

Susan Pockett

Does Consciousness Cause Behaviour?

Abstract: *The prevailing folk-psychology of modern Western humans holds that (1) an individual person or self can be identified with that individual's consciousness and thus that (2) consciousness can cause behaviour. In this paper I assemble a body of experimental and philosophical evidence suggesting that both of these beliefs might be mistaken. I discuss some of the practical and philosophical implications of the idea that consciousness does not in fact cause behaviour and conclude that it is not only a serious academic possibility but also an (almost) intuitively acceptable idea that consciousness may be epiphenomenal.*

Introduction

There may be much to recommend Bertrand Russell's remark 'The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm' (Russell, 1913, p. 1). However, arguments about the reality of causation notwithstanding, the fact remains that when a healthy adult human makes a freely initiated body movement, he or she usually *experiences* both a sense of volition — the feeling that s/he (and not someone else) is willing or causing the movement — and a sense of agency — the feeling that s/he (and not someone else) is carrying out the movement. Most persons living in Western cultures in the twenty-first century would identify the 'I' who wills and carries out their voluntary movements with their conscious minds. We identify our self with our individual consciousness and we believe that our consciousness is what causes the voluntary components of our behaviour.

This strong intuitive sense that voluntary actions are willed and carried out by *a consciousness* goes to the heart of the ancient Free Will v. Determinism debate. The problem boils down to the fact that it is difficult to see how consciousness could possibly break in to the deterministic/random chain of events in the brain and impose a new direction on these. If I am my consciousness and I cause my

Correspondence: Susan Pockett, Department of Physics, University of Auckland, Private Bag 92019, Auckland, New Zealand. Email: s.pockett@auckland.ac.nz

Journal of Consciousness Studies, 11, No. 2, 2004, pp. 23–40

behaviour, then even the watered down varieties of Free Will espoused by compatibilists require that consciousness should in some way be able to impose a new direction on brain events.¹ Even if we conceive of choice between alternatives and voluntariness as being possible in the absence of any genuine act of origination,² still *my choosing* to do something requires me (my consciousness) to adjust what is going on in my brain. How could consciousness do this? It's hard to say.³ But it seems that if we allow the possibility that we *don't* have Free Will, we find ourselves in an even worse position. The trouble is that we badly want Free Will. We feel that our autonomy and personhood would be unacceptably diminished if we didn't have it. What's more, our subjective phenomenological experiences, our ethical belief system, the Christian faith (despite a pull towards predestination) and the whole might of the Western legal system combine to assure us that we do have Free Will. It's a large system to buck.

This problem has proved insoluble for so long now that it is becoming reasonable to conclude there might be some radical error in the world-view which generates it. The present paper examines the proposition that there are not one but two such errors, the second of which follows logically from the first. These putative errors are:

- (1) Identification of the subject, agent or self who wills and carries out actions with the *consciousness* of the person performing the task, rather than with their sub- or preconscious brain activity.

-
- [1] There are probably as many definitions of Free Will as there are philosophers, but basically the definitions can be subdivided into two kinds: strong and weak. Those philosophers known as Incompatibilists adopt a strong definition of Free Will. Such a definition requires that humans have the capacity to violate strong Determinism. Strong Determinism is the view that everything which happens in the world is the necessary result (effect) of an earlier event (cause). Taken to its logical conclusion, this implies that everything which happens to, or is done by, any given person was predestined before that person was born. Incompatibilists accept strong Determinism as scientific truth and thus believe that strong Free Will, which would have to break the chain of cause and effect, is impossible. On the other hand, those philosophers known as Compatibilists adopt a weak definition of Free Will. In this kind of definition, the word 'free' simply implies an absence of compulsion. If no one is holding a gun to your head, or if you do not have some inner compulsion like kleptomania, then you are free to do as you like, which means you have free will. The question of the truth or otherwise of strong Determinism is usually glossed over by Compatibilists (although some have made an attempt to invoke quantum mechanical solutions). Some nod to the problem by drawing a distinction between origination (see Footnote 2) and carefully chosen definitions of choice and volition that do not violate Determinism, but they then simply concentrate on the latter and ignore the question of whether genuine origination is possible. Probably a majority of philosophers since Hume and most lay people are Compatibilists. The majority of scientists, whose professional life tends to be based on the assumption of Determinism, do not think about the question at all — but those who do are more likely to be Incompatibilists.
 - [2] An originated action is one that has no causal antecedent. By definition, a truly originated action is impossible in a deterministic world.
 - [3] One possibility is to argue that consciousness is identical with (and/or a different aspect of) some feature of the normal deterministic chain of cause and effect interactions in the brain. This strategy is compatible with the existence of Free Will if one adopts a compatibilist definition of Free Will such that genuine origination is not required, only choice between alternatives, but is even more than usually *incompatible* with the existence of Free Will if genuine origination is required, since consciousness is now actually defined as deterministic.

- (2) The consequent conviction that consciousness *per se* must have a causal role in brain function.

The idea that these two propositions are errors is by no means a new one. A fairly large number of authors have already advanced the idea that one or both of them are false: notably in recent years Velmans (2000; 2002) and Wegner (2002). My contribution here is partly to assemble a body of evidence on the matter which includes some experimental findings and philosophical notions that have not previously been presented in conjunction with each other, and partly to show that the consequences of accepting putative Errors (1) and (2) *as* errors are actually not as major as might be imagined.

A route map of the paper is as follows. First, a number of reasons for considering that it might be a mistake to identify the self-which-initiates-actions with consciousness are outlined. The initial sketch is then filled in with details. From this platform, it is pointed out that a full acceptance of the falseness of Proposition (1) would imply the epiphenomenality of consciousness. Two arguments against epiphenomenality are considered and rejected. Various practical and philosophical implications of accepting the epiphenomenality of consciousness are then discussed and it is shown that the consequences of epiphenomenality are really not as major as might be thought. I conclude that it is a serious possibility that consciousness may be epiphenomenal.⁴

I: Proposition (1) — Consciousness as Agent?

