive or probabilistic) of some other justifiably believed theory that they did. But in that case a justified belief in that other theory would itself depend on the evidence of the same three sources.

It is a fundamental epistemic principle that what we seem to (that is, apparently) experience is probably so – barring counter-evidence; this includes what we seem to observe in the public world, what we seem to experience as conscious events, and the logical consequences we seem to 'see'. This has been called the principle of credulity. If this were not a fundamental epistemic principle, total scepticism would follow. It is a second fundamental epistemic principle that what we seem to (that is, apparently) remember having experienced, we probably did experience - barring counter-evidence. And it is a third fundamental epistemic principle that what people seem to be (that is, apparently are) telling us that they experienced,

they probably did experience, again barring counter-evidence.[7] Beliefs acquired by apparent experience, memory, and testimony are probably true – in the absence of counter-evidence. Science relies on the applicability of these principles to determine what constitutes evidence. A scientist takes his (apparent) observations, experiences and calculations as probably correct, at least when he has looked carefully and checked. Almost all scientific knowledge relies on (apparent) memory (e.g. of the results of experiments or calculations only written up the following day). And for all science, we all rely most of the time on the (apparent) testimony (written and spoken) of observers to have had certain experiences (normally in the form of observations) and of theoreticians to have done certain calculations. And the wider public relies entirely on the (apparent) testimony of scientists with respect both to their calculations and to their experiences.

Beliefs acquired by apparent experience, memory, and testimony are however open to counter-evidence or defeaters. There are two kinds of defeaters– undermining and overriding defeaters. If we have inferred the occurrence of some event *y* from present evidence *x*, then an undermining defeater is evidence (making it probable) that *x* did not occur or is not good evidence for *y*, whereas an overriding defeater is new evidence that *y* did not happen. If for example I apparently experience hearing my telephone ring, and then someone points out to me that the noise is coming from the television set where someone is depicted as hearing a telephone ring, that constitutes an undermining defeater for my apparent experience. It doesn't show that my telephone was not ringing, but it does show that the noise was not evidence that it was, because the noise had a different cause. Again, if I have come to believe that *y* happened because some person apparently testified that he saw *y*, evidence that that person was somewhere else at the relevant time and so could not have seen *y* undermines his evidence, and I no longer have reason to believe that *y* happened. By contrast the apparent testimony of two independent witnesses that they were at the place of the alleged occurrence of *y*, and that they saw that *y* did not happen, overrides the evidence of the original witness. But the evidence constituting the defeater must itself be provided by apparent experience, memory, or testimony. This evidence need not be direct evidence of, for example, the non-occurrence of the event or of the evidence for it – for example in the form of apparent testimony that the testifier was not present at the site of the alleged event; it may be indirect evidence, in the sense that it may be evidence supporting a theory which has the consequence that the event or the evidence for it apparently experienced, remembered or testified to couldn't have happened - for example, evidence supporting a theory that the testifier was blind and so couldn't have seen what he testified to having seen.

Scientists are used to their claims about what theories predict and what has been observed being defeated from time to time. Calculations are sometimes shown to have been inaccurate, when others recalculate them many times. Evidence that when the experiment is done again and again, the original result cannot be replicated is evidence that it was originally misobserved or misreported. But, to repeat the point, the evidence constituting the defeater must itself ultimately come from apparent experience, memory, or testimony.

That someone justifiably believes that a theory predicts certain events and that these events were (in the wide sense) experienced, only if they have apparent experience, memory, or testimony (which is not defeated) that these things are so (or it follows from some other theory justified by these three sources) is hard to deny[8]. This claim is compatible with any mildly plausible account, either internalist or externalist, of what gives apparent experience, memory, and testimony this status. Likewise it is compatible with any view at all about what else (e.g being a simple theory) other than its ability to predict events makes a theory probably true.

Further, I claim, in having beliefs resulting from experience of physical events such as the apparent observation of a desk, we assume that the event (of the presence of the desk) experienced caused the belief (with its accompanying sensations), 'caused' that is in being a necessary part of the total cause. In perception we seem in contact with the event apparently observed. That event seems to force itself upon us; the presence of the desk seems to force itself upon me, and so I have no option but to believe that it is there. That, we assume, is because there is a causal chain from the desk to the belief – only causes exert 'force'. (This holds, we believe these days, because the desk reflects light rays which land on my eyes and cause the belief.) Hence the generally accepted causal theory of perception. (Maybe not any perceptual belief caused by the object apparently observed constitutes an observation of it. Maybe the causal route must not be 'deviant'. But that does not affect my point that a causal route is necessary for perception.) It is natural to suppose that the same goes for our beliefs about our conscious events; that in believing that we are having certain sensations we assume that the belief is forced upon us by those events, and in believing that our calculations are correct we assume that that belief is forced upon us by the calculations – the marks on the paper or in our mind symbolizing the calculations cause us to have the belief that the calculations are correct.

