



HAL
open science

Consciousness and Agency

Tim Bayne, Elisabeth Pacherie

► **To cite this version:**

Tim Bayne, Elisabeth Pacherie. Consciousness and Agency. Jens Clausen; Neil Levy. Handbook of Neuroethics, Springer Netherlands, pp.211-230, 2015, 9789400747067. 10.1007/978-94-007-4707-4_117. hal-03878902

HAL Id: hal-03878902

<https://hal.science/hal-03878902>

Submitted on 10 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Draft version. For purposes of quotation please consult the published version:
Bayne, T. & Pacherie, E. (2014). Consciousness and agency. In Jens Clausen & Neil Levy (eds.), *Springer Handbook of Neuroethics*. Dordrecht: Springer, pp. 211-230.

Consciousness and Agency

Tim Bayne
Philosophy, School of Social Sciences
University of Manchester
M13 9PL
United Kingdom
tim.bayne@manchester.ac.uk

&

Elisabeth Pacherie (Institut Jean Nicod, ENS, EHESS, CNRS, Paris)
CNRS UMR 8129
Institut d'Etude de la Cognition
Ecole Normale Supérieure
29, rue d'Ulm
75005 Paris – France
<pacherie@ens.fr>

Abstract

There are three projects within cognitive science of agency and consciousness that are of particular interest to neuroethics: the *descriptive project*, the *genetic project*, and the *substantive project*. The descriptive project is concerned with characterizing our everyday experience of, and beliefs about, agency. What is the folk view of agency? The aim of the genetic project is to give an account of the psychological mechanisms involved in constructing our experience of, and beliefs about, agency. How is the folk view of agency to be explained? The substantive project is concerned with determining the degree to which our experiences of, and beliefs about, agency are correct, and what degree they might need to be revised in light of findings from the cognitive sciences. Is the folk view of agency basically correct or does it need to be modified in fundamental ways (as “will sceptics” argue)? This entry provides an overview of recent research relating to all three projects.

1. Introduction: Three Projects

It is possible to distinguish three distinct, but related, projects in the burgeoning literature on the relationship between agency and consciousness: the *descriptive* project, the *genetic* project, and the *substantive* project. The descriptive project is concerned with charting the contours of what, following Sellars (1962), can be described as the manifest image of human agency. How do we experience our own agency? What is the content of our intuitive, pre-scientific conception of human agency? The aim of the genetic project is to give an account of the psychological mechanisms involved in the construction of the manifest image of human agency. Why do we have the pre-theoretical conception of agency that we do? The focus of the substantive project is that of determining the degree to which the manifest image of agency is correct. To what extent might the folk conception of agency need to be revised in light of what we have learnt from the cognitive sciences?

Although these projects are formally distinct there are many important connections between them. Most obviously, tackling the genetic project presupposes that one has made a certain degree of progress with the descriptive project, for until one has an account (if only partial) of the content of the manifest image of agency one won't be in a position to know quite what it is that the genetic project is meant to explain. There are also important points of contact between the genetic project and the substantive project, as we shall see in due course.

2. The Descriptive Project

The manifest image of agency comprises two kinds of states: agentive experiences and agentive beliefs (Bayne & Pacherie 2007). Agentive experiences involve the awareness of agency as such. Although some agentive experiences concern the agency of others, we will focus here—as the literature does—on agentive experience that concern one's own agency (Horgan et al. 2003; Bayne 2008; Pacherie 2008; Wakefield & Dreyfus 1991). The literature contains a great variety of terminology here, with authors referring to experiences of deliberation, experiences of decision-making, experiences of intentionality, experiences of freedom, experiences of mental causation, the awareness of movement, the awareness of intentions to act, and the sense of control, sense of effort, and sense of agency (among many others).

It is possible to impose some much-needed structure on this discussion by distinguishing two kinds of elements associated with agentive experience: those that are associated with *all* agentive experiences ('core elements') and those that are associated with only *some* agentive experiences ('non-core elements') (Bayne 2010). One core element of agentive awareness is the sense of oneself as acting. Arguably, this experience cannot exist independently of any experience of what it is one is doing—as a feeling simply 'floating in the air' so to speak—but must be accompanied by a sense of *what* it is one is doing, whether that be moving one's body in a certain way (e.g., moving a finger), performing an action of particular kind (e.g., pressing a key), or trying to realize a certain goal (e.g., starting a new paragraph). Having an agentive experience does not require one to identify the

action the experience is about at a particular level of specification, but it does demand that the action be specified in some way, however vague. And if that is right, then agentive experience includes two core elements: a sense of agency (that is, an awareness of oneself as acting) and some specification of the action the sense of agency is directed towards.

What about the non-core elements of agentive experience? Among the various kinds of experiential states associated with agency are experiences of effort, experiences of deliberation, experiences of decision-making, and experiences of freedom. These experiences do not appear to characterize all forms of agency—for example, spontaneously and unreflectively cleaning one's glasses or brushing the hair from one's eyes might not be accompanied by any sense of effort, deliberation, decision-making or freedom—but they do seem to accompany some of our actions. However, giving a precise characterization of the content of these agentive experiences has proven to be very challenging. For example, although many theorists would agree that there is a sense in which we experience ourselves as free, there is little agreement as to what exactly this sense of freedom involves (Nahmias *et al.* 2004).

Let us turn now to the other component of the manifest image of agency: agentive belief. We take this label to include all our pre-theoretical (or 'folk') beliefs about agency, whether they are held explicitly or whether they are merely implicit in our intuitive responses to situations. Examples of such beliefs might include the belief that we possess free will, that consciousness plays an important role in the control of intentional actions, that deliberation typically improves the quality of decision-making, and so on. However, these characterizations of the folk conception of agency are fairly rough and ready, and it would be nice to have a more precise analysis of what the folk conception of agency involves. A great deal of recent work that has been conducted under the 'experimental philosophy' label has attempted to meet this need (see Sommers 2010 for a review). Two issues in particular have been the focus of recent discussion: (1) What is the folk conception of free will; and (2) What is the folk conception of intentional agency?

One of the central debates in the philosophy of action is that between compatibilists and incompatibilists. Compatibilists hold that free will and determinism are compatible with each other, whereas incompatibilists hold that they are incompatible with each other. The question that concerns us here is not how free will *should* be analysed (whatever precisely that might mean), but what conception of free will the folk have. Are the folk compatibilists or incompatibilists?

