

Psychological Epiphenomenalism

Darryl Mathieson, ANU

Darryl.Mathieson@outlook.com

Abstract

Researchers in the psychological sciences have put forward the thesis that various sources of psychological, cognitive, and neuroscientific evidence demonstrate that being conscious of our mental states does not make any difference to our behaviour. In this paper, I argue that the evidence marshalled in support of this view—which I call *psychological epiphenomenalism*—is subject to major objections, relies on a superficial reading of the relevant literature, and fails to engage with the more precise ways in which philosophers understand mental states to be conscious. I then appeal to work on implementation intentions to demonstrate that an intention’s being “access conscious” enhances its functional role, which makes it more likely that we will successfully carry out our intended behaviour. The result is that consciousness in at least one relevant sense is not epiphenomenal, with further work remaining to be done to show how other kinds of consciousness cause behaviour too.

1. Introduction

It is standardly held that mental states—thoughts, intentions, beliefs, desires—can cause our behaviour. Suppose I desire a cup of tea and believe that I am out of crispy cake rusks. It is reasonable to think that these beliefs and desires are among what causes me to put the kettle on and check the biscuit bucket. Indeed, this is the standard story of action in philosophy (Davidson 1970, Velleman 1992). However, it has also been a truism of psychology for at least a century that most of what goes on inside the human mind is unconscious, and that we can have thoughts, intentions, beliefs, and desires that we are not conscious of which nevertheless influence our behaviour. Suppose, then, that in addition to having various mental states, that we are also conscious (in some sense) of having them. Does being *conscious* of our mental states ever feature in the causes of our behaviour?

According to recent work in the science of consciousness, the answer is no (Oakley & Halligan 2017, Halligan and Oakley 2021). It is argued that various sources of psychological, cognitive, and neuroscientific evidence demonstrate that unconscious processes provide a sufficient causal explanation for our behaviour. Neither our mental states, nor our consciousness of them (in any sense) make any difference to our psychological or behavioural outputs. Both said to be “epiphenomenal”—caused by the brain but causally inert themselves. Notice the two distinct claims here. One is whether mental states cause behaviour, which is such a well-trodden topic in philosophy that I have little to add (see the enormous literature inspired by Kim 1998). Another is whether mental states cause behaviour in virtue of their being *conscious* in some sense. This latter question is what I find most interesting, important, and about which I have something to say. Given that the main motivation for those who respond in the negative involves appealing to various sources of psychological evidence, I call the view *psychological epiphenomenalism*.¹

Conversely, according to what I call the *default view*, it is standardly held that a mental state’s being conscious can sometimes make a difference to our behaviour.² For example, subjects who consciously form specific intentions to perform an action tend to be more successful at carrying out the action (Gollwitzer 1993, 1999; Gollwitzer and Sheeran 2006). As Bargh (2017) has extensively argued, becoming consciously aware of the ways in which the past, present, and expected future affect us can help us to change our behaviour,

¹ I will use Oakley and Halligan’s (2017) and Halligan and Oakley’s (2021) work as target articles in this paper, but for a sense of the broader approaches, motivations, and responses to psychological epiphenomenalism, see Balaguer (2019), Baumeister et al., (2018), Bonicalzi and Haggard (2019), Lavazza (2019), Lumer (2019), Mele (2009, 2018), Nahmias (2002), Niker, Reiner, and Felsen (2018), Stockdale (2022), Velmans (2000), Wegner (2002), Wegner and Wheatley (1999). For the purpose of this paper, I am interested exclusively in psychological epiphenomenalism, not the broader metaphysical considerations for and against epiphenomenalism (not that I am suggesting these are devoid of scientific considerations), see Jackson (1982) and Robinson (2010).

² Psychologists tend to define a state’s being conscious in terms of experiencing that state and being able to report on it, which is a bit less precise than the phenomenal/access/higher-order distinction common in philosophy (more on this shortly). Phenomenal consciousness captures the former usage, while access and higher order capture the latter.

and without this awareness we are apt to be the unwitting tools of unconscious influences (see also Bargh & Hassin 2022). Moreover, some of the best theoretical models of conscious processing suggest that consciousness plays a distinct role of making information globally accessible to multiple brain systems (Baars 1988, Baars et al., 2021, Dehaene & Changeux 2011). After a comprehensive literature review, Baumeister and colleagues (2011) conclude, “The evidence for conscious causation of behaviour is profound, extensive, adaptive, multifaceted, and empirically strong” (p. 331) (see also Baumeister et al. 2018). It appears that various sources of psychological evidence are being used to argue for starkly opposing conclusions about the causal efficacy of consciousness. Who, then, is right?

Much of the disagreement turns on how the relevant psychological evidence is interpreted and how consciousness is understood. I am inclined to think that the psychological epiphenomenalist’s conclusion that unconscious processes sufficiently cause human behaviour is vastly overblown and relies on a superficial reading of the relevant research. Further, there is a lack of engagement with the comprehensive and sophisticated philosophical literature on the three main ways in which mental states can be conscious—an engagement that would make the problem of conscious causation more tractable. For example, mental states can be *phenomenally* conscious just in case there is something it feels like to be in that state (Nagel 1974). Alternatively, mental states can be *higher-order* conscious, just in case the subject is aware of being in that state through introspection (Rosenthal 2005). Finally, mental states can be *access* conscious if they are available for rationally guiding speech and action, or, put somewhat more technically, if the functional role of that mental state—the various causes and effects that the mental state is disposed to produce—is enhanced in specific ways (Block 1995). Access consciousness is the sense that I will defend against psychological epiphenomenalism.

Here is the structure of the paper. In §2 I lay out the three main bodies of evidence used in support of psychological epiphenomenalism. In §3 I show that none of this evidence supports the view. In §4 I appeal to research on implementation intentions as evidence that access conscious mental states can make a substantial difference to the likelihood that we will successfully carry out our intended actions. I consider and respond to three objections in §5 and make some concluding remarks in §6.