Some writers have denied that our beliefs about currently experienced conscious events are caused by those events; they claim that in this special case we have direct access of a non-causal kind to our conscious events[9]. This suggests a qualification on [X] so as to read, instead of 'all conscious events', 'all conscious events except

conscious beliefs about current conscious events'. I shall assume this qualification, since if my arguments show (as I believe that they do) that noone could have a justified belief in epiphenomenalism even with this qualification, it will follow that they cannot have it without the qualification.

So with respect to beliefs resulting from experience (with the above mentioned exception) evidence that such a belief was not caused by a causal chain (of necessary parts of total causes) from the event believed constitutes an undermining defeater for it – as in my example of the telephone ring.

A similar assumption of the existence of causal chains, although longer ones than for experience and ones involving different kinds of event, undergirds our beliefs in the deliverances of apparent memory and testimony. I trust my apparent memory of an event because I assume that that apparent memory was caused by a past apparent experience of the event recalled, and that the experience was caused by the event itself. Thus in trusting my apparent memory that I was in London on Monday I assume that it was caused by my apparent experience on Monday of being in London, itself caused by my being in London. Hence the generally accepted causal theory of memory. (The apparent memory must of course correspond to the previous experience, and maybe the causal route must not be 'deviant'. My point is merely that a causal route is necessary for memory.) Any evidence that the (apparent) memory was planted in me by a hypnotist or a brain surgeon constitutes an undermining defeater for that apparent memory belief.

Similarly in believing someone's apparent testimony to be experiencing or have experienced some event I assume that they say what they do because they are apparently experiencing or apparently remember having experienced that event and have the intention of telling me the truth about it; that is, their apparent experience or memory and their intention causes them to say what they do, 'causes' in the sense of being a necessary part of the total cause. In the case of a past event I believe that their apparent memory was caused by an apparent past experience of the event, the latter being caused by the event itself. So if I get evidence that the words coming out of some person's mouth were not caused by any intention of his (e.g. that the words were caused by a neurophysiologist stimulating that person's neurones to cause his mouth to make the sounds, or simply as in fluent aphasia where a neural malfunction causes a stream of words to come out of a subject's mouth) that evidence constitutes an undermining defeater to belief in the truth of what that person seemed to be saying. (The intention does of course have to be of a particular kind, an intention to tell the truth; and evidence that the person was intending to deceive me would also undermine his testimony.

But my point is simply that evidence that there is no causation at all by the apparent testifier's intention undermines his apparent testimony.) In all of these cases the counter-evidence (in the form of an undermining defeater) must itself come (directly or indirectly) from apparent experience, memory, or testimony[10].

In summary then I am making the epistemic assumption (EA) that:

(1) A justified belief in a scientific theory (which is not itself a consequence of any higher-level theory in which the believer has a justified belief) requires a justified belief that the theory makes true predictions.

(2) A justified belief that a theory makes true predictions is (unless this is a consequence of some other theory in which the believer has a justified belief) provided by and only by the evidence of apparent experience, memory, and testimony that the theory predicts certain events and that these events occurred.

(3) Such justification is undermined by evidence that any apparent experience was not caused by the event apparently experienced, any apparent memory was not caused by an apparent experience of the event apparently remembered, and any apparent testimony was not caused by the testifier's intention to report his apparent experience or memory.

I hope that the few examples by which I have illustrated its application show the centrality of (EA) in our noetic framework. The fundamental criterion (FC) behind (EA) is that justified belief that some event occurred requires the assumption that that event is (privilegedly) accessible to or causes effects (privilegedly) accessible to the believer (unless it is justifiably believed to be the consequence of some theory which predicts events justifiably believed to occur on grounds independent of that theory). Then justified belief that a theory makes true predictions requires (unless justified by a higher-level theory) the assumption that both a scientist's awareness of the calculations that the theory predicts certain events and the events predicted are accessible or cause effects accessible to the believer. (FC), I suggest, is a criterion central to our judgments about the credibility of a scientific theory.

## IV Consequences for Epiphenomenalism

(Unless her belief depends for its justification on some higher-level theory of which epiphenomenalism is a consequence) a believer could be justified in believing epiphenomenalism only in virtue of having evidence of when (relative to brain events) various conscious events occur, which I shall call (α) type evidence. For both conjuncts

of epiphenomenalism ([X] and [Y]) make predictions about and only about such evidence.