Attempts to answer this question paint a rather mixed picture (Nichols 2011). Eddy Nahmias and his colleagues have argued that the folk have a predominantly compatibilist conception of free will (Nahmias *et al.* 2005; Turner & Nahmias 2006). They presented undergraduates with vignettes outlining a world in which human agency (along with everything else) was perfectly predictable, and asked them whether free will was possible in such a world. They found that 76% of their subjects claimed that free will was possible in such a world, suggesting that the folk are 'natural compatibilists'.

However, other studies have suggested that the folk have a predominantly incompatibilist conception of free will. In one study, Nichols (2004) presented a group of four and five year olds with scenarios involving physical events and moral choices, and asked them questions on the model of the following:

Moral choice: "If everything in the world was the same right until she chose to steal, did Mary have to choose to steal?"

Physical event: "If everything in the world was the same right until the time that the water boiled, did the water have to boil?"

Nichols found that the children were more likely to say that the physical event had to happen than they were to say that the moral choice events had to happen, and he concluded that the folk are intuitive incompatibilists.

In work designed to examine the tension between these findings, Nichols and Knobe (2007) explored the idea folk responses to question about free will depend on precisely how such questions are framed. Nichols and Knobe described two universes to their subjects, a fully deterministic universe (A) and a universe in which everything except human decisions are deterministically fixed (B). The subjects are first asked which of these two universes most closely resembles are own, with over 90% of subjects choosing universe B. Subjects are then split into two conditions: an abstract condition and a concrete condition. In the former, subjects are asked whether it is possible for someone in universe A to be fully morally responsible for their actions, with 86% of the subjects saying "no". But in the concrete condition, 72% of subjects said that a man in universe A who kills his wife and three children in order to be with his secretary is morally responsible for his actions. In other words, the concreteness (and in particular affective character) of a scenario has an impact on the judgments that subjects make about it.

Just what we should conclude from all of this is, as yet, quite unclear. One possibility is that there is no standard folk conception of free will; instead, significant numbers of folks are compatibilists and significant numbers are incompatibilists; indeed, there is some evidence that individual differences in conception of free will covary with personality differences (Feltz & Cokely 2009). Another possibility is that (most) of us don't have a single unified conception of free will but instead have a few of free will that incorporates both compatibilist and incompatibilist elements. Deciding between these (and other) possibilities will no doubt be on the agenda of experimental philosophy for some time to come.

Let us turn now to the folk conception of intentional action. A central question in the analysis of intentional agency is whether foreseen side effects of an action are ever brought about intentionally. Some philosophers have argued that side effects are never brought about intentionally, whereas others have argued that side effects can be brought about intentionally. In an influential series of studies, Joshua Knobe (2003a; 2003b; see also Malle & Knobe 1997) asked what the folk conception of intentional action is. Knobe discovered that people are much more likely to ascribe intentionality to side effects when those effects are negative (e.g. harming the environment) than when they are positive (e.g. helping the environment). Interesting, the folk conception of intentional agency is very

different from the one which philosophical treatments of intentional agency would have lead one to predict.

3: The Genetic Project

The second of the three projects that we have distinguished is the genetic project: what accounts for the content of the manifest image of agency?

One important question here concerns the relationship between agentic experience and agentic belief. It is plausible to suppose that the content agentic experience plays an important role in determining the content of agentic belief, and that much of what we intuitively believe about agency is grounded in our first-person experiences of agency. On this view, explaining why the manifest image of agency has the content that it does is primarily a matter of explaining why we have the agentic experiences that we do.

This 'empiricist' account of the folk conception of agency is not the only option on the table. Another possibility is that certain aspects of folk beliefs about agency might be grounded in our moral commitments. For example, Nichols (2004) has suggested that the incompatibilist elements in the folk view of free will might derive from perceived constraints on moral responsibility. The idea is that we reason as follows: we are subject to moral requirements, we could be subject to moral requirements only if we possessed incompatibilist free will, so we possess incompatibilist free will. Note that Nichols does not endorse this argument but merely suggests that it might play a role in explaining the widespread appeal of incompatibilism.

Let us leave these questions to one side and turn to the question of how agentic experience itself might be generated, for it is this issue that has received the lion's share of attention in the recent literature. A number of accounts of agentic experience can be distinguished, but they all have a shared commitment to the idea that the sense of agency is produced when there is a match between cues x and y . What distinguishes these accounts from each other is their conception of: (1) the nature of the cues being compared; (2) the nature of the processes involved in the production of the sense of agency; and (3) how closely these processes are related to action production and control processes.

Two theoretical positions define the two ends of the spectrum of possibilities: the motor prediction view and the cognitive reconstruction view. On the motor prediction view, the sense of agency is generated by processes dedicated to action control. On the cognitive reconstruction view, the sense of agency is generated by a general-purpose process of retrospective causal inference.

The motor prediction view is inspired by computational theories of motor control. According to these theories, when the motor system generates a motor command, an efference copy of this command is sent to forward models whose role is to generate predictions about its sensory consequences in advance of actual execution. Error signals arising from the comparison of desired, predicted, and actual states (as estimated from sensory reafferences) are used to make corrections and adjustments. The motor prediction view holds that the signals used for motor control also provide cues to agency (Frith *et al.* 2000). In particular, it holds (1) that awareness of initiating an action is based on a representation of the predicted consequences of making that action, rather than its actual

consequences, and on the congruence between the predicted state and the desired state, and (2) that for this experience of agency to continue, the predicted consequences would also have to remain congruent with the sensory reafferences when they become available.

Claim (1)—and therefore the possibility that the sense of agency can emerge in advance of actual sensory effect and be based on premotor processes alone—is supported by evidence that awareness of initiating a movement in healthy subjects is reported by the agent between 80-200 milliseconds before the movement actually occurs (Libet *et al.* 1983; Libet 1985; Haggard & Eimer 1999). Evidence for claim (2)—that the sense of agency also depends on the congruence between predictions and sensory reafferences—comes from studies where these reafferences are artificially manipulated by introducing temporal delays and spatial distortions of feedback. These studies demonstrate that the sense of agency is gradually reduced as these discrepancies increase (Fournieret & Jeannerod 1998; Knoblich & Kircher 2004; Sato & Yasuda 2005).