2. Psychological Epiphenomenalism

According to epiphenomenalists, everyday experience furnishes the strong but misleading impression that being conscious of our mental states causes our behaviour because our consciousness of our mental states typically occurs prior to the subsequent action. But what we are not conscious of are the various unconscious mental operations within us, and so it is claimed that we misattribute the cause of our actions to being conscious of our mental states, rather than the unconscious processes that precede them. Given the mounting psychological evidence detailing how much of what goes on inside us is unconscious, it has been argued that we should abandon the default view that consciousness is sometimes among the causes of our actions in favour of the view that a mental state's being conscious (in any sense) is never causally efficacious and that unconscious processes sufficiently cause actions instead.

The evidence used by Oakley and Halligan (2017) and Halligan and Oakley (2021) to reject the default view can be roughly grouped into three categories, all of which have the general aim of undermining the necessity of consciousness. First, research has demonstrated that various implicit cognitive and behavioural abilities remain operational despite damage to conscious functionality. This is especially clear in patients with various impairments due to injury and neurological deficits like physical trauma, strokes, disease, tumours, visual and spatial neglect, prosopagnosia, amnesia, dyslexia, blindsight, and hypnosis (Bargh & Hassin

2022, Frigato 2014, Litman & Jaffe 2022, Oakley & Halligan 2017, Reber 1992, Schacter & Graf 1986, Weiskrantz 1986). For example, patients with bilateral lesions in the V1 area of the primary visual cortex report no conscious visual experience. Despite being cortically blind, however, various unconscious visual functions still operate, producing a phenomenon called blindsight (Weiskrantz 1986). When prompted, blindsight patients can identify visual objects, follow targets with their eyes or fingers, and can even successfully navigate narrow hallways cluttered with obstacles—challenging the belief that perceptions must be conscious to affect or produce actions.

Second, experimental evidence from attempting to quantify the timing of when mental states (typically intentions and decisions) become conscious has also demonstrated that a significant amount of work takes place unconsciously in processes that appear to be consciously initiated and/or controlled (Fried et al. 2011, Haggard 2005, Libet 2001, Pockett 2006, Roediger et al. 2008, Soon et al. 2008).

For example, Libet (1985) had participants flex their wrist whenever they consciously intended to do so, and to take note of where a revolving point was on a clock when they became consciously aware of the intention. Using electroencephalography, readiness potentials (preparatory brain activity) were found to occur in the motor-cortex one-third of a second before participants became conscious of their intentions or decisions. Subsequent replications by Haynes (2011) and Soon and colleagues (2008) using fMRI found that they could predict with 60% accuracy which of two buttons subjects would press based on neural activity occurring in various cortical areas up to ten seconds before participants consciously intended to act. Using depth electrodes, Fried and colleagues (2011) found that they could predict subjects' impending decisions to press a key with 80% accuracy, 700ms before subjects consciously decided. According to Halligan and Oakley (2021), the evidence shows

Penultimate draft—appearing in the *Journal of Consciousness Studies*—please cite published version

that unconscious processes are what initiate and cause our actions, and that the consciousness of our decisions occurs too late in the process to count as a cause.

It is one thing to accept that fast and efficient unconscious processes are responsible for ‘low level’ mental operations like constructing perceptual experiences, and it may not be surprising that they can initiate and cause simple motor actions like wrist flexing. However, increasing research demonstrates that unconscious processes are also capable of performing many high-level complex functions like abstract thinking, self-control, decision making, remembering, making social comparisons, inferences, learning, goal pursuit, problem solving, planning, and reasoning (Bargh 2017, Bargh & Morsella 2008, Bargh & Williams 2006, Dijksterhuis & Aarts 2010, Dijksterhuis & Nordgren 2006, Hassin 2013, Sklar, Kardosh & Hassin 2021, Wegner 2002, Wilson 2002). Based on a review of the cognitive, social psychological, and neuroscientific data, Hassin (2013) proposed that unconscious processes can perform the same, fundamental, high-level functions that conscious processes can perform (see also Goldstein & Hassin 2017). Oakley and Halligan (2017) take the evidence to motivate a new model of the mind, where unconscious processes do all the causal work (see Fig 1).