The second conjunct [Y], that no conscious event causes any physical event, makes (α) type predictions to the effect that the occurrence of any conscious event makes no difference to the pattern of later brain events. It predicts that whether or not some type of conscious event occurs during the first part of some sequence of brain events will make no difference to whether or not the sequence is completed (and so cause public behaviour). If this prediction were tested for a large random sample of different types of sequences of brain events (especially those ending with events which are supposed to be caused by intentions) and different types of conscious events (especially intentions), and found to be correct, this would be strong evidence for [Y] To test such predictions, a scientist would have to learn about the times of occurrence of various conscious events. The paradigm way to learn about this is from apparent experience, memory, and testimony about when conscious events occurred. But a scientist could learn about this from the predictions of some other theory, and in that case that theory would be a theory about when conscious events occurred and could be justifiably believed only on evidence of the same kind.

Yet if apparent testimony is to constitute evidence that conscious events occurred, the scientist must -by (EA) - assume that the subjects are caused to say what they do by a belief that the conscious events occurred and an intention to tell the truth about their belief – a causal route which must go through a brain event. But if [Y] were true, no conscious events will cause any brain event to cause the subjects to say what they do. Yet no theory could be justifiably believed on the basis of evidence about the occurrence of events about the occurrence of which we could have evidence only if we assume [Y] to be false. Hence epiphenomenalism (of which [Y] is a conjunct) couldn't be justifiably believed on the basis of apparent testimony. A scientist might remember his own conscious events. (By EA) someone is justified in trusting his apparent memories on the assumption that they are caused by his past experiences. But if [X] is true, all apparent memories of past events are caused immediately by and only by brain events. So – if someone believes [X] - in order to be justified in trusting his apparent memories he must assume that the brain events which are their immediate causes are themselves caused (whether immediately or via a chain of intermediate brain events) by his past experiences. So if someone believes [X], trusting his apparent memories would also involve assuming conscious event- brain event causation, and so not believing [Y]. Hence if a scientist is to be justified in trusting his apparent memories, he must assume that either [X] or [Y] and so epiphenomenalism as a whole, is false. Hence apparent memories of past experiences cannot

provide a justified belief that epiphenomenalism makes true predictions, any more than can the apparent testimony of others. The qualified version of [X] would still permit a scientist to have a justified belief about which conscious event he was currently experiencing. But the evidence of one private event currently experienced by a scientist would hardly constitute enough evidence of successful predictions to make it (together with any amount of evidence about brain events) at all probable that [Y] is true, and so give even that scientist a justified belief in epiphenomenalism. So noone could have a justified belief in epiphenomenalism on the basis of type () evidence for it, because epiphenomenalism rules out the availability of enough evidence of that kind[11]. And even if someone doesn't believe [X] in its full generality, it is implausible to believe that apparent memories of events more than a few seconds previously are caused by past experiences without the route of causation going through the brain; and so the above arguments still show that noone can have enough evidence of type (α) to provide a justified belief in [Y].

It might however seem that someone could have a justified belief in epiphenomenalism because of a justified belief in some general theory about the world of which epiphenomenalism was a consequence. The obvious example of such a theory is a theory that the physical realm is causally closed; but all that is needed is a narrower deterministic theory that every brain event has as an immediate necessary and sufficient causal condition some other brain (or other physical) event. That theory would entail [Y]; for if every brain event has another brain (or other physical) event as its immediate necessary and sufficient causal condition, no brain event can have a conscious event as its necessary causal condition; overdetermination would be excluded. It might be thought that we could establish that deterministic physical theory on evidence solely about which brain events occur when (relative to other brain or other physical events), which I will call evidence of type (β). If we found that for any random sample of brain events (and especially ones supposed to be caused by conscious events), that each of them is related to some brain (or other physical) event as its immediate necessary and sufficient cause in a way calculable from such a theory, that would seem to be powerful evidence in favour of [Y].

Someone could justifiably believe certain brain events to be occurring on the evidence of apparent experience (a current observation). But to get enough evidence to acquire a justified belief that the deterministic physical theory is true, a scientist would require evidence provided by apparent memory of past observations and apparent testimony by others to having observed various brain events in the past. But a justified belief in the deliverances of apparent memory of past experiences and apparent testimony to them is – (by EA) - undermined by evidence that they are not caused by experiences of

those events. So – given (EA)- there could not be a justified belief in a physical theory which entailed [Y].