There are a number of empirical and philosophical observations which, taken together, suggest that it might be a mistake to identify the agent who wills actions with the subject's consciousness. In outline, these are:

- (1) The senses of volition and agency are conscious experiences. Like any other conscious experience, they are generated by the normal functioning of specific parts of the brain.
- (2) If these parts of the brain are physically damaged or temporarily inactivated by physiological or psychological intervention, the senses of agency or volition are not felt, even when the subject has actually willed and/or carried out a movement. Thus the *experience* of volition is not necessary for the performance of actions that would normally be considered volitional.
- (3) Conversely, under certain pathological or experimental conditions, the senses of volition or agency may be felt when the subject has not actually willed or caused a particular movement. The experience of volition can be felt incorrectly.
- (4) Under normal circumstances, the brain activity which immediately precedes a voluntary bodily movement begins before the subject experiences the sense of having willed the movement. In other words, it seems that the

[4] Of course this would still not completely solve the Free Will/Determinism problem — there remains the question of whether human brains in a pre- or sub-conscious mode have Free Will. It would, however, alter the way in which the question can legitimately be framed.

movement is initiated by the brain subconsciously, before the brain regions giving rise to the senses of volition and agency are activated.

- (5) Simple introspection shows that rapid reactions occur before we become conscious of acting and that quite complex overlearned acts such as typing can be performed intentionally (I am not typing this manuscript by accident) but without ongoing awareness of the many individual movements which comprise the act.
- (6) There exists a philosophical argument to the effect that consciousness itself can not 'act' in the philosophical sense, meaning that it has no power to *originate* any physical or mental activity.
- (7) Because it is presently impossible to generate conscious experiences in the absence of neural processing, there is presently no evidence to show that consciousness *per se* (over and above the neural processing that accompanies it) is essential for any biological function.

Evidence for each of the above statements is as follows.

(1) *Brain areas responsible for sensations of agency and will*

The anatomy of agency. Brain imaging studies suggest that the sense that a particular action was carried out by the subject and not by somebody else is generated in the anterior part of the insula (McGuire *et al.*, 1996; Fink *et al.*, 1996; Kircher *et al.*, 2000, 2001; Ruby and Decety, 2001; Farrer and Frith, 2002). The insula is a structure lying deep in the floor of the central sulcus, overlying and more or less coextensive with the claustrum and putamen of the lentiform nucleus in the basal ganglia. The function of the insula is not well established, but lesions to it provoke somatic hallucinations in epileptic patients (Roper *et al.*, 1993), and it has been reported to be involved in integrating visual and auditory signals of movement (Lewis *et al.*, 2000). Farrer and Frith (2002) suggest that

one aspect of the experience of agency that we feel when we move our bodies through space is the close correspondence between many different sensory signals. In particular there will be a correspondence between three kinds of signal: somatosensory signals directly consequent upon our movements, visual and auditory signals that may result indirectly from our movements, and last, corollary discharge associated with motor commands that generated the movements. A close correspondence between all these signals helps to give us a sense of agency . . . When subjects attribute an action to themselves . . . attention is directed toward representations integrating the many sensory signals associated with the action. These representations are found in the anterior insula.

In contrast, imaging studies suggest that the main brain area involved in generating the sense that a particular action is carried out by somebody other than the subject is the right inferior parietal lobe (McGuire *et al.*, 1996; Spence *et al.*, 1997; Ruby and Decety, 2001; Farrer and Frith, 2002). The inferior part of the parietal cortex in particular is active both when the sense that someone else controls the action is correct and also when it is incorrect (as in mistakes by the subject during experimental manipulations, or schizophrenic delusions of alien

control). Lesions in the right parietal cortex can lead to disorders such as a form of the neurological syndrome known as ‘the alien hand’ in which the patient’s hand is perceived to be under the control of someone else (Leiguardia *et al.*, 1993; Bundick and Spinella, 2000). Other patients with right parietal lesions have been described as perceiving their limbs not only as not belonging to them, but as actually belonging to other people (Critchley, 1953; Nightingale, 1982; Daprati *et al.*, 2000).

The anatomy of will. Some brain imaging experiments have been interpreted as showing that actions are initiated in the dorsolateral prefrontal cortex (DLPFC) of the brain (Frith *et al.*, 1991; Jahanshahi *et al.*, 1995). However, Frith (Frith and Gallagher, 2002) now feels that these experiments do not tell the whole story and that ‘we haven’t identified the key element of will, if there is such a thing’. Deecke (1996) points out that voluntary decisions involve answering three questions: what to do, how to do and when to do. It seems likely that each of these questions may be answered in a different area of the brain.

The path taken by neural information from the DLPFC to the muscles is not completely clear, but it involves the sub-cortical basal ganglia, the cerebellum (sometimes), the supplementary motor area (SMA), cingulate motor area (CMA), premotor area (PMA) and primary motor area (M1) of the cortex and finally the pyramidal and/or extrapyramidal tracts leading to the motor neurons in the spinal cord and brainstem. The precise order in which these areas are activated is the subject of some debate.

Interestingly, activity in only two of the brain areas involved in the generation of movements seems to decline as a task is learned well enough to become automatic and thus not to require conscious monitoring. These two areas are the prefrontal cortex and the supplementary motor area (Deiber *et al.*, 1991; Seitz and Roland, 1992; Jenkins *et al.*, 1994). It is thus tempting to speculate that consciousness of the ‘what’ and ‘how’ of a voluntary decision to act may be generated in various areas of the prefrontal cortex, while the conscious correlate of the ‘when’ may be generated in the SMA, as essentially suggested by Deecke and his colleagues (Deecke, 1996).

In this context, it would be most interesting to know the precise timing of SMA activation in relation to the timing of conscious awareness of the decision to act in Libet-type experiments such as those described in section (4). Unfortunately, precise localization of the neural activity measured by EEG is fraught with difficulty and the time resolution of imaging methods like fMRI has, at least until now, not been adequate to answer this question.