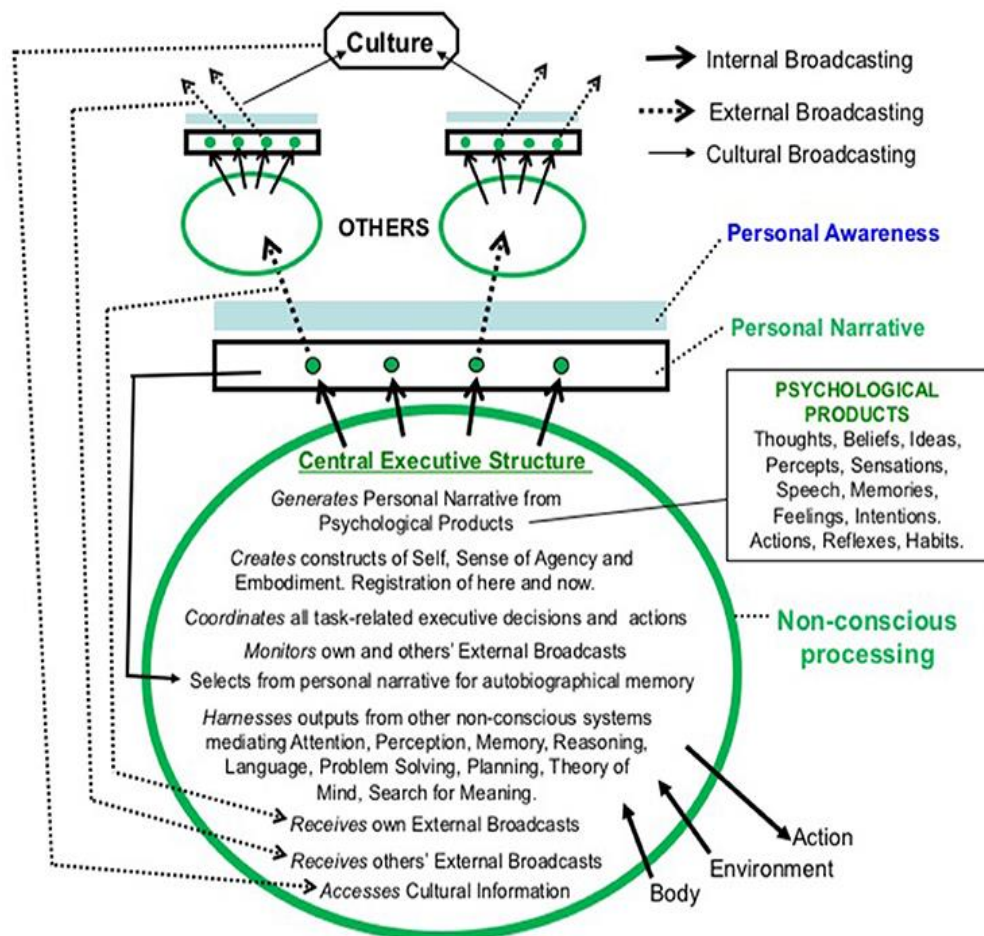


Fig. 1. The Oakley-Halligan Model (Oakley & Halligan 2017).

According to this model, all neuropsychological processing takes place in unconscious systems called the *Central Executive Structure* (CES). Just like the central nervous system coordinates and regulates bodily functions like temperature, breathing, and digestion, so too the CES produces and controls all actions like reasoning, problem solving, decision making, and planning, and selects (“internally broadcasts”) some of the psychological products (thoughts, sensations, intentions, beliefs, memories, emotions) for inclusion in a self-referential *Personal Narrative*. The personal narrative is an unconsciously generated, continuously updated, post-hoc story about what we do that can then be communicated (“externally broadcast”) by the CES to other individuals. *Personal Awareness* is a passive observer of this personal narrative. A distinction is drawn between the contents of

consciousness and consciousness itself, and both are argued to be epiphenomenal end-products of unconscious processes. As Oakley and Halligan (2017) state:

In this view, both the personal narrative and the associated personal awareness are end-products of widely distributed, efficient, non-conscious processing that arrives too late in the psychological process cycle for there to be a reason to infer the necessity of an additional independent executive or causal capacity to either of them (p. 13).

An example will help illustrate how this model works. Suppose I spot a homeless person on the street and think to myself, *I'm going to offer them something to eat or drink*, and then act accordingly. The default view is that being conscious of my deliberation was among what caused me to do what I did when I did it. However, according to the epiphenomenal model, both my thoughts and my consciousness of them were passive by-products of neurocognitive processing in the CES and played no role in causing my action. It is the unconscious workings of the CES which caused my action and constructed the personal narrative about it, and I mistakenly believed that being conscious of my mental states was among the causes of my action. This is an easy mistake to make, says the psychological epiphenomenalist, because the consciousness of our mental states tends to precede the action, and the content of the mental states that we are conscious of usually match the action to be performed.

3. Objections to the Evidence

Each of these sources of evidence used to support epiphenomenalism are subject to major objections. I will address the scientific evidence in reverse order this time.

First, let us begin with the claim that consciousness is not necessary for any high-level functions. In the spirit of taking psychological epiphenomenalism at its strongest point, let us grant for the sake of argument that Hassin (2013) is right that unconscious processes can

perform every fundamental, high-level function traditionally thought to require consciousness.³ Even if consciousness is not necessary for a given behaviour, x , on a given occasion, it does not follow that consciousness is epiphenomenal more generally. It is plausible to conclude that if, on a given occasion, event x is sufficiently caused by an unconscious process, then consciousness, on that given occasion, does not cause x . Here is what you cannot infer: if, on some given occasion, event x is sufficiently caused by an unconscious process, then on every occasion event x is sufficiently caused by an unconscious process.

Put slightly differently, even if unconscious processes can do everything that conscious processes can, it does not follow that they always will. In fact, Hassin (2013) himself makes it clear that the YIC principle (“Yes It Can” – that unconscious processes can perform the same, fundamental, high-level functions that conscious processes can perform) “does not imply that consciousness does not play a role in our lives. Rather, it sends us to look for the differences elsewhere: not in the functions themselves but in how they kick in and play out (p. 196).”⁴ The YIC principle—even if it were true—should not be recruited in support of epiphenomenalism.

Alternatively, the psychological epiphenomenalist might intend the YIC principle as a kind of causal exclusion argument (Kim 1998), which could be formulated as follows: if our behaviour is sufficiently caused by unconscious processes, then to avoid causal overdetermination, consciousness must be excluded from playing a causal role. Suppose C is

³ Note that although Hassin’s work is used by psychological epiphenomenalists to support their view, Hassin (2013, p. 202) explicitly denies that his principle entails epiphenomenalism. It is also worth acknowledging that Hassin’s principle is controversial. For objections, see Baumeister et al. (2018) and Hesselmann and Moors (2015).

⁴ Another way of putting this principle is that conscious processes can be more effective than unconscious processes (and vice versa) depending on the different circumstances and constraints that are present.

an instantiated conscious process, *U* is an instantiated unconscious process, and *B* is the behavioural output. The psychological epiphenomenalist may assert the following:

- (1) *C* is causally sufficient for *B*
- (2) *U* is causally sufficient for *B*
- (3) *C* is distinct from *U*
- (4) If *x* is causally sufficient for *y*, then nothing distinct from *x* is causally sufficient for *y*.

Not all of these can be true at once. (4) is the exclusion principle about causation, which is often thought of as obviously true. (3) is obviously true since the relevant distinction is that *C* is conscious while *U* is not. The psychological epiphenomenalist might argue that (2) follows from Hassin's YIC principle. Therefore, (1) must be false, and consciousness is causally excluded. But here is why this exclusion reasoning does not work: (2) is false.

Hassin's YIC principle is expressed as a modal "can", but the psychological epiphenomenalist will need something stronger than a modal claim, namely, that unconscious processes *in fact* cause the same things as conscious processes. Hassin's principle is the more modest claim that unconscious processes *can* be causally sufficient for a given behaviour, and that there is no function that conscious processes can perform that unconscious processes cannot. But despite performing the same functions, there are differences in the ways that conscious versus unconscious processes carry out a given function, with circumstances and constraints leading to differences in efficacy. To generate exclusion reasoning, the outcomes or effects must be the same. If they are not, there is no threat of causal overdetermination.