However a modified understanding of memory and testimony is possible, which keeps apparent memory and testimony as sources of justified belief, and is still compatible with the fundamental criterion (FC) (lying behind (EA)) that (barring justification by a justified theory) justified belief in the occurrence of an event is dependent on the assumption that that event is accessible to or causes an effect accessible to the believer. One could understand memory simply as memory of the occurrence of events, and not only of events which are experiences of the occurrence of events. A subject could be said to 'remember' past physical events in virtue of those events causing traces in his brain, which at a later time cause the apparent memory of those events without any mental-to-physical causation being involved. People sometimes become aware later of details of some event which they observed and of which they were not at the time aware; and it does not seem too unnatural use of the word 'remember' to say that they 'remembered' those details. And we could come to understand testimony to amount merely to the public utterance of sentences reporting that an event occurred caused by a chain of events in the utterer, itself caused by the event reported, a chain which need not include any conscious events. The 'testimony' would not be testimony that the testifier had observed the events, but merely testimony that the events had occurred. This certainly seems to involve giving a stretched meaning to 'testimony', but relying on apparent testimony of this kind to the occurrence of physical events would still be compatible with the fundamental criterion (FC). Given these modified senses of memory and testimony, someone could have an apparent memory of or receive apparent testimony to the occurrence of brain events without making any assumption about anyone's conscious events causing brain events. Thus someone's eyes could receive light rays from events in the brains of others and – because those brain events caused brain events in that person – subsequently report them, without that causal chain proceeding through any conscious events. Given this modified understanding of apparent testimony and memory, anyone could have justified beliefs in the occurrence of any set of brain events which occurred without presupposing causation of the physical by the mental; and so come to believe in the occurrence of the brain events ((β) type evidence) predicted by a deterministic physical theory. This modified understanding of memory and testimony would not however make any difference to the unavailability of apparent memory and testimony to provide justified beliefs about the occurrence of conscious events, and so could not provide type (α) evidence in favour of epiphenomenalism. For even on the modified understanding of these notions apparent memory and testimony about conscious events is undermined

by evidence that conscious events do not cause brain events.

However, if someone believed [Y] on the basis of (β) type evidence in favour of a deterministic physical theory, they could not in that case have a justified belief in [X]. For that would require justified beliefs about which conscious events occurred after which brain events, and – given [Y] - apparent memory or testimony could not provide this. It would be open to the believer to suppose that conscious events were often uncaused or caused each other. Nevertheless it is reasonable to suppose that [Y] is the core claim of epiphenomenalism; and it might seem that (given the modified understanding of 'memory and 'testimony') someone could have a justified belief in [Y] in virtue of having a justified belief in a deterministic physical theory of which it was a consequence.

There is however a further problem in supposing that we could have a justified belief that some deterministic physical theory gave true predictions about relations between brain (or other physical) events. This is that we would also need a justified belief, not merely that certain relations between brain events occurred, but also a justified belief that these relations were predicted by that deterministic theory. But anyone who had not calculated for himself what that theory predicted about the relations between brain events must depend on the evidence provided by the apparent testimony of scientists to have calculated this and 'seen' (that is, had a conscious belief) that that was what the theory predicted, that is evidence of the conscious events of scientists. But if the deterministic physical theory were true, the scientist would not have been caused to give that testimony by any conscious event – neither by his intention to tell the truth or even merely by his conscious belief about what the theory predicted. Hence noone could justifiably believe what the scientist reported about his calculations, and so believe that the theory made the predictions which he claimed that it did (as well as believing that the predicted events occurred), since believing what the scientist reported would undermine the credibility of his apparent testimony to it. Scientists normally check each other's calculations, but for the same reason – if the deterministic physical theory were true - no scientist could rely on the testimony of another scientist to have made the same calculation as he had. Neither could any scientist rely on his own testimony to himself recorded in a diary that he had previously calculated the consequences of the deterministic theory. And (although he might see no need to assume [X] in all its fullness) unless – implausibly - a scientist is prepared to suppose that his past experiences cause his apparent memories more than a few seconds later without the causal chain proceeding through brain events, could he be justified in relying on the evidence of his own apparent memory about his calculations. Only if a scientist could hold in his mind at one time all his calculations from which it apparently followed that

the deterministic theory predicted certain events, could he have a justified belief that that theory made successful predictions, and so a justified belief in [Y] and so in epiphenomenalism. For most scientific theories and most scientists, this is most unlikely.