In contrast, the cognitive reconstruction view downplays the contribution of the motor system to the sense of agency and proposes that it is inferred retrospectively from the existence of a match between a prior thought and an observed action. Thus, on Wegner's 'theory of apparent mental causation' (Wegner 2002), a general-purpose causal inference process is at play. If an action is consistent with a prior thought of the agent and other potential causes of the action are not present or salient, a sense of agency for the action will be induced.

There is also empirical evidence that high-level inferential processes play a role in determining the sense of agency for an action. Studies of Wegner and colleagues have demonstrated that cognitive cues can alter the sense of agency for an action independently of changes in sensorimotor and perceptual cues. For instance, in their 'I-Spy' study (Wegner & Wheatley 1999), a participant and a confederate of the experimenter had joint control of a computer mouse that could be moved over any one of a number of pictures on a screen. When participants had been primed with the name of an item on which the mouse landed, they expressed a stronger sense of agency for the action of stopping on that object (when in fact the stop had been forced by the confederate). Further studies also suggest that subliminally priming an outcome just before the outcome is produced can enhance the sense of agency for that outcome (Aarts *et al.* 2005) and that priming an outcome relatively far in advance can augment self-agency as well, but only if the outcome is attached to positive affect (Aarts *et al.* 2009).

There is now a growing consensus that the motor prediction view and the cognitive reconstruction view are not mutually exclusive but complementary and that intrinsic cues (cues provided by the motor system) and extrinsic cues (such as cognitive primes) both contribute to the sense of agency (Pacherie 2008; Sato 2009; Synofzik *et al.* 2008; Moore *et al.* 2009; Moore & Fletcher 2012).

Researchers are now trying to develop integrative frameworks and to get a better understanding of how all these agency cues interact.

One way to try and combine the motor prediction view and the cognitive reconstruction view is to appeal to the distinction between pre-reflective agentive experiences and reflective agentive beliefs or judgments (Bayne &

Pacherie 2007; Gallagher 2007; Haggard & Tsakiris 2009) and to argue that while motor processes contribute mainly to feelings of agency, interpretive processes contribute mainly to judgments of agency. This conceptual distinction is echoed methodologically in the ways agency is measured in experimental studies. While some studies (Farrer *et al.* 2003; Metcalfe & Greene, 2007; Sato & Yasuda 2005) investigate agency by asking participants to explicitly judge whether they caused a particular sensory event, other studies use implicit agency measures such as intentional binding and sensory suppression. Intentional binding is a phenomenon, first reported by Haggard and his colleagues (Haggard *et al.* 2002), whereby an action and its external sensory consequences are compressed together in subjective time. As intentional binding occurs only for voluntary actions (Tsakiris & Haggard 2003) and is furthermore modulated by the statistical relation between events (Moore & Haggard 2008), it is considered to provide an implicit measure of agency. Sensory attenuation of self-produced action effects has also been used as an implicit measure of agency. When the internally generated motor predictions about the sensory consequences of one's on-going actions and their actual sensory consequences are congruent, the sensory percept is attenuated, thereby enabling a differentiation between self-generated and externally generated sensory events (Blakemore *et al.* 2002; Cullen 2004). However, recent studies showing that prior authorships beliefs can modulate both sensory attenuation and intentional binding (Desantis *et al.* 2011; Desantis *et al.* 2012), suggest that drawing a sharp distinction between feelings of agency supported by motor processes and judgments of agency supported by interpretive processes may be over-simplistic.

A promising approach is to appeal to a Bayesian integrative framework involving a hierarchy of prediction and model building. Thus, Moore & Fletcher (2012) propose that the sense of agency is determined by a Bayesian process of cue integration, where the predictions generated at higher levels of the hierarchy provide the priors for the lower levels; i.e. constrain the interpretation of cues available at lower levels. In this model, cue integration is itself the product of both the strength of the priors and the weights attached to the available cues as a function of their reliability. When priors are weak—as, for example, when one is quite unsure what the effects of pressing a button will be—one may still have a strong sense of agency for the ensuing consequence, provided that perceptual reafferences carrying information about it are very reliable. Conversely, if my priors are very robust, I may have a strong sense that I produced a certain effect in the world, even though the feedback I get is weak or ambiguous. When both priors and reafferent cues are weak, my sense of agency may be correspondingly weakened. While this Bayesian approach does not allow for a sharp distinction between agentive experiences and agentive judgments, it can accommodate the idea that high-level priors exert more influence on agentive judgments than on agentive experiences.

4. The Substantive Project

We turn now to the third of three projects that we outlined in the introduction: the substantive project. Briefly put, the aim of the substantive project is to determine the degree to which the manifest image of agency is correct. How accurate are our experiences of and beliefs about agency? The substantive project has dominated discussions of the relationship between agency and

consciousness over the last two or so decades. At the centre of these discussions is a position that has been dubbed *will scepticism*. Will sceptics argue that important elements of the manifest image of agency are at odds with the scientific image of agency, and as such should be rejected or at least revised.

The most popular form of argument for will scepticism attempt to put pressure on the folk conception of agency by trying to show that some of its commitments are false. For example, arguments for will scepticism that appeal to Libet's (1985) influential studies on the readiness potential claim that these studies are at odds with the folk commitment to the idea that freely willed actions are initiated by an act of conscious volition.

These arguments share the following 'two-pronged' structure. The first prong involves the claim that the folk conception of agency is committed to ascribing a certain feature ('feature X') to human agency. The second prong involves a claim to the effect that human agency does not in fact possess feature X. Cognitive science is relevant to the evaluation of each of these two prongs. Most obviously, it is relevant to the evaluation of the second prong, for the claim that human action lacks certain features is subject to the tribunal of empirical inquiry. But cognitive science is also relevant to the evaluation of the first prong, for the question of what precisely the folk conception of agency is committed to falls under the purview of the descriptive project, and that—as we have seen—falls within the domain of cognitive science.

The argument for will scepticism that appeal to Libet's experiments regarding free will and the readiness potential (Libet 1985, Libet *et al.* 1983) is one of the most the most widely-discussed in the current literature (see, e.g., Banks & Pockett, 2007; Mele 2009; Nahmias, 2013; Sinnott-Armstrong & Nadel, 2011).. In these experiments, subjects were asked to flex their wrist at will and to note when they felt the urge to move by observing the position of a rapidly rotating dot on a special clock. While subjects were both acting and monitoring their urges (intentions, decisions) to act, Libet used an EEG to record the activity of prefrontal motor areas. On average, participants reported the conscious intention to act, what Libet called the W-judgement, about 200 ms before the onset of muscle activity. By contrast, the EEG revealed that preparatory brain activity, termed by Libet type II readiness potential (RP), preceded action onset by about 550 ms. In other words, their brain started preparing the action at least 350 ms before the participants became aware of the intention to act. In fact, for reasons that we need not explore here, Libet claimed that this gap was likely to be closer to 400 ms in length.