(2) *Voluntary movement without the experience of volition*

There are a number of conditions under which apparently purposeful movements are made in the absence of any sense of volition. In some of these circumstances, the subjects report that their movement was initiated by some agent outside themselves. This may be conceived to be the hypnotist when hypnosis is involved, some form of alien being in schizophrenia, or simply some other

unknown person in cases of neurological damage like those described in section (1). However, simple performance of an overlearned task such as typing can also give rise to a series of quite complicated movements which the subject later reports not having been conscious of performing. In these circumstances the subject is more likely to say they performed the actions ‘automatically’ or ‘unconsciously’ than to ascribe volition to some outside agency, but the upshot is similar. Actions which appear at least from the outside to be completely volitional are performed without the subject’s being consciously aware of willing them.

(3) *Experience of volition without voluntary movement*

Conversely, there are also a number of conditions under which the subject reports having consciously willed an action which they actually did not or were not able to make.

Wegner (Wegner and Wheatley, 1999; Wegner, 2002) contends that the experience of willing an act arises simply from interpreting one’s thoughts as the cause of the act. Like any other cause-effect attribution, he says, this one is based on three factors:

- (i) the timing of the perceived cause (which must occur before the perceived effect, but not too far before it),
- (ii) the consistency of the perceived cause with the perceived effect and
- (iii) the exclusivity of the perceived cause (i.e. the absence of any other possible cause of the perceived effect).

Wegner and Wheatley (1999) demonstrate the role of priority or timing in a psychological experiment where subjects mistakenly end up believing that they willed a particular action, when in fact they were forced to perform it and simply led to think about it just before it happened. The conclusion drawn by the authors is that

the unique human convenience of conscious thoughts that preview our actions gives us the privilege of feeling we wilfully cause what we do. In fact, unconscious and inscrutable mechanisms create both our conscious thought about action and create the action as well, and also produce the sense of will we experience by perceiving the thought as the cause of action. So although our thoughts may have deep, important and unconscious causal connections to our actions, the experience of conscious will arises from a process that interprets these connections, not from the connections themselves. Believing that our conscious thoughts cause our actions is an error based on the illusory experience of will — much like believing that a rabbit has indeed popped out of an empty hat.

Another example of the experience of volition in the absence of any actual movement is provided by subjects who are paralysed, either experimentally or by accident of nature. Such subjects still experience volition, but the volition has no observable effect. Patients who are immobilized by the motor neuron disease ALS (amyotrophic lateral sclerosis) do still generate motor readiness potentials a.k.a. Bereitschaftspotentials when they ‘will’ movements they are unable

actually to make (Westphal *et al.*, 1998) and it seems likely that these EEG event-related potentials are at least connected with the experience of volition.

Interesting observations on the sense of effort experienced by paralysed subjects during ineffective willed movements are provided by Gandevia (1982). He reports that

Patients who became suddenly hemiplegic, without sensory symptoms, noted that attempts to move when first paralysed were not accompanied by a sense of effort, but that attempts to move when movement first returned were accompanied by distinct sensations of effort or heaviness. This absence of a sense of effort . . . did not occur either in a group of patients with pure lower motor neuron paralysis without sensory signs or in patients with clinically complete spinal transection.

On the basis of this Gandevia suggests that the sense of effort may be due to efferent activity in the cortex. Presumably, when the motor cortex is first compromised by a stroke, efferent activity ceases completely (because the neurons in the cortical area generating this activity have died from lack of oxygen), so there is no sensation of effort. But as adjacent cortical areas take over the function of the affected part, they initially generate an excess of efferent commands, which constitute a sensation of extreme effort.

Gandevia's later experiments on exercise-induced fatigue in normal subjects also demonstrate the lack of any fixed relationship between the sensation of effort or volition and the muscular outcome. He finds that exercise-induced fatigue is at least partly due to central rather than peripheral processes, and that the central fatigue is at least partly due to events 'upstream' of the motor cortex (Taylor *et al.*, 2000; Gandevia, 2001). In other words, central drive diminishes with fatigue, despite the subject's best efforts to deliver the maximal possible muscular force. It appears that the 'will' can become physiologically tired, just like the muscles.

In yet another demonstration of the lack of correspondence between the experience of volition and what actually happens, Ammon and Gandevia (1990) show that transcranial magnetic stimulation of the frontal and/or prefrontal cortex produces a significant preference for selection of one hand over the other in a forced-choice task. Single magnetic stimuli subthreshold for hand movement were used and the resulting hand preference depended on the direction of the induced current. Most interestingly, although the subjects were aware of the stimulation (because of its effects on scalp muscles under the coil) they were completely unaware that it had influenced their decisions, which they felt to have been made in an entirely natural way.

(4) Movement-related brain activity starts before the experience of volition

Benjamin Libet has famously shown that in normal subjects, the brain-generated Bereitschaftspotential or readiness potential (RP) which precedes a spontaneous voluntary movement begins on average 343 ms before subjects become conscious of deciding to make the movement (Libet *et al.*, 1982; 1983). Libet himself clearly perceives that his finding negates the idea of the

movement's being *initiated* by consciousness. However he recoils from the full force of this conclusion and adopts the position that consciousness may still be able to step in and veto the action before it is performed — that even if we don't have free will, we do have free won't (Libet, 1985; 1999).

Because of the importance of the conclusions that can be drawn from them, these experiments have been extensively discussed in the past (see, for example, the commentaries on Libet, 1985). More recent work shows that the experiments are repeatable in several independent laboratories (Keller and Heckhausen, 1990; Haggard and Eimer, 1999; Trevena and Miller, 2002) and that they do stand up methodologically, despite a flurry of recent experimental and theoretical work on possible timing errors (Joordens *et al.*, 2002; Pockett, 2002a,b; Trevena and Miller, 2002). The movements studied do comply with an everyday (i.e. relatively weak) definition of freely willed actions. However it seems to me quite likely that the subjects may have interpreted the experimental instructions as requiring them to set their brain motor systems in a threshold state and then wait for a random neural event to initiate each action, rather than actually deciding voluntarily when to make the movement. Thus the reported time of the decision to move may actually have been the time at which the subject *became aware* that this random neural event had happened and the action was under way. Additionally, even if this interpretation is completely wrong and each movement really was preceded by a definite decision to move, this decision could only have been concerned with the 'when' of the movement, not the 'what' or the 'how' of it. Both *what* the movement would be and *how* to make it had been decided well in advance. These experiments certainly show that sometimes (and only sometimes, even in these experiments) 'decisions' about *when* to make a particular movement are made preconsciously. But they do not address at all the question of whether larger decisions about what to do and how to do it are routinely made preconsciously.