To borrow a few examples from Hassin (2013), the vast automatization literature suggests that the more automatic a process becomes, the more likely it will be to occur unconsciously. Once we have overlearned something like, for example, driving a manual

transmission car, we are able to turn much of the cognitive workload over to unconscious processes which will perform the same functions as those previously mediated by conscious processes. Similarly, even though unconscious processes can help us to learn complicated rules more effectively than conscious processes, the latter do better at following task-switch cues (Van Gaal et al. 2010). Monitoring mental contents can also occur consciously and unconsciously (Morewedge & Kahneman 2010, Wegner 1994), but the narrower capacity limitations of conscious monitoring will render it less effective under certain circumstances than unconscious monitoring. This is obviously not an exhaustive list, but the bottom line is that there are situations where an unconscious process is not sufficient to achieve the efficacy of a behavioural output mediated by consciousness, and vice versa. I will revisit this point in more detail in §4 when I discuss implementing intentions in action.

Second on the list are the neuroscience experiments on the timing of consciousness. As Stockdale (2022) has persuasively argued, Libet-style experiments cannot be used to support epiphenomenalism given that a core part of their experimental design assumes consciousness is causally efficacious. How so? Recall that participants are asked in all these experiments to report when they first became consciously aware of the relevant mental state (urge, intention, decision). Conscious awareness must be causally efficacious in these experiments for there to be a reliable correlation between subjects' conscious awareness of their intention to perform a motor action and accurately reporting when this occurred. The method that experimenters use to establish when the conscious intention occurred assumes that the subject's consciousness of her intention causes her to take note of and report the position of the dot on the clock (or whatever the relevant indicator is across different experimental designs). Otherwise, no reliable conclusions could be drawn about when conscious awareness occurs relative to the neural activity and the resultant overt action.

Rather than awareness of the intention and the subsequent reporting standing in some causal relation, epiphenomenalists might maintain a reliable correlation by arguing that prior unconscious processes independently cause the consciousness of the intention and the act of reporting. If this is right, we do not need to assume that the consciousness of the intention is what causes the reporting. This is an empirical claim, and proponents of epiphenomenalism would then need to produce evidence of consciousness occurring without relying on the action of reporting consciousness. However, there is currently no way to conclusively tell when a subject becomes consciously aware of a particular event without relying on their overt self-reports, in which case we cannot exclude conscious awareness from being among the causes of reporting when it occurred.

At this point, psychological epiphenomenalists might fall back on the claim that the consciousness of the intention or decision occurs too late to be among the causes of behaviour. Now, Libet-style experiments certainly show that being conscious of our intention to act does not initiate the entire intention or decision-making process (not that this should surprise anyone). But the consciousness of our intention not being the *initiator* of the entire process leading to action does not show that it is not among the later *causes* of the action. Processes have parts, some distal, some proximal, and the consciousness of our intention is plausibly among the more proximal causes of the action (see also Mele 2009, 2010, 2018). You cannot infer from the datum that unconscious neural activity occurs prior to the consciousness of our intentions that the consciousness of our intentions is thereby excluded from the causal chain—even if it occurs quite close to the action itself. Even Libet himself maintained that we could veto impending decisions or intentions to act once we were conscious of them.⁵

⁵ A common error in the psychological literature that often comes up in response to this point is to argue that the veto was itself caused by unconscious processes, which shows that the countermanding intention (and our

Third, research on neurological deficits and physical impairments to conscious access mechanisms in conditions like blindsight do not offer support to epiphenomenalism either. Suppose a subject has a lesion in the cortex, how do experimenters know whether the conscious access mechanisms have been impaired? Part of the evidence comes from what area the lesion is in, to be sure (lesions in the visual cortex are likely to correlate with some loss of visual consciousness), but the subject must also report a lack of conscious visual experience. There is simply no way of conclusively ascertaining to what degree the individual's conscious perception is impaired without relying on what they self-report (Dehaene et al., 2006). Once again, we must assume that their conscious experiences are causally efficacious for there to be a reliable correlation between their visual awareness (or lack thereof) and the resultant reporting.

Moreover, examples like blindsight do not show that conscious perception, when it does occur, has no different effects. Instead, it is much more plausible that the perceptual states in the blind sighted play similar but less effective functional roles compared to subjects whose perceptual experiences are conscious. The access and phenomenally unconscious visual states of blind sighted patients can sometimes help them to navigate around objects down a narrow hallway, albeit in a much more limited and slow manner than if these states were access (and perhaps phenomenally) conscious.

It is beyond the scope of one paper to exhaustively cover all the evidence that psychological epiphenomenalists appeal to, although it is also worth noting that much of what is left has been comprehensively rebutted elsewhere. For example, Wegner (2002) and

consciousness of it) did not cause the vetoing of the action. This relies on a faulty inference about causation, namely, that if A causes B, then B cannot be the cause of C. But just because something is caused, it does not follow that it cannot itself be a cause of something else. See Mele (2009, p. 70-74) for a comprehensive refutation of this fallacious reasoning. Alternatively, epiphenomenalists might assert that A independently causes both B and C, severing the causal connection between B and C. This view—dubbed “bypassing”—has also been heavily criticized (Nahmias 2002, 2005, 2011).

Penultimate draft—appearing in the *Journal of Consciousness Studies*—please cite published version

colleagues’ (Wegner and Wheatley 1999) much-discussed work on apparent mental causation argues that unconscious processes cause actions independently of the intentional conscious thoughts that we mistakenly believe to cause action. I have nothing new to add to the comprehensive critiques that already exist of this work, which is where I will direct interested readers (Malle 2006, Mele 2009, 2018, Nahmias 2002, 2005). Still, in the spirit of fairness, I will issue a challenge: bring forth any psychological evidence you like, and I am confident that a careful analysis will show that epiphenomenalism will not follow. Instead, the opposite is the case—we can use psychological research to demonstrate the ways in which consciousness does make a difference to what we do. Herein lies the purpose of the remainder of the paper.