I conclude that, given the fundamental criterion (FC) which guides the acceptability of scientific theories, (with the above very small exception) no one could have a justified belief that any deterministic physical theory of which epiphenomenalism was a consequence made certain predictions, and so noone could have a justified belief in [Y]. Hence more generally (with this very small exception) noone can have a justified belief in epiphenomenalism.

## V Recent neurophysiology

The research programme initiated by Benjamin Libet seeks to provide evidence of type (α) showing (i.e. providing a justified belief) that a sample of brain events of one kind which cause intentional actions (i.e. actions which the agent had or believed that he had the intention to perform) are not caused by intentions[12]. In this final section I consider whether it does show that; and I apply my earlier arguments to show that even if it did show that, this result could not be extrapolated to establish overall epiphenomenalism..

In the original and most influential Libet experiments[13] participants were instructed to move their hand at a moment of their choice within a period (e.g. 20 seconds). They watch a very fast clock, and report subsequently the moment at which they first had the 'intention' (or whatever – see below) to move the hand. They reported the 'intention' to move the hand as (on average) occurring 200 msecs before the onset of muscle activity initiating a hand movement. However electrodes placed on their scalp recorded (on each occasion of hand moving) a build up of 'readiness potential' (RP), which was evidence of a particular kind of brain event (which I'll call $B_1$) occurring an average 550 msecs before the muscle activity. Experiments of other kinds, Libet claimed, showed that subjects report the time of sensations as occurring 50 msecs before the time of brain events which caused them[14.] That led Libet to hold that subjects misjudge the time of all conscious events by 50 msecs, and so he concluded that the 'intention' first appeared 150 msecs before the muscle activation. So, many have argued, this showed that $B_1$ caused the hand movement, and that the 'intention' was a mere epiphenomenon.

One problem with Libet's experiments is that Libet and other experimenters describe the conscious event which the subjects report and which I have just described as an 'intention', sometimes instead

as the onset of a 'wish', an 'urge' a 'wanting' or a 'decision'[15]. There is clearly a big confusion here between passive inclinations which are designated synonymously by 'wishes', 'wantings', 'desires' or 'urges', apparent active 'intentions' (of which the subject is not necessarily conscious all the time they operate), and conscious formations of intentions called 'decisions'. Since however these results were obtained from subjects who were instructed not to pre-plan their movements and to act spontaneously, then (on the reasonable assumption that subjects followed the instructions) any decision (i.e. conscious formation of an intention) and any intention must have been a decision or intention for immediate implementation. And since intentions (or at any rate ones concerning basic actions of this present kind) are the sort of things of which we can become conscious at any moment if we so choose and since subjects were asked to look out for their intentions, the first awareness of an intention must have coincided with a decision to act. So any confusion in subjects' reports would be about whether they were reporting the time at which they first became aware of a desire to act, or the time at which they first become aware of an intention to act (i.e. a decision). This is a serious confusion, but since the desire must precede the decision (desires lead to the formation of intentions, not vice versa), it does not affect the main result that (if subjects' reports of timing of either event are anywhere near accurate) the decision to act follows $B_1$. So – to simplify discussion – I will suppose the conscious event dated by the subjects (in these experiments) to be a decision, which would be the same as the beginning of an intention. But since this intention needed only to last a very short time to produce its effect, I will use the word 'intention' to describe this event.

So, if the subjects' reports are at all accurate there is a succession of events: a brain event ($B_1$), then a (conscious) mental event (the intention which I'll call $M_2$), and then some brain event (which I'll call $B_3$) which directly causes the muscle activity and so the movement. Many neurophysiologists proceed from that to reach the extraordinary conclusion that the intention does not cause the movement. Thus Roediger, Goode, and Zaromb ([2008] p.208) conclude that Libet's data 'contradict the naïve view of free will – that conscious intention causes action. Clearly conscious intention cannot cause an action if a neural event that precedes and correlates with the action comes before the conscious intention'.[16] But that is a totally unjustified conclusion, since it is equally compatible with all the data and the most natural explanation of them to suppose that $B_1$ causes (in the sense of being a necessary causal condition for) the 'conscious intention'($M_2$), and that the intention causes the brain event ($B_3$) which directly causes the movement. Causation is transitive. If I flip the light switch and thereby cause the light bulb to light-up, that doesn't rule out the possibility that my flipping the switch

caused an electric current to pass to the bulb and that the current caused the bulb to light-up. Despite this obvious point many neurophysiologists prefer one of two rival explanations of the data over the natural explanation. One is that an earlier brain event ($B_1$) causes both the intention ($M_2$) and (in 'parallel') a sequence of brain events leading to $B_3$ which causes the hand movement without the intention causing any brain event; and the other is that the intention never occurred, but that after the hand movement the subject came to believe falsely that it occurred[17].