As a number of commentators have pointed out, Libet's paradigm is subject to a number of methodological problems (see e.g. the commentaries on Libet 1985). To take just one example of these problems, Libet's paradigm requires subjects to divide their attention between the position of dot on the clock-face and their own agency. The demand to divide one's attention between two perceptual streams in this way is a notorious source of error in temporal-order judgements. Despite these difficulties, Libet's basic findings have been replicated by a number of laboratories using studies that are free of these methodological difficulties.

Although there is some variability between studies, the claim that 'Libet-actions'—that is, simple and (relatively) spontaneous motor actions—involve an

RP whose onset precedes the time of the subjects' W judgement by about 400 ms or so is largely undisputed. What is in dispute are the implications of these results for questions concerning free will.

Libet denied that his results establish free will scepticism, for he argued that the gap of 150 ms between the agent's conscious decision and the onset of the action allowed for a kind of free will in the form of conscious veto. However, many theorists have seen in Libet's work the death-knell of free will. In their review of his work, Banks and Pocket (2007: 658) describe Libet's experiments as providing "the first direct neurophysiological evidence in support of [the idea that perceived freedom of action is an illusion]."

Unfortunately, few sceptics have said exactly how Libet's data is supposed to undermine free will. Here is one way in which Libet's data might be thought to put pressure on free will:

- (1) The actions studied in the Libet paradigm are not initiated by conscious decisions but are instead initiated by the RP.
- (2) In order to exemplify free will an action must be initiated by a conscious decision.
- (3) So, the actions studied in the Libet paradigm are not freely willed. [From (1) and (2).]
- (4) Actions studied in the Libet paradigm are central exemplars of free will (as intuitively understood), and so if these actions are not freely willed then no (or at least very few) actions are freely willed.
- (5) So no human actions are freely willed. [From (3) and (4).]

We will refer to this as *the sceptical argument*. The sceptical argument is valid, so if it is to be resisted we need to reject one (or more) of its premises. Let us begin by considering (4).

The scope of free will

Are the actions that form the focus of the sceptical argument—'Libet-actions'—paradigm examples of our intuitive notion of free will? Libet himself had no doubts about the answer to this question—he took himself to have studied an "incontrovertible and ideal example of a fully endogeneous and 'freely voluntary' act" (Libet *et al.* 1983: 640)—but not everyone shares this view. Adina Roskies, for example, claims that Libet actions are at best 'degenerate' examples of free will, and suggests that we ought to focus on actions that are grounded in our reasons and motivations if we are interested in "how awareness and action are related insofar as they bear on freedom and responsibility" (2011: 19).

To make progress here we need a taxonomy of action types. One useful distinction is between *automatic* actions and *willed* actions. Automatic actions flow directly from the agent's standing intentions and pre-potent action routines. Many of our everyday actions—washing the dishes, answering the telephone, reaching for a door handle—are automatic. Our awareness of various features of our environment together with over-learned action schemas conspire to trigger the appropriate actions with only the minimal participation of conscious deliberation or decision on the part of the agent. Willed actions, by contrast, require the intervention of executive processes. Some willed actions—what we

call 'disinterested actions'—involve only decision. Consider the experience of finding oneself in a restaurant confronted by a number of equally attractive—or, as the case may be, unattractive—options on the menu. One needs to make a choice, but it does not matter what one orders. Other willed actions—what we call 'deliberation actions'—involve both decision and deliberation. Consider Sartre's case of the young man who must choose whether to look after his aged mother or join the resistance. Here, the function of decision-making is not to select from amongst a range of options between which one is relatively indifferent (as is the case in contexts of disinterested actions), but to draw on one's reasons in making a good decision.

Are Libet-actions automatic or willed? Although they are embedded in a wider agentive context—a context that includes a conscious decision to produce an action of a certain type within a certain temporal window—Libet-actions are not plausibly regarded as automatic. Unlike standard examples of automatic actions, Libet actions are not triggered by an external cue. They may not be the 'ideal examples' of fully spontaneous agency that Libet took them to be, but Libet-actions do seem to be genuine instances of willed agency nonetheless.

But although Libet-actions involve an act of will they do not involve deliberation—at least, not immediately prior to the action. They are 'disinterested' rather than 'deliberative' actions, for the agent has no reason to flex their wrist at one particular time rather than the other, or to flex it in one way rather than another. Indeed, Libet-experiments are explicitly constructed so as to minimize the rational constraints under which the subject acts. We might think of Libet-actions as manifesting the liberty of indifference.

With the foregoing in hand, let us return to the question of whether Libet-actions are paradigms of free will (as we intuitively conceive of it). Are disinterested actions our central exemplars of free will, or does that epithet belong to deliberative actions? Philosophers do not agree on the answer to this question, and the systemic research that would be required in order to settle this dispute has not been carried out. That being said, we suspect that Roskies is right to identify the central or core cases of free will—at least, the kind of free will that is most intimately related to moral agency—with deliberation and rational reflection.

But even though Libet-actions might not be paradigms of free agency, it seems clear that they *do* fall within the scope of our pre-theoretical notion of free will. As such, the free will sceptic is perfectly within his or her rights to claim that if Libet-actions—and indeed disinterested actions more generally—are not free then an important component of our common-sense conception of free will would be threatened. In sum, although (4) is unacceptable as stated, the sceptical argument is not thereby rendered impotent, for the question of whether Libet-actions manifest free will is itself an important one. Libet-actions might not qualify as ideal examples of free will, but they do provide the free will sceptic with a legitimate target.

The initiation of free action

Let us turn now to the second premise of the sceptical argument:

(2) In order to exemplify free will an action must be initiated by a conscious decision.

We can think of (2) as the 'conceptual' step of the sceptical argument, for its plausibility turns chiefly on the contours of our everyday (or 'folk') notion of free will.

But is (2) true? In order to engage with this question, we need to consider what it means for an action to be initiated by a conscious decision. According to one view, an action is initiated by a conscious decision only if it has as its point of origin in a conscious decision that is itself uncaused. Is this how we should understand (2)?