4. Conscious Implementation Intentions

Let me begin by noting that the kind of consciousness at issue in arguments about psychological epiphenomenalism is state consciousness. State consciousness is an umbrella term which refers to mental states (as opposed to, for example, creatures) being conscious, which can be further broken down into the three more specific senses I outlined in the introduction (phenomenal, higher order, or access). While it is standardly held that mental states like thoughts, intentions, and decisions can sometimes cause actions, it is an important and interesting further question as to what causal work (if any) is being done by that mental state’s being either phenomenally, higher-order, or access conscious. As it turns out, there is impressive experimental evidence demonstrating that being access conscious of our intentions makes a difference to our actions, so that is the sense I will be concerned with defending.

In what follows, I will motivate this alternative view to epiphenomenalism a bit more by way of a story, lay out the experimental evidence, and then explain why access

consciousness is the most relevant kind of consciousness that is playing a causal role.

Consider the following story.

Torn. Every Saturday, Torn walks to his local supermarket to buy groceries. Outside the automatic door sits a homeless person, who Torn usually ignores. Later that evening, Torn feels torn about his lack of interaction: *What if I was in that situation? He must have been hungry or thirsty. I could at least have offered to get him something to eat or drink.* After some deliberation, Torn comes to a resolution: *If a homeless person is sitting outside the grocery store next Saturday, before I enter, I am going to make sure to ask if they would like something to eat or drink.* Sure enough, the following Saturday Torn sees the homeless man, and this triggers his memory to offer the man something to eat or drink, which he does.

What causes Torn's action? The most obvious answer is that his deliberation caused him to do what he did when he did it, in which case his mental states made a difference to his action, and thereby count as a cause. However, it seems highly implausible that if Torn had not *consciously* deliberated, that he would have been just as likely to do what he did when he did it, which means that there is something about Torn's consciousness of his mental states that caused his action. I will say more about what work is being done by the "conscious" part shortly. First, I will remind us of what the psychological epiphenomenalist is committed to here and introduce some experimental evidence in support of my view to make it worthy of a more careful explanation. The psychological epiphenomenalist is committed to the following counterfactuals, where *U* is an unconscious implementation intention, *C* is Torn's consciousness of his implementation intention, and *B* is his behaviour of offering the homeless person something to eat or drink.

1. If *U* and *C* were to occur, *B* would be no more likely to occur when it did because of *C* occurring.

2. If *U* were to occur, but *C* did not, then *B* would be just as likely to occur when it did.

Both are intuitively unacceptable. To say that Torn's consciousness of his implementation intention made the subsequent action no more likely to occur when it did seems wildly implausible. Similarly, if Torn's thinking was not conscious (suppose instead he had identical but unconscious intentions to help the homeless man), while he might have still acted on this intention at some point, it is unlikely that he would have acted *when* he did. David Lewis's (2000) counterfactual analysis of causes as difference-making influences offers an instructive guide here. On this view, a cause *C* causing an event *E* is for *C* to make a difference to what happens such that, if *C* had not occurred, *E* would either have (a) not occurred, (b) occurred at a different time, or (c) occurred in a different manner. The results of several controlled studies will help to explain why Torn's intention being conscious was a difference-making cause of his action according to (b), in that if his intention was not conscious, he would not have done what he did *when* he did it.

Enter implementation intentions. Conceptualized and developed by Gollwitzer (1993, 1999), implementation intentions promote the attainment of goal intentions ("I intend to *x*") by specifying when, where, and how we will perform the intended goal ("when situation *x* arises, I will perform response *y*"). By specifying the intended behaviour in detail and linking it to the chosen situation, this strategically makes use of the automatic ways that the environment cues our behaviour, making it much more likely that the goal intention will be effective in action once the situation is encountered. Unconscious, automatic processes then take over (e.g., automatic memory retrieval upon encountering the specified environmental stimulus), at which point "action initiation becomes swift, efficient, and does not require conscious intent" (Gollwitzer 1999, p. 495).

The evidence for the efficacy of forming and successfully executing implementation intentions is impressive. One study involved participants who set themselves the goal of performing a breast self-examination (BSE) in the following month. Of the women who reported strong goal intentions to perform the BSE, 100% did so if they made specific implementation intentions, compared to 53% of those who had strong goal intentions alone. Another study examined whether students' participation in performing 20 minutes of vigorous exercise in the following week could be increased through implementation intentions. The initial 29% compliance rate was raised to 39% upon providing a motivational intervention about how exercise reduces the risk of heart disease. Complementing this intervention with forming implementation intentions to exercise raised the compliance rate to 91%. A meta-analysis of 94 independent tests of this kind revealed that forming implementation intentions resulted in a positive effect size of medium to large magnitude on goal attainment (Gollwitzer and Sheeran 2006, p. 69).

Back to Torn. Recall, the psychological epiphenomenalist is committed to the claim that Torn's implementation intention being conscious was not a difference-making cause of his action. According to the epiphenomenalist, the fact that Torn's implementation intention was conscious made it no more likely that he would do what he did when he did it, and if he had not been conscious of his implementation intention, he would have been no less likely to act it out when he did. I suggested that both conclusions are intuitively implausible, and now we can put a finer point on this. In the above experiments, all the subjects who formed and reported specific implementation intentions were more likely to carry out their intended behaviour than those who did not form and report implementation intentions—and all of the subjects who formed implementation intentions were conscious of doing so. The evidence suggests that subjects who consciously make, report, and successfully carry out their implementation intentions would be less likely to do what they did when they said they

Penultimate draft—appearing in the *Journal of Consciousness Studies*—please cite published version

would do it in the absence of having such conscious intentions—meaning that the fact that their intentions to act were conscious is causally relevant to their behaviour.

Now, what exactly is it about an implementation intention's being *conscious* that plays a causal role here? Recall again that philosophers distinguish several different ways that mental states can be conscious, so let us apply those distinctions to implementation intentions. An implementation intention is *phenomenally* conscious if there is something it feels like to be in the state of having an implementation intention. Alternatively, an implementation intention is *higher-order* conscious if the subject is aware (i.e., forms a higher-order belief) that they are in this state via introspection. Both plausibly apply to implementation intentions, but I think the third way in which a state can be conscious—namely, access consciousness—is most relevant for present purposes, because an implementation intention's being access conscious enhances its functional role.