(5) *Introspection shows consciousness neither necessary nor sufficient*

We can probably all recall some situation in which we reacted to a potentially life-threatening event with such lightning speed that we found our foot already on the brake pedal (or whatever) by the time we became conscious of the danger. Consciousness is clearly not necessary for the initiation of *reactions*. But perhaps it is not legitimate to consider reactions if we are concerned to elucidate the initiation of voluntary actions.

One voluntary action performed by all of us every day is the act of getting out of bed in the morning. This is the subject of a famous piece of introspection by that master of the art William James (James, 1890, Part II, p. 524). No matter how hard he tries, James finds himself unable to capture the moment at which he decides to get out of bed on a cold morning. His repeated observation is that after a prolonged period of procrastination, the action suddenly just happens. It seems to be initiated preconsciously. However it is interesting to note that James, like Libet, then remains unable even to consider the possibility that his own

observations might imply the acausality of consciousness. On the contrary, he spends a whole chapter of his magnum opus (James, 1890, pp. 128–44) determinedly attempting to refute what he calls ‘the automaton-theory’ as put forward by the likes of T.H. Huxley. This is not a new debate.

But perhaps even the act of getting out of bed is too far down on the scale from completely involuntary (reflex withdrawal from a noxious stimulus) to completely voluntary (writing a twenty-page paper). Let us consider the act of speaking. All commentators who argue for the importance and efficacy of consciousness agree that one of the main features of human life for which consciousness is absolutely required is the production of language. But Velmans (2002) puts forward a number of arguments and pieces of evidence to the effect that one becomes conscious of what one wants to say *only after one has said it*. Writing is just an extension of speaking. It is certainly my direct experience at this moment that the finger movements I am making in typing these words are not directed consciously. More controversially, but still (I aver) definitely, it is my current experience that the undeniably hard work involved in deciding what words to type in this situation is also happening preconsciously. Certainly I could not type this paper if I were completely *unconscious*. However it seems to me a genuine possibility that this is simply because the nervous system state necessary for actions such as typing is also a state which involves the generation of conscious sensations. This does not necessarily imply a causal relationship between the conscious sensations and the production of words. I know what I think only when I see what I write.

Finally, let us consider a little more explicitly the kind of operation that we feel most distinguishes humans, in all our conscious glory, from mere machines — the act of scientific or artistic creation. A common feature of unusually creative persons is the frequency of the ‘Aha!’ experience in their lives. The approach taken by such people to the solution of a difficult problem tends to be to input all relevant data, mull over the problem for a bit and then forget about it, confident that the answer will suddenly pop into consciousness fully formed, usually while the subject is consciously engaged in doing something else. All the work of producing a novel and creative solution to the problem has gone on *subconsciously*. Only the solution, which may be considered by others to be the fruit of genius, the tutelary spirit of the person, is presented to the conscious mind.

So what is the difference between actions that are perceived by the actor to be voluntary and those that are not? As Wegner (2002) points out, what gives us the impression that *we have* consciously caused something is nothing more than the correct temporal conjunction between thoughts about the act and the occurrence of the act. If the act occurs without this conjunction, it seems as though someone else did it. For example, it is very notable that acts of creation are often perceived by the creator not to be generated by themselves. Explanations of who *did* do the work of creation have historically varied from the artist’s Muse to the mathematician’s subconscious mind, but in all cases of genuine creation some degree of magic is felt to have been at work. The creator tends to have the eerie feeling that ‘It wasn’t me who did it’.

(6) *Philosophical argument that consciousness cannot act*

The philosophical argument in question is put forward in slightly different fashions by both Ryle (1949) and Smith (1999). The basis of the argument is the proposition that, as Ryle puts it: 'One thing I cannot prepare myself for is the next thought I am going to think.' Smith expresses the matter as follows:

Is it possible for me to know (be conscious of) [the] next word [I am going to speak]? The answer is trivially 'No' because of the way 'next word' is defined, namely, a word which has *not yet appeared* in I_0 [my consciousness]. One might incorrectly argue that if I were, say, reciting some familiar verse, surely I would know the next word at any point in the verse even though it had not actually appeared in I_0 [my consciousness]. In truth, although the next word is certainly in my long-term memory, it is, by definition of 'next', logically impossible for me to be conscious of it.

This argument leads Smith to the logical premise that '*it is impossible for me to know my next conscious experience*'.

Smith's next question is 'If I cannot know my next conscious experience, *what determines what it will be?*' His answer is that the contents of my consciousness at any given time are determined by processes in my brain that took place earlier, but of which I was not conscious. Thus he comes to the conclusion that my consciousness 'cannot "do" anything such as steering my thoughts . . . If I subjectively feel, as we all do, that I_0 [my consciousness] controls my thoughts and behaviour, it can only be because that feeling is itself a part of the consciousness generated by my brain.'

(7) *Lack of any unequivocal demonstration that consciousness per se, as opposed to the neural processing that covaries with it, is necessary to explain anything*

It has been claimed by various investigators that consciousness is necessary for identification of complex or novel stimuli, for choice, for learning, for memory and for the organization of complex, novel responses, particularly those requiring planning. Velmans (1991) reviews evidence that consciousness performs none of these functions.

Baars (1988) lists nine 'major functions' of consciousness: definition and context-setting; adaptation and learning; editing, flagging and debugging; recruiting and control function; prioritizing and access-control; decision-making or executive function; analogy-forming function; metacognitive or self-monitoring function; autoprogramming and self-maintenance function. In fact however, while he demonstrates at length that awareness often *accompanies* all of these functions, Baars provides no evidence at all that consciousness itself, as distinct from the neural processing that goes with it, is actually necessary for any of them. Possibly he draws no distinction between consciousness and its neural correlates because, like many neurophysiologists or psychologists respectively, he operates on the assumption that either the neural processing itself or the function performed by the neural processing *is* consciousness.