Certain incompatibilists might think so. More to the point, certain kinds of incompatibilists might argue that the *folk* are implicitly committed to this claim, and thus any evidence to suggest that this claim is false would require that our folk conception of free will be modified in some way. However, as we noted in discussing the descriptive project, it is far from clear just what 'the folk' conception of free will is. Although the folk can be prompted to give incompatibilists responses in certain contexts, they can also be prompted to give compatibilists responses in others, and thus it remains an open question just how deeply committed the folk are to incompatibilism.

Let us turn to another reading of (2). One might argue that all it is for an action to be initiated by a conscious decision is for that action to have its point of origin in that decision, without also requiring that that decision is itself an uncaused event. Is (2) plausible even on this weaker reading of it?

Note first that the very idea that an action can always be traced back to a *single* point of origin is open to challenge. Rather than thinking of actions as originating with particular discrete events, we might do better to conceive of them as the outcome of multiple events and standing states, no single one of which qualifies as 'the' point of origin of the action. Just as the Nile has more than one tributary, so too many of our actions might result from multiple sources.

Secondly, to the extent that free actions can be traced back to a point of origin, it is by no means obvious that this point of origin must always be a conscious decision (Levy 2005). Consider a thoughtless comment that is uttered on the spur of the moment and without forethought. Despite the fact that such an utterance is not consciously initiated, one might think that the notion of free will has some kind of grip in such contexts. But, the objection continues, if that is right, then (2) is too demanding: freely willed actions need not be initiated by conscious decisions.

In response to these points, the advocate of the Libet argument might argue that even if it's not an essential feature of all freely willed actions that they have their point of origin in a conscious decision, it is a feature of the kinds of (supposedly free) actions that *Libet* studied. Unlike those actions that we experience as automatic, Libet-actions are accompanied by the 'phenomenology of conscious initiation': one experiences themselves as deciding to act here-and-now. And, the will sceptic might continue, if the neural data demonstrate that the action has been initiated before the agent is aware of their decision then this sense of origination is illusory.

Some authors would take issue with this characterization of the agentive experience that accompanies Libet actions. For example, Terry Horgan (2010) acknowledges that one would experience oneself as beginning to actively undertake an action at some specific moment in time, but he denies that this phenomenology would involve any sense that one's behaviour is caused by one's mental states. Instead, he suggests, one would experience oneself as 'authoring' the behaviour. Horgan's comments raise deep and important issues, but we lack the space to engage with them here. Instead, we will turn to the first premise of the Libet argument and the question of whether Libet actions really are initiated by the readiness potential rather than the agent's conscious decision.

Conscious decisions and the readiness potential

The first premise of the sceptical argument is as follows:

- (1) The actions studied in the Libet paradigm are not initiated by conscious decisions but are instead initiated by the RP.

In order to evaluate (1) we need to consider what it is for an event to initiate an action. Let us say that ε initiates α only if there is a robust correlation between ε -type events and α -type events, such that in normal contexts there is a high probability that an ε -type event will be followed by an α -type event. (The notion of origination clearly requires more than this, but it is implausible to suppose that it requires less than this.) So, if the RP initiates the agent's action, then, we ought to expect RP events to be 'immediately' followed by the appropriate action, unless something unusual happens (such as the person being struck by lightning). Or, to put it the other way round, we should expect that when there is no movement, there is also no RP event. Is this the case?

As several commentators have observed, the back-averaging techniques used to measure RPs do not allow us to answer this question. Because the RP on any one trial is obscured by neural noise, what is presented as 'the RP data' is determined by averaging the data collected on a large number of trials. In order to compute this average, the EEG recordings on different trials need to be aligned, and this requires some fixed point—such as the onset of muscle activity or some other observable behaviour on the part of the subject—that can be identified across trials. This technique has two main drawbacks. First, as Roskies (2011) and Trevena and Miller (2002) note, because it involves averaging across a number of trials certain aspect of the data might be statistical illusions. In other words, features of the relationship between (say) the RP and the W judgement might characterize the averaged data even though they do not characterize any of the individual trials that contribute to that grouped data. Second, because action onset serves as the needed fixed point for the alignment of EEG recording, any RPs that are not followed by an action simply won't be measured, and so we don't know how robust the correlation between the RP and Libet-actions is (Mele 2009).

There are indirect reasons for thinking that the relation between the RP and subsequent action may not be as tight as that which would need in order to say that the RP is the point of origin of the action. Firstly, we know that the nature of the experimental context can significantly affect both the temporal properties and the strength of the RP signal. Subjects who are highly motivated to perform the task produce a large RP, whereas the RP almost disappears in subjects who

have lost interest in the task (McCallum 1988; Deecke *et al.* 1973; see also Rigoni *et al.* 2011). Secondly, it is possible to make willed responses to stimuli in very much less than 550 ms, which indicates that a type II RP is not 'the' point of origin even where it occurs. Thirdly, another neural event—the *lateralized* readiness potential (LRP)—appears to be more strongly coupled to agency than the (generalized) RP is. Whereas the (generalized) RP is symmetrically distributed over both hemispheres, the LRP is restricted to the hemisphere contralateral to the hand that is moved. Haggard and Eimer (1999) found that the LRP was more robustly correlated with the subsequent action than the RP as well as tightly coupled to the W judgments that subjects make. However, a version of the Libet argument in which (1) is replaced with a corresponding claim about the LRP does not possess even the surface plausibility that (1) does. (Note, however, that a recent study by Schlegel *et al.* (2013) failed to replicate Haggard and Eimer's finding, and found no within-subject covariation between LRP onset and W judgment, leading them to conclude that neither RP onset nor LRP onset cause W.)

In a recent experiment, Schurger and colleagues (2012) used a modified Libet task to circumvent the limitations due to back-averaging techniques. Their aim was to test the proposal that RPs correlate with predecision activity rather than with activity which coincides with, or is subsequent to, the agent's decision (as Libet thought). Schurger and colleagues proceeded by assuming that the decisions of the participants in Libet's experiment can be modelled—as neural decision tasks typically are—in terms of an accumulator-plus-threshold mechanism: decisions are made when relevant evidence accumulated over time reaches a threshold. What is unique to Libet's task is that subjects are explicitly instructed not to base their decision on any specific evidence. Schurger and colleagues propose that the motor system constantly undergoes random fluctuations of RPs and that this random premotor activity is used as a substitute for actual evidence. According to their stochastic decision model, the decision process, given Libet's instructions, amounts to simply shifting premotor activation up closer to the threshold for initiation of the movement and waiting for a random threshold-crossing fluctuation in RP. Time-locking to movement onset ensures that these fluctuations appear in the average as a gradual increase of neuronal activity, when in fact what is measured are simply random fluctuations of RPs that happened to cross a decision threshold.