Consider Torn's implementation intention to offer the homeless man something to eat or drink. One of the functional roles of this implementation intention is to dispose Torn to carrying out the action if he wants to. If the implementation intention is access conscious, then the functional role will be enhanced. This enhancement can manifest in at least three different ways (Stoljar 2023).

First, the state will be more likely than it otherwise would be to produce its effects. Torn having his implementation intention means there is some likelihood that he will act it out. When that implementation intention is *conscious*, however, the likelihood increases. Second, the state will have a wider functional role than it normally would. Torn's implementation intention has an associated functional role, but when the implementation intention is *conscious*, its various causes and effects will be wider than they otherwise would be. Third, the state will be more likely to involve attention than it otherwise would be. Torn

having his implementation intention means there is some likelihood that he will focus on, be concerned with, or concentrate on it or the related action. When his implementation intention is *conscious*, however, that attentional involvement increases. He will be more likely to pay attention to his implementation intention (by, for example, concentrating on and being concerned with it when he went home later that evening) than he otherwise would be.

Of these three ways that the functional role of a mental state can be enhanced by its being access conscious, the first sense—an increased *likelihood* of the mental state performing its functional role—is the most obvious way that Torn's implementation intention, being conscious, made a difference to his behaviour. On this view of access consciousness, the following counterfactual conditionals would hold, where *T* is Torn's implementation intention, and *F* is the functional role of that mental state, namely, disposing Torn to help the homeless man.

1. If *T* is access conscious, then *F* will be more likely to occur.
2. If *T* is access unconscious, then *F* would be less likely to occur.

This is exactly consistent with our intuitions about Torn's conscious deliberation being among the causes of his behaviour and aligns neatly with the experimental evidence on implementation intentions. None of this is to deny that Torn's implementation intention might have been causally efficacious by being conscious in other ways too. It is plausible that there was something it felt like for Torn to form his implementation intention, meaning that it was also phenomenally conscious, and it is plausible that he would have been aware of his intention too, making it higher-order conscious. There are many different senses in which a state can be conscious, and it remains an open question as to how each of them effects behaviour. What is clear is that Torn's implementation intention was conscious in the access

sense at the very least, and its being so made a difference to his behaviour. Therefore, the consequent of the preceding conditionals is that access consciousness is not epiphenomenal.

5. *Objections*

Three objections merit a response before closing. First, the psychological epiphenomenalist might object that access consciousness does not vindicate the causal powers of consciousness, per se, but a certain kind of information processing. Rather than attributing a kind of consciousness to ‘access conscious’ mental states, it might be thought that this label simply amounts to picking out an enhanced functional role which could be achieved unconsciously.

Here is why this objection is misplaced. When we are in the state of having an access conscious implementation intention, we are in fact aware of something—in a very general sense of ‘aware of’—in a particular way. Namely, the state of affairs that would obtain if our implementation intention was effective in action. We can then appeal to an idea connected to the higher-order account discussed earlier, which says that a mental state is conscious if the subject is aware of being in that state in a particular way. Given that having an access conscious implementation intention means being aware of something in a particular way, it follows that we are conscious of something. The only difference between the two accounts is that a mental state is conscious according to the higher order view if the subject is aware of the state itself, whereas according to the access view one is conscious of the state of affairs that would obtain if their intention was effective in action. The bottom line is that being in an access conscious state means that you are indeed conscious of something (Stoljar 2023).

Second, one might worry that establishing the causal efficacy of access consciousness does not vindicate the most mysterious kind of consciousness, namely, phenomenal consciousness, and that the latter is the real target of psychological epiphenomenalism.

Halligan and Oakley will not like this, as their epiphenomenal view encompasses both phenomenal and access consciousness. One last quote will illustrate this.

Accordingly, while accepting that one can draw a qualitative distinction between P and A consciousness, we hold that neither engage any cognitive executive functions *per se*. As such, the traditional parsing of psychological states in terms of the presence or absence of consciousness is a relative distinction as both are carried out by underlying brain systems. Perpetuating this distinction serves to constrain the scientific understanding of psychological processes by excluding the reality that non-conscious brain processes are responsible for *all* psychological processes including conscious awareness itself (Halligan & Oakley 2021, p. 4).

It is far from clear that unconscious processes being “responsible for” (i.e., causing) consciousness shows that the latter plays no causal role of its own (causes have causes, see Mele 2009, p. 72). It is also very unclear why the phenomenal/access distinction constrains scientific understanding (as I have argued, the opposite is true—a careful distinction makes the question of conscious causation more scientifically tractable, not less). What is clear, however, is that access consciousness is a target of (at least this version of) psychological epiphenomenalism. Nevertheless, phenomenal consciousness certainly does tend to be the target of epiphenomenalism more broadly, so let me very briefly touch on two points here. First, it is worth noting that psychological epiphenomenalists are writing sceptical papers trying to show that phenomenal consciousness does not cause behaviour, and it is perfectly reasonable to suppose that their phenomenally conscious experiences are among what is influencing them to do just that. Second, as Block (1995) suggests, there may be something about phenomenal consciousness that acts as the means by which other kinds of consciousness (e.g., access consciousness) operate. For example, given that there is plausibly something it felt like for Torn to deliberate and form his implementation intention, this

qualitative feel could have made the implementation intention more access conscious—in which case phenomenal consciousness would feature in the causal chain.

Third, the psychological epiphenomenalist might respond that Torn's conscious experience of forming and implementing his intention is an epiphenomenal state (on Oakley and Halligan's account, it would be a feature of the personal narrative). Instead, unconscious processes might be said to cause the conscious intention and the corresponding overt action.