Even if it were shown that $B_1$ causes a sequence of brain events which are necessary for the bodily movement when that constitutes an intentional action (in virtue of the agent having or believing that he had the intention to make that movement), that wouldn't show that the intention was not also a necessary part of the cause. To show that you would need to show that $B_1$ causes the very same sequence of brain events with or without subjects having the requisite intention (to produce that bodily movement) and so with or without the bodily movement constituting an intentional action. This could be shown only if a scientist could prevent the occurrence of the intention, without thereby automatically preventing the occurrence of the sequence caused by $B_1$. This experiment could only be performed if the intention is not caused by any brain event (and then [X] would be false), or is caused by a brain event (that is, has a brain event as a necessary causal condition) which was not itself part of the sequence from $B_1$ to $B_3$. For only if the intention is not caused by an event which belongs to that sequence (although perhaps by a brain event itself caused by an event in the sequence), could the intention be prevented without preventing the occurrence of the sequence of events which caused the intention. If nevertheless the intention was prevented, and $B_1$ still caused exactly the same sequence of events culminating in $B_3$ and the hand movement, that would show that the intention was not a necessary part of the cause of the hand movement. It would show that the sequence of brain events alone causes the very same movement as would constitute an intentional action if the agent had (or believed that he had) an intention to cause it. As far as I know, no one has attempted to show this. If this were shown, we would have evidence against the natural interpretation of the Libet experiments, that a brain event causes the intention which causes the brain event which causes the bodily movement. This interpretation, I repeat, is to be read in the minimal sense as the hypothesis that $B_1$ is a necessary causal condition of $M_2$ which is a necessary causal condition of $B_3$ which causes the movement; it is compatible with $B_1$ and a sequence of brain events caused by it also being a necessary condition for the occurrence of $B_3$.

Experimenters seeking to establish a scientific theory, such as those performing Libet-type experiments, assume that they have ac-

cess to the conscious lives of many different subjects (and so evidence of type α about them), in order to test the prediction discussed in the last paragraph, that the same sequence of brain events would occur in the absence of the intention, without which the experimental results do not show that the intention does not cause the movement. (We can now ignore the situation of the solitary scientist relying on his own apparent memories, the consequences of which have already been adequately explored.) The only way for experimenters to acquire this information about the conscious events of subjects is from what those subjects tell them (or by a higher-level theory itself justified by what subjects say). So experimenters assume that subjects' beliefs about their conscious events (including their memory beliefs) are correlated with their testimony (in the sense that the testimony is a true report of their beliefs). The normal reason for assuming this is provided by (EA) – subjects' intention to tell the truth about their beliefs plus their beliefs cause the testimony. If we assume that the correlation holds for this reason, then we would already be assuming the falsity of [Y] in one respect in order to test the crucial prediction necessary to provide justification of either of the interpretations of the Libet experiments which claim that intentions do not cause the hand movements. We can only justifiably believe that intentions do not cause the hand movements if we justifiably believe that they do cause the apparent testimony about them.

However we might have good grounds to believe that in the particular circumstances of Libet-type experiments, apparent testimony is not caused by the intention to produce it, while nevertheless being in general reliable (i.e. correctly reporting the testifiers' beliefs). But these grounds could only be provided by a wider scientific theory about when apparent testimony to a belief about a testifier's conscious life was or was not correlated with the occurrence of that belief, and someone's apparent memory of their past conscious life was true. A justified belief in that scientific theory would require a justified belief that the theory made true predictions. The predictions would need to be predictions of when on other occasions subjects' apparent testimony was correlated with their beliefs and their own apparent memories were true. . But in order to have a justified belief that these predicted correlations occurred we must rely ultimately on apparent testimony and memory and so – by (EA) - assume that subjects' apparent testimony was caused by intentions to report true beliefs, and apparent memory was caused by the conscious events apparently remembered.

I conclude that the Libet-type experiments have not so far shown that in their experimental circumstances intentions do not cause bodily movements; and – even if the crucial predictions necessary to show this proved correct – that would only show that epi-

phenomenalism held in these circumstances on the assumption that in general it was false[18].

## NOTES

1. Robinson [2007], (in *Stanford Encyclopedia of Philosophy*) defines 'epiphenomenalism' as 'the view that mental events are caused by physical events in the brain, but have no effects upon any physical events'. But all discussions bring out that the 'mental events' involved are 'conscious events'. So Robinson [2009] (in *The Oxford Companion to Consciousness*) defines 'epiphenomenalism' as 'the view that conscious events do not themselves have effects', adding that 'some of [our] neural events are..causes of our conscious events.' The first clause of this second definition entails that the causation of conscious events by brain events is immediate causation.