Thus the two models predict the same premotor activation buildup when a movement is produced, but whereas on Libet's postdecision interpretation of this buildup there should be no premotor activity (and hence no RPs) when no movement is produced, on the predecision interpretation there should be continuous random fluctuations in RPs even when no movement is produced. Schurger and colleagues reasoned that it should be possible to capture these fluctuations by interrupting subjects in a Libet task with a compulsory response cue and sorting trials by their reaction times. On the assumption that the interrupted responses arise from the same decision accumulator as the self-initiated ones, response times should be shorter in trials in which the spontaneous fluctuations of RPs happened to be already close to threshold at the time of the interruption. On the assumption that close to threshold activity reflects spontaneous fluctuations of RPs rather than mounting preparation to

move building over the course of the entire trial, slow and fast reaction times should be distributed equally across time within trials. To test these predictions, they therefore devised what they called a Libet's Interruptus task, where they added random interruptions to trials. They found, as they had predicted, that slow and fast responses to interruptions were distributed equally throughout the time span of the trial.

According to the predecision model, Libet's contention that the neural decision to move happens much before we are aware of an intention or urge to move is unfounded. The neural decision to move isn't made when a RP starts building up, since spontaneous fluctuations of RPs happen all the time but when a random fluctuation in RP crosses a threshold. The reason we do not experience the urge to move earlier is simply that the decision threshold has not yet been crossed and thus the decision has not yet been made. While Schurger and colleagues take no stand on the exact temporal relation between the conscious urge to move and the neural decision to move, their results cast serious doubt on Libet's claim that the neural decision to move coincides with the onset of the RP and thus on his further claim that since RP onset precedes the urge to move by 350 ms or more, conscious intentions play no role in the initiation of the movement. If instead the neural decision to move coincides with a much later threshold-crossing event, it remains at least an open possibility that this event coincides with and constitutes the neural basis of a conscious urge to move. In any case, Schurger and colleagues also insist that this threshold-crossing event should not be interpreted as *the* cause of the movement but rather as simply one of the many factors involved in the causation of self-initiated movements.

Taken together, these points suggest that the RP is unlikely to qualify as 'the' point of origin of the action. If the RP has a psychological interpretation—and it is far from clear that it does—then we should perhaps think of it as the neural correlate of an 'urge' or 'inclination' to act, rather than as the neural basis of the decision to act now (Gomes 1999; Mele 2009). The RP may be one of the many tributaries that contribute to an action, but it is not its 'origin' in any intuitive sense of that term.

Other sources of will scepticism

In a series of papers and most influentially in his book *The Illusion of Conscious Will*, Daniel Wegner has argued that central components of the folk psychological conception of agency are inaccurate and should be jettisoned. As he puts it, the conscious will 'is an illusion'. Precisely what Wegner means by describing the conscious will as an illusion is open to some debate (Bayne 2006; Mele 2009; Nahmias 2002), but we take his central claim to be this: agentic experience misrepresents the causal path by means of which one's own actions are generated.

One reason that Wegner gives for thinking that the conscious will 'is an illusion' involves the idea that agentic experiences are theoretically mediated. As he puts it, "[Conscious will is an illusion] in the sense that the experience of consciously willing an action is not the direct indication that the conscious thought has caused the action." (Wegner 2002: 2) As we have seen, there is very good reason to think that agentic experiences are theoretically mediated, but it

is difficult to see why will scepticism should follow from this. Even if the folk are intuitively inclined to think that our access to our own agency is direct and unmediated—and we're far from certain that such a view *is* part of the folk conception of agency—there is little reason to think that such a view is part of the core conception of agency.

Another sense in which one might regard the conscious will as illusory involves the idea that experiences of doing are systematically, or at least frequently, *non-veridical*: Experiences of doing misrepresent our agency and the structure of our actions. This seems to be Wegner's main line of argument for will scepticism, and *The Illusion of Conscious Will* contains extensive discussion of dissociations between the exercise of agency and the phenomenology of agency. Some of these cases appear to demonstrate that we can experience ourselves as doing something that someone else is doing (and that we are not). Wegner calls such cases *illusions of control*. The I-Spy experiment (discussed earlier) is an example of an illusion of control. Other dissociations involve experiencing someone (or something) else as the agent of what one is doing. Wegner calls such phenomena *illusions of action projection*. Among the most fascinating of the various illusions of action projection that he discusses is facilitated communication, a practice that was introduced as a technique for helping individuals with communication difficulties. Facilitators would rest their hands on the hands of their clients as the client typed a message. Although the facilitators experienced no sense of authorship towards the typed message, there is ample evidence that the content of "facilitated" messages derived from the facilitator rather than the client (Wegner et al. 2003).

How might these dissociations support the case for will scepticism? On one reading of his argument, Wegner is mounting an inductive generalization: since some experiences of conscious will are non-veridical it is reasonable to infer that most, and perhaps even all, such experiences are. But this argument seems weak, for the fact that experiences of agency *can* be non-veridical shows that the mechanisms responsible for generating such experiences are *fallible* does not show that they are *unreliable*. Another way to read the argument from dissociations is as an inference to the best explanation. The argument proceeds as follows: Since the phenomenology of agency plays no direct role in the genesis of action where such experiences are absent, we have good reason to think that it plays no direct role in the genesis of action when such experiences are present. As Wegner himself puts it, "If conscious will is illusory, automatism are somehow the 'real thing', fundamental mechanisms of mind that are left over once the illusion has been stripped away. Rather than conscious will being the rule and automatism the exception, the opposite may be true." (2002: 143)

What does it mean to say that automatism are the *fundamental* mechanisms of mind? To the extent that automatism are action-generation procedures that do not involve intentional states of any kind then there may be a tension between automaticity and the experience of conscious will, but Wegner provides little evidence for the view that our actions are usually automatic in this sense of the term. If, on the other hand, automatism are action-generating procedures that are non-consciously initiated then there is ample reason to describe much of what we do as automatic in nature. But on this conception of an automatism there is no conflict between automaticity and the feeling of doing. So there is no

argument from automaticity (thus conceived) to the claim that the experience of conscious will is an illusion.