Epiphenomenalists then owe us a more thorough explanation of what is going on in the scientific literature on implementation intentions. What explains their increased efficacy? Presumably not that subjects later consciously remembered, thanks to consciously specifying in advance and later consciously perceiving the environmental cues, to perform the action. In my view, the fact that the subjects who had implementation intentions (and all of whom were conscious of having them) performed much better than those in the control groups suggests that the intentions themselves *and* the subjects' consciousness of them were causally efficacious. In the language of access consciousness, consciously specified implementation intentions enhance their functional role. If it is not access (or some other kind of) consciousness doing some causal work here, then what is? Remember, the main selling point of psychological epiphenomenalism is its appeal to the empirical evidence. While there is evidence that unconscious anger can make it more likely that people will form implementation intentions compared to unconscious sadness (Maglio, Gollwitzer, & Oettingen 2013), there is no experimental evidence (that I am aware of) showing that subjects form and execute their goal intentions just as effectively by unconsciously (as opposed to consciously) furnishing them with implementation intentions (see also Mele 2009).⁶

⁶ Moreover, even if evidence were produced demonstrating that implementation intentions of which subjects were not conscious were equally as effective as implementation intentions of which subjects were conscious, this would not exclude consciousness from the causal chain. Instead, there are arguably just different ways of acquiring or bringing it about that one forms implementation intentions, some of which might be conscious

Instead, it is much more plausible—and perfectly consistent with the scientific evidence—to accept that when people consciously make implementation intentions, the intention's being conscious (in at least one of the three senses I have discussed) is among the causes of their behaviour. It might be helpful to think about this in the context of your own life. We set intentions for ourselves and fail at achieving them all the time. What should we do? According to psychological epiphenomenalists, consciously thinking about your goal intentions is of zero use. On the alternative view I have been offering, consciously thinking about your goal intentions can make a difference to what you do, providing you consciously think about them in the right way. Consciously and sincerely specifying when, where, and how you will perform the action you desire means you will be more likely to do what you said you would do when you said you would do it. For myself, I have found this strategy enormously effective in reminding me to offer homeless people something to eat or drink whenever I see them outside the supermarket.

6. Conclusion

Given the primacy of unconscious processes, psychological epiphenomenalism plays an important role in reminding us not to take for granted how a mental state's being conscious could count as a cause of our behaviour. However, the right conclusion to draw from the last several decades of psychological research into the unconscious mind is not epiphenomenalism, and none of the evidence cited by various psychological epiphenomenalists supports the view. I hope to have made it clear that consciousness being causally efficacious is perfectly consistent with the various sources of psychological evidence typically marshalled against it. Further important work defending the causal efficacy of other ways in which mental states can be conscious remains to be done.

(e.g., by actively thinking, see Gollwitzer 1999, Gollwitzer & Sheeran 2006), and others not (e.g., by being unconsciously primed, see Maglio, Gollwitzer, & Oettingen 2013).

References

- Baars, B. 1988. *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.
- Baars, B, N. Geld, and R. Kozma. 2021. Global workspace theory (GWT) and prefrontal cortex: Recent developments. *Frontiers in Psychology* 12: 1–6.
- Banks, W. 2006. Does consciousness cause misbehaviour? In *Does Consciousness Cause Behaviour? An Investigation of the Nature of Volition*, eds. S. Pockett, W. Banks, and S. Gallagher, 235–56. Cambridge, MA: MIT Press.
- Bargh, J. 2017. *Before you know it: The unconscious reasons we do what we do*. New York: Touchstone.
- Bargh, J, and R. Hassin. 2022. Unconscious processes in-situ: The kind of awareness that really matters. In *The Cognitive Unconscious: The First Half Century*, eds. R. Allen and A. Reber, 199–222. New York: Oxford University Press.
- Bargh, J, and E. Morsella. 2008. The unconscious mind. *Perspectives on Psychological Science* 3: 73–9.
- Bargh, J, and E. Williams. 2006. The automaticity of social life. *Current Directions in Psychological Science* 15: 1–4.
- Baumeister, R, E. Masicampo, and K. Vohs. 2011. Do conscious thoughts cause behaviour? *Annual Review of Psychology* 62: 331–61.
- Baumeister, R, S. Lau, H. Maranges, and C. Clark. 2018. On the necessity of consciousness for sophisticated human action. *Frontiers in Psychology* 9: 1–6.

Balaguer, M. 2019. Free will, determinism, and epiphenomenalism. *Frontiers in Psychology* 9: 1–14.

Block, N. 1995. On a confusion about a function of consciousness. *Behavioural and Brain Sciences* 18: 227–47.

Bonicalzi, S, and P. Haggard. 2019. From freedom from to freedom to: New perspectives on intentional action. *Frontiers in Psychology* 10: 1–14.

Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Oxford University Press.

Dehaene, S, and J. Changeux. 2011. Experimental and theoretical approaches to conscious processing. *Neuron* 70: 200–27.

Dehaene, S, J. Changeux, L. Naccache, J. Sackur, and C. Sergent. 2006. Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Science* 10: 203–11.

Dijksterhuis, A, and H. Aarts. 2010. Goals, attention, and (un)consciousness. *Annual Review of Psychology* 61: 467–90.

Dijksterhuis, A, and L. Nordgren. 2006. A theory of unconscious thought. *Perspectives on Psychological Science* 1: 95–109.

Fried, I, R. Mukamel, and G. Kreiman. 2011. Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron* 69: 548–62.

Frigato, G. 2014. Reflections on the existence of different types of parallel consciousness. *Journal of Consciousness Studies* 21: 131–51.

- Goldstein, A, and R. Hassin. 2017. Commentary: Definitely maybe: Can unconscious processes perform the same functions as conscious processes? *Frontiers in Psychology* 8: 1–4.
- Gollwitzer, P. 1993. Goal achievement: The role of intentions. *European Review of Social Psychology* 4: 141–85.
- Gollwitzer, P. 1999. Implementation intentions: Strong effects of simple plans. *American Psychologist* 54: 493–503.
- Gollwitzer, P. and P. Sheeran. 2006. Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology* 38: 69–119.
- Haggard, P. 2005. Conscious intention and motor cognition. *Trends in Cognitive Sciences* 9: 290–95.
- Halligan, P, and D. Oakley. 2021. Giving up on consciousness as the ghost in the machine. *Frontiers in Psychology* 12: 1–16.
- Hassin, R. 2013. Yes it can: On the functional abilities of the human unconscious. *Perspectives on Psychological Science* 8: 195–207.
- Haynes, J. 2011. Decoding and predicting intentions. *Annals of the New York Academy of Sciences* 1224: 9–21.
- Hesselmann, G, and P. Moors. 2015. Definitely, maybe: Can unconscious processes perform the same functions as conscious processes? *Frontiers in Psychology* 6: 1–5.
- Jackson, F. 1982. Epiphenomenal qualia. *The Philosophical Quarterly* 32: 127–36.