2. I understand by 'predict', 'make probable the occurrence of'; and leave open the question whether the predictions need to concern events observed (or otherwise experienced) after the formulation of the theory.

3. Staudacher [2006] ) suggests a reliabilist account of justification which includes among the reliable correlations which give rise to justified beliefs, a supposed reliable correlation between a subject's conscious events and that subject's later beliefs about the content of those conscious events; these justified beliefs could then provide evidence in the form of evidence about the subject's conscious events which could be used to check the predictions of epiphenomenalism. But correlations don't just happen; there has to be some causal mechanism which sustains them. The mechanism must be either that conscious events cause the later beliefs, or that there is a common cause of the earlier conscious events and later beliefs about them. If mechanism is the former, it rules out epiphenomenalism. If the mechanism is the latter, it presupposes epiphenomenalism and so cannot be used to check its predictions.

4. While this paper repeats (in section III) some arguments against epiphenomenalism discussed by others to the effect that any justification of the second conjunct [Y] of epiphenomenalism by evidence about the occurrence of conscious events is self-defeating, these discussions seldom mention the need for conscious events to cause brain events if we are to remember our own past conscious events. See e.g. Pauen [2006] or Robinson [2009]), but contrast Hyslop [1998] (see my note 12.) These arguments are not normally presented in the framework of a wider epistemology (as in my section III), and the discussions do not normally consider whether someone might hold a justified belief in epiphenomenalism in virtue of holding, solely on physical evidence, a justified belief that the physical realm is causally closed.

5. I use here the way of distinguishing mental from physical events made in Swinburne [1997] Part I.

6. Burge [1993] points out that we do not normally use (apparent) memory and testimony as 'evidence' for an inference, but rather take their reliability for granted in extending our knowledge. But he also points out that that reliability can on any occasion be called into question.

7. Some philosophers hold that the principle that what people seem to be telling us that they experienced or learnt, they probably did experience or learn is not an a priori principle, but one for which we have evidence: we

have apparent memories that in the past apparent testimony has proved correct. Thus Hume [1902] section 10, pt. 1: 'The reason why we place any credit in witnesses and historians, is not derived from any connection, which we perceive a priori, between testimony and reality, but because we are accustomed to find a conformity between them'. If that were so, the authority of testimony would still remain, given that it is supported by apparent memory. But if in fact the evidence of apparent memory were to suggest that apparent testimony is not reliable, then each person would have to depend for his justified beliefs on his own apparent experience and memory; and then, it will be apparent, only a few people could have a justified belief in any scientific theory and then only one weakly justified.

8. Part I of Robert Audi's textbook *Epistemology* [1998] entitled 'Sources of Justification, Knowledge, and Truth' has chapters on 'Perception', 'Memory', 'Consciousness', 'Reason', and 'Testimony'. What I have called 'experience' covers what he calls 'perception', 'consciousness', and 'reason'.

9. See, for example, Chalmers [1996] ch 5.

10. Audi [1998] defends the causal nature of perception, memory, consciousness, and (in effect) testimony. Thus (his p.28) 'perception is a kind of causal relation', (p.56) 'causal connections to the past are essential to genuine remembering', (p.81) 'the process by which introspection leads to introspective beliefs..is..causal', (p.137) 'with testimonially grounded knowledge..there must be a certain kind of unbroken chain from the belief constituting that knowledge to a source of the knowledge in some other mode'.

11. Hyslop ([1998] p.67) claims that we can have justified beliefs in the occurrence of our own past conscious events, given an 'an epiphenomenalist twist...to the notion of remembering'. Presumably the twist consists in holding apparent memories veridical merely in virtue of their believed correspondence with past experiences. But of course it flies in the face of (EA) to suppose that our beliefs that our apparent memories are veridical if we also believe that the appropriate causal relation does not exist.

12. Many of the articles in the popular press about this programme and also many of the technical scientific reports regard this programme as providing evidence that we do or do not have 'free will'. They often ignore the points that there are different ways of understanding 'free will', and that there is a difficult philosophical issue about the kind of free will (if any) which would make its possessor morally responsible for his or her actions. Clearly [Y] entails that our public intentional actions are not free in almost any sense, although it leaves it open whether our 'will' (that is, our intentions) is free. Clearly too [X], entails that we do not have 'free will' in a certain sense, but for a 'compatibilist' that is not the sense relevant to moral responsibility.