We do not deny that the phenomenology of agency *can* be illusory. Consider, for example, the experience of intentionality. An experience of intentionality will be non-veridical if the action in question is not guided by an intention, or if it is guided by an intention other than the one that it seems to have been produced by. The phenomenon of confabulation suggests that at least one if not both of these conditions occur. But I think that there is little reason to assume that either kind of mistake is at all common in everyday life. Confabulation is striking precisely because it is unusual.

In fact, Wegner's own account of the genesis of the experience of doing suggests that such experiences will normally be veridical. According to the matching model, we experience ourselves as doing X when we are aware of our intention to X as being immediately prior to our X-ing, and when we are not aware of any rival causes of our X-ing. Now, if we experience ourselves as having an intention to X then it probably is the case that we do have the intention to X. (After all, it seems reasonable to suppose that introspection is generally reliable. At any rate, Wegner is in not position to challenge the reliability of introspection, for he himself assume the reliability of introspection insofar as he takes subjects to be reliable in reporting their experiences of conscious will.) But if one has an intention to X, and if one has in fact X-ed, and if one's intention to X is immediately prior to one's X-ing, then it is highly likely that one's intention to X is involved in bringing about one's X-ing. It would be remarkable if one had an intention to raise one's hand just prior to raising one's hand but the intention played no causal role in the raising of one's hand. Far from showing that experiences of agency are illusory, Wegner's own model of how such experiences are generated predicts that they will normally be veridical.

Conclusion

Discussions of human agency in the last decades have been dominated by the issue of whether the scientific image of human agency undermines the manifest image of human agency, and in particular its commitment to free will. For the most part, cognitive scientists have argued that the folk view of agency is undermined by cognitive science, whereas philosophers have generally denied that there is any such tension, either on the grounds that the scientists in question have misinterpreted the neuroscientific findings or on the grounds that they have assumed a tendentious account of the folk conception of agency.

Although neither camp can claim a decisive victory in this debate, it has prompted theorists to take a closer and more nuanced look at both agentic belief and experience, showing them to be more subtle and less monolithic than previously thought. It is also clear that both elements of the manifest image of agency have complex aetiologies, and that the sense of agency in particular depends on the integration of multiple agency cues. Finally, recent research in the neuroscience of decision and action control reveals both the multilevel structure of these processes and the flexibility of their organization.

These advances suggest that we should resist the temptation to think in dichotomous terms and that we may need to replace such questions as whether we have free will, whether we are natural incompatibilists, and whether our

sense of agency is veridical with questions that do not lend themselves to yes/no answers. As Roskies (2010) notes, it is also clear that the time is ripe for philosophers, cognitive scientists and neuroscientists to engage in more constructive endeavours and use all of the tools at their disposal to advance a positive conception of human agency, one that does justice both to its strengths and to its limitations.

References

- Aarts, H., Custers, R., & Marien, H. (2009). Priming and authorship ascription: When nonconscious goals turn into conscious experiences of self-agency. *Journal of Personality and Social Psychology*, 96, 967-979.
- Aarts, H., Custers, R., & Wegner, D. (2005). On the inference of personal authorship: Enhancing experienced agency by priming effect information. *Consciousness and Cognition*, 14 (3): 439-458.
- Banks, B., & Pockett, S. (2007). Benjamin Libet's work on the neuroscience of free will. In M. Velmans & S. Schneider (Eds.), *The Blackwell Companion to Consciousness*. Blackwell.
- Bayne, T. (2006). Phenomenology and the feeling of doing: Wegner on the conscious will, In S. Pockett, W. P. Banks and S. Gallagher (eds.) *Does Consciousness Cause Behavior?* Cambridge, MA: MIT Press, pp. 169-186.
- Bayne, T. (2008). The phenomenology of agency, *Philosophy Compass*, 3: 1-21.
- Bayne, T. (2011). Libet and the case for free will scepticism. In R. Swinburne (ed.) *Free Will and Modern Science*. Oxford: Oxford University Press, pp. 25-46.
- Bayne, T. & Pacherie, E. (2007). Narrators and comparators: The architecture of agentive self-awareness, *Synthese*, 159: 475-91.
- Blakemore, S.-J., Wolpert, D.M., & Frith, C.D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Sciences*, 6(6): 237-242.
- Cullen, K.E. (2004). Sensory signals during active versus passive movement. *Current Opinion in Neurobiology*, 14: 698-706.
- Deecke, L., Becker, W., Grözing, B., Scheid, P. & Kornhuber, H.H. (1973). Human brain potentials preceding voluntary limb movements. In W.C. McCallum & J.R. Knott (eds.) *Electroencephalography and Clinical Neurophysiological Supplement: Event-related Slow Potentials of the Brain: Their Relations to Behavior* (Vol. 33), Elsevier: Amsterdam, pp. 87-94.
- Desantis, A., Roussel, C., & Waszak, F. (2011). On the influence of causal beliefs on the feeling of agency. *Consciousness and Cognition*, 20(4): 1211-1220.

- Desantis, A., Weiss, C., Schutz-Bosbach, S., & Waszak, F. (2012). Believing and Perceiving: authorship belief modulates sensory attenuation. *PLoS ONE*, 7(5): e37959.
- Farrer, C., Franck, N., Georgieff, N., Frith, C.D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of agency: A positron emission tomography study. *Neuroimage*, 18: 324–333.
- Feltz, A. & Cokely, E. (2008). Do judgments about freedom and responsibility depend on who you are?: Personality differences in intuitions about compatibilism and incompatibilism. *Consciousness and Cognition*, 18: 342-50.
- Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, 36(11): 1133-1140.
- Frith, C., Blakemore, S., & Wolpert, D. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society B*, 355(1404): 1771-1788.
- Gallagher, S. (2007). The natural philosophy of agency. *Philosophy Compass*, 2(2): 347-357.
- Gomes, G. (1999). Volition and the readiness potential. *Journal of Consciousness Studies*, 6 (8-9): 59-76.
- Haggard, P., Clark, S., & Kalogeras, J. 2002. Voluntary action and conscious awareness. *Nature Neuroscience*, 5(4): 382-385.
- Haggard, P. & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research*, 126: 128-133.
- Haggard, P. & Tsakiris, M. (2009). The experience of agency: feeling, judgment and responsibility. *Current Directions in Psychological Science*, 18, 4: 242-246.
- Horgan, T. (2010). The phenomenology of agency and the Libet results. In W. Sinnott-Armstrong & L. Nadel (Eds) *Conscious Will and Responsibility: A Tribute to Benjamin Libet*. New York: Oxford University Press.
- Horgan, T., Tienson, J., & Graham, G. (2003). The phenomenology of first-person agency. In S. Walter and H.-D. Heckmann (Eds) *Physicalism and Mental Causation: The Metaphysics of Mind and Action*. Exeter, UK: Imprint Academic, pp. 323-40.
- Knobe, J. (2003a). Intentional action and side effects in ordinary language. *Analysis*, 63, 190-193.
- Knobe, J. (2003b). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*, 16: 309-324.