- Kim, J. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.
- Lavazza, A. 2019. Why cognitive sciences do not prove that free will is an epiphenomenon. *Frontiers in Psychology* 10: 1–11.
- Lewis, D. 2000. Causation as influence. *The Journal of Philosophy* 97: 182–97.
- Libet, B. 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioural and Brain Sciences* 8: 529–66.
- Libet, B. 2001. Consciousness, free action, and the brain. *Journal of Consciousness Studies* 8: 59–65.
- Litman, L. and S. Jaffe. 2022. Implicit cognition in the face of neurological disorders: Implications for neural mechanisms and evolution. In *The Cognitive Unconscious: The First Half Century*, eds. R. Allen and A. Reber, 62–88. New York: Oxford University Press.
- Lumer, C. 2019. Unconscious motives and actions – agency, freedom, and responsibility. *Frontiers in Psychology* 9: 1–16.
- Maglio, S, P. Gollwitzer, and G. Oettingen. 2013. Action control by implementation intentions: The role of discrete emotions. In *Decomposing the Will*, eds. A. Clark, J. Kiverstein, and T. Vierkant, 221–243. New York: Oxford University Press.
- Malle, B. 2006. Of windwills and straw men: Folk assumptions of mind and action. In *Does Consciousness Cause Behaviour? An Investigation of the Nature of Volition*, eds. S. Pockett, W. Banks, and S. Gallagher, 207–231. Cambridge, MA: MIT Press.
- Mele, A. 2009. *Effective intentions: The power of conscious will*. New York: Oxford University Press.

- Mele, A. 2010. Conscious deciding and the science of free will. In *Free Will and Consciousness: How Might They Work?* eds. R. Baumeister, A. Mele, and K. Vohs, 43–65. Oxford: Oxford University Press.
- Mele, A. 2018. Free will, moral responsibility, and scientific epiphenomenalism. *Frontiers in Psychology* 9: 1–8.
- Morewedge, C. and D. Kahneman. 2010. Associative processes in intuitive judgement. *Trends in Cognitive Sciences* 14: 435–40.
- Nagel, T. 1974. What is it like to be a bat? *The Philosophical Review* 83: 435–50.
- Nahmias, E. 2002. When consciousness matters: A critical review of Daniel Wegner’s the illusion of conscious will. *Philosophical Psychology* 15: 527–54.
- Nahmias, E. 2005. Agency, authorship, and illusion. *Consciousness and Cognition* 14: 771–85.
- Nahmias, E. 2011. Intuitions about free will, determinism, and bypassing. In *The Oxford Handbook of Free Will: Second Edition*, ed. R. Kane, 555–577. Oxford: Oxford University Press.
- Niker, F, P. Reiner, and G. Felsen. 2018. Perceptions of undue influence shed light on the folk conception of autonomy. *Frontiers in Psychology* 9: 1–11.
- Oakley, D, and P. Halligan. 2017. Chasing the rainbow: The non-conscious nature of being. *Frontiers in Psychology* 8: 1–16.
- Pockett, S. 2006. The neuroscience of movement. In *Does Consciousness Cause Behaviour? An Investigation of the Nature of Volition*, eds. S. Pockett, W. Banks, and S. Gallagher, 9–24. Cambridge, MA: MIT Press.

Reber, A. 1992. The cognitive unconscious: An evolutionary perspective. *Consciousness and Cognition* 1: 93–133.

Robinson, W. 2010. Epiphenomenalism. *WIREs Cognitive Science* 1: 539–47.

Roediger, H, M. Goode, and F. Zaromb. 2008. Free will and the control of action. In *Are We Free? Psychology and Free Will*, eds. J. Baer, J. Kaufman, and R. Baumeister, 205–25. New York: Oxford University Press.

Rosenthal, D. 2005. *Consciousness and mind*. Oxford: Oxford University Press.

Schacter, D. and P. Graf. 1986. Preserved memory in amnesiac patients: Perspectives from research on direct priming. *Journal of Clinical and Experimental Neuropsychology* 8: 727–43.

Sklar, A, R. Kardosh, and R. Hassin. 2021. From non-conscious processing to conscious events and back: A minimalist approach. *Neuroscience of Consciousness* 7: 1–18.

Soon, C, M. Brass, H. J. Heinz, and J. D. Haynes. 2008. Unconscious determinants of free decisions in the brain. *Nature Neuroscience* 11: 543–45.

Stockdale, B. 2022. The Libet paradigm and a dilemma for epiphenomenalism. *Philosophical Psychology*. doi: 10.1080/09515089.2022.2130744

Stoljar, D. 2023. What consciousness is. In *What is Consciousness? A Debate*, eds. A. Kind and D. Stoljar. Routledge.

Van Gaal, S. V. Lamme, J. Fahrenfort, and K. Ridderinkhof. 2010. Dissociable brain mechanisms underlying the conscious and unconscious control of behaviour. *Journal of Cognitive Neuroscience* 23: 91–105.

Velmans, M. 2000. *Understanding consciousness*. London: Routledge.

Velleman, J. 1992. What happens when someone acts? *Mind* 101: 461–81.

Wegner, D. 1994. Ironic processes of mental control. *Psychological Review* 101: 34–52.

Wegner, D. 2002. *The Illusion of Conscious Will*. MIT Press.

Wegner, D. and T. Wheatley. 1999. Apparent mental causation: Sources of the experience of will. *American Psychologist* 54: 480–92.

Weiskrantz, L. 1986. *Blindsight: A case study and implications*. Oxford: Oxford University Press.

Wilson, T. 2002. *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: Harvard University Press.