13. See Libet [2004], ch.4. There has been much later experimental work developing Libet's work – see the surveys in Hallett [2007] and Haggard [2008]. In my view none of this later work makes any substantial difference to the points made in this paper. Some of this work involves the identification of the brain area which is the primary location of the event which I label below '$B_1$'. Other of this work affects only [X]. Libet claimed that humans have the power to 'veto' their intentions caused by brain events, and thus argued that his results did not show that we do not have 'free will'. The results reported by Brass and Haggard [2007] have some tendency to suggest that the 'veto's' also are caused by brain events; and the results of Soon and others [2008] tend to suggest that subjects have already a strong inclination to bring about one intentional action rather than a different one well before forming the intention to do so. However the results of Trevena and Miller

[2010] tend to suggest that $B_1$, the supposed cause of RP, doesn't cause the intention but is merely the cause of or evidence of the subject paying attention.

14. Libet [2004], p.128. See Mele [2009] chapter 6 for discussion of some of the literature on the temporal relations between the time of the onset of an intention, the time of the subject's first awareness of this onset, and the time which the subject later believed to be the time of his first awareness of this. Mele allows the possibility that the intention might begin before the subject was aware of it. I argue in the next paragraph of this paper that in the experiments discussed the intention is a conscious intention, and so that that is not a possibility.

15. See Mele [2006], pp.32-34.

16. For a collection of similar quotations from neurophysiologists see Mele

[2009], pp.70-72.

17. The former interpretation is the one preferred by most neurophysiologists – thus Hallett [2007] 1181, 'the data favour the parallel model'.

18. I am grateful to Mark Wynn and to Glenn Spigel for correspondence which led me to develop my ideas on this topic, and to Al Mele for his comments on an earlier draft of this paper.

## REFERENCES

Audi, R. (1998), *Epistemology* (London: Routledge).

Burge,T. (1993) 'Content Preservation', *Philosophical Review*, 102, pp.457-488.

Brass, M and Haggard, P. (2007) 'To do or Not to do: The Neural Signature of Self-Control.' *Journal of Neuroscience* **27**, pp.9141-9147.

Chalmers, D. (1996), *The Conscious Mind* (Oxford: Oxford University Press).

Haggard, P.(2008) 'Human Volition: Towards a Neuroscience of Will', *Nature Reviews, Neuroscience*, **9**, pp.934-946.

Hallett, M. (2007) 'Volitional Control of Movement: The Physiology of Free Will' *Clinical Neurophysiology* **118**, pp.1179-1192.

Hume, D. (1902) *An Enquiry Concerning Human Understanding* (reprinted from 1777 edition) second edition (Oxford: Clarendon Press).

Hyslop, A. 'Methodological Epiphenomenalism', *Australasian Journal of Philosophy*, 76 (1), pp.61-70.

Libet, B. (2004): *Mind Time* (Cambridge, Mass: Harvard University Press).

Mele, A.R. (2006) *Free Will and Luck* (Oxford: Oxford University Press).

Mele, A.R. (2009) *Effective Intentions* (Oxford: Oxford University Press).

Pauen, M. (2006) 'Feeling Causes', *Journal of Consciousness Studies,* **13**, pp. 129-152.

Robinson, W.S. (2007) 'Epiphenomenalism' in *Stanford Encyclopedia of Philosophy* (http://plato.stanford.edu)

Robinson, W.S (2009) 'Epiphenomenalism' in (ed.) T.Bayne and others, *The Oxford Companion to Consciousness* (Oxford: Oxford University Press).

Roediger, H.L., Goode,M.K., and Zaromb,F.M. (2008) 'Free Will and the Control of

Action', in (ed) J.Baer, J.C.Kaufman, and R.F.Baumeister, *Are We Free?*

(Oxford: Oxford University Press).

Soon, C.S, Brass, M, Heinze, H.J, and Haynes, J-D (2008) 'Unconscious Determinants of Free Decisions in the Human Brain', *Nature Neuroscience*, **11**, pp. 543-545.

Staudacher, A. (2006) 'Epistemological Challenges to Qualia-Epiphenomenalism', *Journal of Consciousness Studies*, **13**, pp.153-175.

Swinburne, R. (1997) *The Evolution of the Soul*, revised edition (Oxford: Oxford University Press)

Trevena, J. A., and Miller, J. O. (2010). 'Brain preparation before a voluntary action: Evidence against unconscious movement initiation', *Consciousness & Cognition,* **19 (1***),* pp.447-456.