- Knoblich, G., & Kircher, T. T. J. (2004). Deceiving oneself about being in control: Conscious detection of changes in visuomotor coupling. *Journal of Experimental Psychology-Human Perception and Performance*, 30(4): 657-666.
- Levy, N. (2005). Libet's impossible demand. *Journal of Consciousness Studies*, 12: 67-76.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, 8: 529-566.
- 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). The unconscious initiation of a freely voluntary act. *Brain*, 106: 623-642.
- Malle, B. & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, 33: 101-121.
- McCallum, W.C. (1988). Potentials related to expectancy, preparation and motor activity. In T.W. Picton (ed.) *EEG Handbook Volume 3: Human Event-Related Potentials*. Amsterdam: Elsevier, pp. 427-534.
- Mele, A. (2009). *Effective Intentions: The Power of Conscious Will*. New York: Oxford University Press.
- Metcalfe, J., & Greene, M.J. (2007). Metacognition of agency. *Journal of Experimental Psychology: General*, 136(2): 184-199.
- Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: A review of cue integration approaches, *Consciousness and Cognition*, 21(1): 68-59.
- Moore, J.W., & Haggard, P. (2008). Awareness of action: Inference and prediction, *Consciousness and Cognition*, 17(1): 136-144.
- Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues, *Consciousness and Cognition*, 18: 1056-64.
- Nahmias, E. (2002). When consciousness matters: A critical review of Daniel Wegner's *The Illusion of Conscious Will*, *Philosophical Psychology*, 15/4: 527-41.
- Nahmias, E. (2013). Is free will an illusion? Confronting challenges from the modern mind sciences. In W. Sinnott-Armstrong (ed.) *Moral Psychology, Volume 4: Freedom and Responsibility*. Cambridge, MA: MIT Press.
- Nahmias, E., Morris, S. Nadelhoffer, T. & Turner, J. (2004). The phenomenology of free will, *Journal of Consciousness Studies*, 11(7-8): 162-79.
- Nahmias, E., Morris, S., Nadelhoffer, T. & Turner, J. (2005). Surveying freedom: Folk intuitions about free will and moral responsibility, *Philosophical Psychology*, 18(5): 561-584.
- Nichols, S. (2004). The folk psychology of free will: Fits and starts, *Mind and Language*, 19: 473-502.

Nichols, S. (2011). Experimental philosophy and the problem of free will. *Science*, 331(6023): 1401-3.

Nichols, S. and Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, 41, 663-685.

Pacherie, E. (2008). The phenomenology of action: A conceptual framework, *Cognition*, 107(1): 179-217.

Pockett, S. & Purdy, S. (2011). Are voluntary movements initiated preconsciously? The relationships between readiness potentials, urges, and decisions. In W. Sinnott-Armstrong and L. Nadel (Eds.) *Conscious Will and Responsibility*. New York: Oxford University Press, pp. 34-46.

Rigoni D., Kühn S., Sartori G., Brass, M. (2011). Inducing disbelief in free will alters brain correlates of preconscious motor preparation: the brain minds whether we believe in free will or not. *Psychological Science*, 22/5: 613-8.

Roskies, A. (2010). How does neuroscience affect our conception of volition? *Annual Review of Neuroscience*, 33: 109-30.

Roskies, A. (2011). Why Libet's studies don't pose a threat to free will. In W. Sinnott-Armstrong and L. Nadel (eds.) *Conscious Will and Responsibility*. New York: Oxford University Press, pp. 11-22.

Sato, A. (2009). Both motor prediction and conceptual congruency between preview and action-effect contribute to explicit judgment of agency. *Cognition*, 110(1), 74-83.

Sato, A., & Yasuda, A. (2005). Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*, 94(3), 241-255.

Schurger, A., Sitt, J. & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement, PNAS.

Sellars, W. (1962). Philosophy and the scientific image of man. In R. Colodny (ed.) *Frontiers of Science and Philosophy*. Pittsburgh, PA: University of Pittsburgh Press, pp. 35-78.

Sommers, T. (2010). Experimental philosophy and free will. *Philosophy Compass*, 5/2: 199-212.

Synofzik, M., Vosgerau, G. & Newen, A. (2008). Beyond the comparator model: A multi-factorial two-step account of agency. *Consciousness and Cognition*, 17 (1): 219-239.

Trevena, J.A. and Miller, J. (2002). Cortical movement preparation before and after a conscious decision to move, *Consciousness and Cognition*, 11: 162-90.

Tsakiris, E., Haggard, P. (2003). Awareness of somatic events associated with a voluntary action. *Experimental Brain Research*, 149(4): 439-446.

Turner, J. & Nahmias, E. (2006). Are the folk agent-causationists? *Mind and Language*, 21(5): 597-609.

Wakefield, J. & Dreyfus, H. (1991). Intentionality and the phenomenology of action. In E. Lepore and R. van Gulick (Eds) *John Searle and His Critics*. Oxford: Blackwell, pp. 259-70.

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: 480-491.

Wegner, D.M., Fuller, V.A., Sparrow, B. (2003). Clever hands: uncontrolled intelligence in facilitated communication. *Journal of Personality and Social Psychology*, 85 (1): 5-19.

Cross-references

Experimental Philosophy and Free Will
Freewill, Agency and the Brain
Mental Causation
Neuroscience, Free Will and Responsibility

Keywords

Agentive experience
Agentive belief
Free Will
Compatibilism
Libet
Incompatibilism
Experimental Philosophy
Wegner
Will scepticism