The point is this. Unless one makes one of the above assumptions about the nature of consciousness (and they are only assumptions), when a certain set of neural processes always accompanies a particular conscious experience *it is impossible in principle* to distinguish between a serial model, in which the conscious experience *per se* is an essential step in ongoing function, and a parallel model, in which the conscious experience is merely an offshoot that happens always to occur as a secondary result of the processing. It doesn't help to say that if conscious experience is not present, the nine functions listed above (or any other functions you care to name) do not occur. On the parallel model, any brain state in which consciousness is absent must also be a state that disallows the neural processing underlying *both* the function *and* the experience.

It does not wash to say that because one can report the contents of consciousness verbally, consciousness must be capable of causal action on at least the language subsystems of the brain. In fact, one of the defining features of conscious experience is that it is essentially private — we *cannot* report its contents. We literally cannot tell anyone else, verbally or in any other way, what it is like to have our conscious experiences. If you doubt this, try to imagine how you would go about explaining to someone who has been blind since birth what it is like to see a computer screen. If that seems too hard, simply try explaining to someone who is red-green colour blind what the colour red looks like to you. It can't be done. All we can do is assign labels to the brainstates evoked in us by certain circumstances and assume that the sensations experienced by others in similar circumstances will be roughly similar to ours. All phenomenological description and all psychophysical measurement relies not on any direct communication about qualia or sensory experiences themselves, but on *comparisons* between different experiences, each of which is itself intrinsically unreportable. Such are the frustrations of the situation that some commentators, e.g. Dennett (1994), have been led to deny the very existence of qualia.

II: Proposition (2) — Consciousness as Epiphenomenon?

If it really is a mistake to believe that our conscious thoughts cause our actions, it seems likely that it is also a mistake to believe that consciousness is capable of any direct causal action on the brain at all. We are led inexorably to the proposition that consciousness may actually be an epiphenomenon.

This is certainly a counterintuitive and probably to most people a highly unpalatable notion. One's first impulse is to squash the idea. But how? A number of arguments against epiphenomenalism have been put forward over the years and effectively rebutted (Robinson, 2003). I will deal here only with two of the major ones, on which I feel I have something new and useful to say.

The first is a subjective, introspective argument. We simply *know*, from the inside, that consciousness causes our acts. It seems absurd to deny this — it's obvious. But is it really? I believe this intuition has been dealt a solid blow by the cumulative weight of the evidence presented above. I leave it to the reader to decide whether or not this blow is mortal, but it does seem beyond debate that, as

Claxton (1999) puts it, '[conscious] intentions are neither necessary nor sufficient for intelligent actions to occur'.

A second major argument against the epiphenomenality of consciousness is that if consciousness were an epiphenomenon it would have no function, which immediately raises the question *why does it exist at all?* Inability to provide a satisfactory answer to this question might be seen as evidence against the proposition that consciousness is epiphenomenal.

Satisfactory or not, answers to the question of *why consciousness exists at all* come in (at least) three classes:

- (a) What can be called for want of a better term the nihilist answer. This is that there is no reason to assume that everything which exists has a *raison d'être* — one may as well ask why the universe exists. It's not a productive question.
- (b) The religious answer. One version of this would be that consciousness was created by Consciousness (a.k.a. God, Yahweh, Allah or Brahmin). We are not able to comprehend, this answer would run, either why God exists or why He should have created human consciousness, and we shouldn't ask. Alternatively, a less mainstream version of the religious answer might be that human consciousness was created by a bored minor deity who only wanted someone to play with.
- (c) The biological answer. This is likely to be something along the lines that consciousness must have some useful function in enabling animals to survive and reproduce, otherwise it wouldn't have evolved. From this it follows that because I have not been able to give a plausible account of the *function* of an epiphenomenal consciousness, all the other arguments in this paper must be wrong. (To which it is apposite to reply that despite a multitude of words on the topic there also remains a startling lack of consensus on what the function of a *non*-epiphenomenal consciousness might be — which would seem to be odd if it actually had one).

How are we to evaluate these answers? It must be admitted that (a) as it stands bears a disturbing resemblance to the former version of (b) i.e. that we can't know why consciousness exists and shouldn't ask. To a member of the intrinsically curious species *H. sapiens* this does not seem a very satisfactory position to take, but it might be right. The latter version of answer (b) is entertaining, but unlikely to be taken seriously in our present culture.

The biological answer, which *is* likely to be taken seriously in our present culture, seems at first to be a killer for the notion that consciousness is an epiphenomenon. This answer carries with it all the gravitas accumulated over the last century and a half by the Theory of Evolution, and it will probably be the answer subscribed to by a large percentage of the readers of this paper. But is answer (c) as it stands the last word on the matter from the scientific point of view? It is not.

The loophole is that, even granting the literal truth of the theory of evolution, not every characteristic that has evolved in biological organisms need

necessarily have done so because it assisted in the leaving of offspring. Logically, it is possible for some characteristics to have evolved which are simply innocent by-products of other features that *are* very advantageous biologically. It is quite consistent with evolutionary theory that such by-products could have neutral or even slightly negative value in the selection stakes and still have appeared during the course of evolution. Viewed in this light, the confident assertion that consciousness is too *biologically expensive* to have evolved if it weren't of some use (Searle, 2001) seems to me completely unproven. Consciousness need 'cost' nothing at all in addition to the neural processing which happens, as a historical accident perhaps, to generate it.

Thus a combination nihilist/biological position does seem to me to be tenable, if not quite what we might have expected (or perhaps wanted) to believe. It does at least provide a consistent and I believe plausible answer to the question of why consciousness should exist at all if it is nothing but an epiphenomenon.

This means that both of what I see as the major arguments against epiphenomenality are at least flawed, if not conclusively floored.

III: Implications of Epiphenomenalism

Opponents of these ideas might argue that while I may have shown it is *sometimes* an error to attribute volition and agency to consciousness, I have by no means shown that this is *always* the case. True. I have not shown that it is always an error to believe that one's consciousness initiates one's actions, and the logical impossibility of showing that anything is always the case is no excuse. However, it might legitimately be suggested that the tenacity with which one clings to this particular logical lifeline may depend largely upon how serious one perceives to be the implications of letting go. Let us consider these implications.

(a) *Implications for human self-concept.* At first blush, acceptance that it is an error to identify one's self exclusively with one's consciousness seems to require a shift in world-view of Copernican proportions. It means that our own consciousness may no longer occupy the centre of our personal universe. Our conscious selves are no longer *in control*. They are nothing more than impotent products of our wilful and wayward subconscious brains. To many, this idea may be so unpalatable as to be downright unthinkable.

However, a little further reflection might possibly allow the emergence of a less hysterical view on the matter. Perhaps accepting that Proposition (1) is wrong would not really require so big a change in self-image after all. Whether or not one accepts all his tenets, Sigmund Freud did initiate a general acceptance that a great deal of what causes our actions is not only subconscious but quite inaccessible to the conscious mind (unless of course one chooses to pay for several years worth of psychoanalysis). The real pity is that in Freud's *zeitgeist*, subconscious impulses are generally taken to be dark and/or dangerous. In fact they are not normally so at all. If one accepts the general correctness of Mr Darwin's Theory of Evolution, it is plain that brains must have evolved to further the interests of the animals they control, not to defeat them. Of course, if one rejects

evolution and prefers a religious account of the construction of the world, some of the more brimstone-laden versions of the major Western religions do support the Freudian vision of an unconscious Id steeped in Original Sin, kept in check only by the strenuous efforts of the conscious Ego. But even on religious ground, Buddhists and Hindus have held for centuries that the entity Freud called the ego is just an illusion and that the ground-state of being is bliss rather than evil. Perhaps all we really need to do is relax, renounce the common tenets of the fundamentalist versions of Puritanism, Catholicism, Freudianism and Islam, and trust that our brains (a.k.a. our subconscious minds) have evolved to work for our benefit rather than our destruction. Then we can safely stop *wanting* to control everything on a conscious level. It's not so hard really.

(b) *Implications for philosophy.* The main philosophical implication of accepting that Proposition (1) is an error is that we would then have to accept that Free Will as we have previously conceptualized it (i.e. as involving choices made by consciousness) is nothing but an illusion. Philosophically speaking, this would at first sight seem to imply that determinism wins the long battle between it and free will. No problem for the fairly large number of philosophers and/or deeply theistic persons who have already come down on the side of determinism anyway, for one reason or another, but deeply upsetting for those who are unable either to ignore their own direct experience or to give up the feeling that, as James (1890) puts it, 'The whole sting and excitement of our voluntary life . . . depends on our sense that in it things are *really being decided* from one moment to another, and that it is not the dull rattling off of a chain that was forged innumerable ages ago'.

A closer look at the issue, however, shows that accepting that Proposition (1) is an error does not necessarily imply the correctness of determinism at all. It simply shifts the locus of free will from consciousness to sub- or pre-conscious brain processes. We are still left with the question of whether or not it is possible for *brains* to make genuine choices, which are not preordained by physical law and which could have gone otherwise. Although it has little or no direct bearing on the present argument, I cannot resist the opportunity to expound on my own current position on this, as follows.

What might be called philosophical determinism assumes that in any given situation there is only one possible outcome that is consistent with the laws of physical causality. But in fact, in systems as complicated as the brain or the weather, it is not true in practical or operational terms that there is only one macroscopic outcome consistent with the microscopic laws of physical causality. Even assuming the truth of philosophical determinism, these systems can be considered to be *operationally* non-deterministic. The science of nonlinear dynamics reveals that there are bifurcation points in the temporal evolution of very complex systems at which a very minor alteration can cause one or another of the macroscopic competing influences to win out and flip the system into one or another state. Although nonlinear dynamical systems are still conceptualized mathematically as being philosophically deterministic — i.e. it is assumed (in the interests of allowing the use of various mathematical techniques to describe

the system) that if it were possible to turn back time and rerun any given situation again with all the conditions exactly the same the outcome would be the same — the complexity of the real-world situation is such that at bifurcation points there *is* what could be considered to be a genuine possibility that the thing could go either way. Certainly any prediction that can be made as to the outcome can only be in terms of statistical probabilities, contingent on the relative strengths of the various influences involved. So, if (i) the brain can validly be considered a non-linear dynamical system and (ii) we join the Oxford Dictionary in defining the word ‘choose’ simply as meaning ‘select out of a greater number’, it seems that in practice there may well be scope for brains to make genuine choices between alternatives while still obeying all known physical laws. All we are renouncing by accepting Proposition (1) as an error is the idea that these choices are made *by consciousness*. All we are embracing is the idea that they are made by our brains, sub- or sometimes pre-consciously.

(c) *Implications for the legal system.* As with the implications for our self-concept, the practical implications for the legal system of accepting Proposition (1) as an error seem at first to be quite major, but turn out actually to be not much of a big deal after all. As described above, we are not being asked to accept that normal persons cannot make genuine choices — we are only being asked to accept that these choices are not made *by consciousness*. They are made by brains, pre- or subconsciously.

The main implication of this for the legal system is probably that the concept of diminished responsibility would have to go. The whole mind-brain entity that comprises a person would have to be held responsible for that person’s choices and acts, not just the conscious part. But is this really so big a change? At present, in most civilized jurisdictions (i.e. those in which capital punishment is not considered to be an option), the only *practical* difference between the treatment of someone found guilty of first degree murder and someone found not guilty of murder by reason of insanity lies in which institution serves as their prison — a penitentiary or a mental hospital. The fact that the perpetrator did not consciously decide to do the deed does not stop society from recognizing the twin needs to sanction the act (in the legal sense) and prevent its recurrence by locking the killer away.

In the matter of lesser crimes, the main consequences of accepting the importance of the subconscious generation of actions would be some further blurring of the already blurred line between intentional and accidental acts. It would probably end up being logically necessary to hold persons responsible for all of their non-accidental acts, *whether or not they were aware of intending to perform them*. In one sense this might make it more difficult to determine whether an outcome was intentional or accidental, but in another sense it does perhaps clarify what should happen when an act was clearly not accidental but the perpetrator truthfully says they don’t know why they did it. In such a case they should hardly be held less responsible than they would be for a consciously premeditated act.

Thus libertarians need not fear that accepting the pre-conscious origin of voluntary actions would lead to a situation where no one could be held responsible

for anything they did. On the contrary, the probable outcome of such an acceptance is that people would be held more responsible than ever.

Conclusions

I have summarized a number of persuasive pieces of empirical and philosophical evidence to suggest that consciousness is often not the direct cause of behaviour which would universally be accepted as ‘voluntary’, or freely willed. I have even shown that the notion that consciousness is epiphenomenal is not as unacceptable on an intuitive, everyday level as it first appears — that we could, perhaps, given a little fortitude, live with the idea. The remaining question would seem to be this. Is the position that consciousness is an epiphenomenon empirically testable? Is it in principle possible to show that any particular instance of behaviour *has* been directly caused by consciousness *per se*, as distinct from the neural processing that underlies it?

At first sight, the answer to this question seems to depend on what consciousness turns out to be. If consciousness turns out to be identical with neural processing, as neural identity theorists would have it, then of course the answer is no, because there was no difference between the consciousness and the processing in the first place. But if consciousness turns out not to be identical with neural processing, if it turns out to be something other, something perhaps generated by and then superimposed on neural processing, then yes, it may be possible to show that consciousness can directly cause behaviour. The experiment would simply be to generate a lump of consciousness artificially, throw it at a brain and show that the consciousness causes the brain to make its body do something. But then again, this would not be terribly convincing — a nay-sayer could complain ‘All you’ve done is stimulate the brain artificially. You can cause behaviour by stimulating the brain with a simple jolt of electricity, yet no one thinks a simple jolt of electricity is consciousness’. Also of course, it may not be possible for the lump of artificial consciousness to cause bodily movement anyway — because the truth may be that consciousness really can’t cause behaviour. This is harder than it looks.

In conclusion, I fear that the answer to the question ‘is the position that consciousness is an epiphenomenon testable empirically?’ may well be ‘no, it isn’t’. I would like to be able to test the proposition scientifically, but I’m simply not smart enough to think of a way to do it. So this is still just philosophy.

References

- Ammon, K. and Gandevia, S.C. (1990), ‘Transcranial magnetic stimulation can influence the selection of motor programmes’, *Journal of Neurology, Neurosurgery and Psychiatry*, **53**, pp. 705–7.
- Baars, B.J. (1988), *A Cognitive Theory of Consciousness* (Cambridge: Cambridge University Press).
- Bundick, T. Jr. and Spinella, M. (2000), ‘Subjective experience, involuntary movement and the posterior alien hand syndrome’, *Journal of Neurology, Neurosurgery and Psychiatry*, **68**, pp. 83–5.

- Claxton, G. (1999), 'Whodunnit? Unpicking the "seems" of free will', *Journal of Consciousness Studies*, **6**, pp. 99–113.
- Critchley, M. (1953), *The Parietal Lobes* (New York: Hafner Press).
- Daprati, E., Sirigu, A., Pradat-Diehl, P., Franck, N. and Jeannerod, M. (2000), 'Recognition of self-produced movement in the case of severe neglect', *Neurocase*, **6**, pp. 477–86.
- Deecke, L. (1996), 'Planning, preparation, execution and imagery of volitional action', *Cognitive Brain Research*, **3**, pp. 59–64.
- Deiber, M.P., Passingham, R.E., Colebatch, J.G., Friston, K.J., Nixon, P.D. and Frackowiak, R.S.J. (1991), 'Cortical areas and the selection of movement. A study with positron emission tomography', *Experimental Brain Research*, **84**, pp. 393–402.
- Dennett, D. (1994), 'Instead of qualia', in *Consciousness in Philosophy and Cognitive Neuroscience*, ed. A. Revonsuo and M. Kampinnen (Hillsdale NJ: Lawrence Erlbaum).
- Farrer, C. and Frith, C.D. (2002), 'Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency', *Neuroimage*, **15**, pp. 596–603.
- Fink, G.R., Markowitsch, H.J., Reinkemeier, M., Bruckbauer, T., Kessler, J. and Heiss, W.D. (1996), 'Cerebral representations of one's own past: neural networks involved in autobiographical memory', *Journal of Neuroscience*, **16**, pp. 4275–82.
- Frith, C.D., Friston, K.J., Liddle, P.F. and Frackowiak, R.S.J. (1991), 'Willed action and the prefrontal cortex in man: a study with PET', *Proceedings of the Royal Society of London Series B*, **244**, pp. 241–6.
- Frith, C. and Gallagher, S. (2002), 'Models of the pathological mind', *Journal of Consciousness Studies*, **9**, pp. 57–80.
- Gandevia, S.C. (1982), 'The perception of motor commands or effort during muscular paralysis', *Brain*, **105**, pp. 151–9.
- Gandevia, S.C. (2001), 'Spinal and supraspinal factors in human muscle fatigue', *Physiological Reviews*, **81**, pp. 1725–89.
- Haggard, P. and Eimer, M. (1999), 'On the relation between brain potentials and awareness of voluntary movements', *Experimental Brain Research*, **126**, pp. 128–33.
- Jahanshahi, M., Jenkins, H., Brown, R.G., Marsden, C.D., Passingham, R.E. and Brooks, D.J. (1995), 'Self-initiated versus externally triggered movements. 1. An investigation using measurement of regional cerebral blood-flow with PET and movement-related potentials in normal and Parkinsons-disease subjects', *Brain*, **118**, pp. 913–33.
- James, W. (1890), *The Principles of Psychology* (New York: Henry Holt & Co).
- Jenkins, I.H., Brooks, D.J., Nixon, P.D., Frackowiak, R.S.J. and Passingham, R.E. (1994), 'Motor sequence learning. A study with positron emission tomography', *Journal of Neuroscience*, **14**, pp. 3775–90.
- Joordens, S., Van Duijn, M. and Spalek, T.M. (2002), 'When timing the mind one should also mind the timing: bias in the measurement of voluntary actions', *Consciousness and Cognition*, **11**, pp. 231–40.
- Keller, I. and Heckhausen, H. (1990), 'Readiness potentials preceding spontaneous motor acts: voluntary vs involuntary control', *Electroencephalography and Clinical Neurophysiology*, **76**, pp. 351–61.
- Kircher, T.T.J., Senior, C., Phillips, M.L., Benson, P.J., Bullmore, E.T., Brammer, M., Simmons, A., Williams, S.C.R., Bartels, M. and David, A.S. (2000), 'Towards a functional neuroanatomy of self processing: effects of faces and words', *Cognitive Brain Research*, **10**, pp. 133–44.
- Kircher, T.T.J., Senior, C., Phillips, M.L., Rabe-Hesketh, S., Benson, P.J., Bullmore, E.T., Brammer, M., Simmons, A., Bartels, M. and David, A.S. (2001), 'Recognizing one's own face', *Cognition*, **78**, B1–B15.
- Leiguardia, R., Starkstein, S., Noguez, M., Berthier, M. and Arbelaz, R. (1993), 'Paroxysmal alien hand syndrome', *Journal of Neurology, Neurosurgery and Psychiatry*, **56**, pp. 788–92.
- Lewis, J.W., Beauchamp, M.S. and DeYoe, E.A. (2000), 'A comparison of visual and auditory motion processing in human cerebral cortex', *Cerebral Cortex*, **10**, pp. 873–88.
- Libet, B. (1985), 'Unconscious cerebral initiative and the role of conscious will in voluntary action', *Behavioral and Brain Sciences*, **8**, pp. 529–66.
- Libet, B. (1999), 'Do we have free will?', *Journal of Consciousness Studies*, **6**, pp. 47–57.
- Libet, B., Wright, E.W. Jr, Feinstein, B. and Pearl, D.K. (1982), 'Readiness potentials preceding unrestricted "spontaneous" vs pre-planned voluntary acts'. *Electroencephalography and Clinical Neurophysiology*, **54**, pp. 322–35.

- Libet, B., Gleason, C.A., Wright, E.W. and Pearl, D.K. (1983), 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential)', *Brain*, **106**, pp. 623–42.
- McGuire, P.K., Silbersweig, D.A. and Frith C.D. (1996), 'Functional neuroanatomy of verbal self-monitoring', *Brain*, **119**, pp. 907–17.
- Nightingale, S. (1982), 'Somatoparaphrenia: a case report', *Cortex*, **18**, pp. 436–67.
- Pockett, S. (2002a), 'On subjective back-referral and how long it takes to become conscious of a stimulus: a reinterpretation of Libet's data', *Consciousness and Cognition*, **11**, pp. 144–61.
- Pockett, S. (2002b), 'Backward referral, flash-lags and quantum free will: a response to commentaries on articles by Pockett, Klein, Gomes and Trevena and Miller', *Consciousness and Cognition*, **11**, pp. 314–25.
- Roper, S.N., Levesque, M.F., Sutherling, W.W. and Engel, J. (1993), 'Surgical treatment of partial epilepsy arising from the insular cortex. Report of two cases', *Journal of Neurosurgery*, **79**, pp. 266–9.
- Ruby, P. and Decety, J. (2001), 'Effect of subjective perspective taken during simulation of action: a PET investigation of agency', *Nature Neuroscience*, **4**, pp. 546–50.
- Robinson, W. (2003), 'Epiphenomenalism', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Spring 2003 edn.), <http://plato.stanford.edu/archives/spr2003/entries/epiphenomenalism>
- Russell, B. (1913), 'On the Notion of Cause', *Proceedings of the Aristotelian Society*, **13**, pp. 1–26.
- Ryle G. (1949), *The Concept of Mind* (London: Hutchinson House).
- Searle J.R. (2001), 'Free will as a problem in neurobiology', <http://www.bitesizebooks.co.uk/searle.htm>
- Seitz, R.J. and Roland, P.E. (1992), 'Learning of sequential finger movements in man: a combined kinematic and positron emission tomography (PET) study', *European Journal of Neuroscience*, **4**, pp. 154–65.
- Smith, R.L. (1999), 'A testable mind-brain theory', *Journal of Mind and Behavior*, **20**, pp. 421–36.
- Spence, S.A., Brooks, D.J., Hirsch, S.R., Liddle P.F., Meehan, J. and Grasby, P.M. (1997), 'A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control)', *Brain*, **120**, pp. 1997–2011.
- Taylor, J.L., Allen, G.M., Butler, J.E. and Gandevia, S.C. (2000), 'Supraspinal fatigue during intermittent maximal voluntary contractions of the human elbow flexors', *Journal of Applied Physiology*, **89**, pp. 305–13.
- Trevena, J.A. and Miller, J. (2002), 'Cortical movement preparation before and after a conscious decision to move', *Consciousness and Cognition*, **11**, pp. 162–90.
- Velmans, M. (1991), 'Is human information processing conscious?', *Behavioral and Brain Sciences*, **14**, pp. 651–726.
- Velmans, M. (2000) *Understanding Consciousness* (London and Philadelphia: Routledge).
- Velmans, M. (2002), 'How could conscious experiences affect brains?', *Journal of Consciousness Studies*, **9**, pp. 3–29.
- Wegner, D.M. (2002), *The Illusion of Conscious Will* (Cambridge MA: MIT Press).
- Wegner, D.M. and Wheatley, T. (1999), 'Apparent mental causation: sources of the experience of will', *American Psychologist*, **54**, pp. 480–91.
- Westphal, K.P., Heinemann, H.A., Grozinger, B., Kotchoubey, B.J., Diekmann, V., Becker, W. and Kornhuber, H.H. (1998), 'Bereitschaftspotential in amyotrophic lateral sclerosis (ALS): lower amplitudes in patients with hyperreflexia (spasticity)', *Acta Neurologica Scandinavica*, **98**, pp. 15–21.

Paper received October 2002, revised May 2003