

Can conscious agency be saved?

Elisabeth Pacherie

► **To cite this version:**

Elisabeth Pacherie. Can conscious agency be saved?. Topoi, Springer Verlag, 2013, pp.1-13. <10.1007/s11245-013-9187-6>. <ijn_00872165>

HAL Id: ijn_00872165

https://jeannicod.ccsd.cnrs.fr/ijn_00872165

Submitted on 11 Oct 2013

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.

Can conscious agency be saved?

Elisabeth Pacherie
Institut Jean Nicod
ENS, EHESS, CNRS, Paris

Can conscious agency be saved?

Abstract

This paper is concerned with the role of conscious agency in human action. On a folk-psychological view of the structure of agency, intentions, conceived as conscious mental states, are the causes of actions. In the last decades, the development of new psychological and neuroscientific methods has made conscious agency an object of empirical investigation and yielded results that challenge the received wisdom. Most famously, the results of Libet's studies on the 'readiness potential' have been interpreted by many as evidence in favor of a skeptical attitude towards conscious agency. It is questionable, however, whether action initiation should be regarded as the touchstone of conscious agency. I shall argue that the traditional folk-psychological view, but also some of the objections leveled against it, rest in part on an over-simplified conception of the structure of agency, that neglects both the role of control processes after action initiation and the role of planning processes before action initiation. Taking these processes into account can lead to a reassessment of the relation between intentions and action and of the role of conscious agency in action production.

Keywords: conscious agency; intentions; action initiation; action control; action planning; Libet.

1. Introduction

Libet's studies on the timing of conscious intentions have been hailed in some quarters as one of the most important psychological experiments ever. Libet's results were claimed to undermine folk-psychological conceptions and traditional philosophical views of free will, of the role of conscious agency in human behavior and of conscious mental causation more generally. In response to these attacks, a number of philosophers have countered that these skeptical claims, at least in their more radical forms, were largely overblown, being both ill supported by the empirical data and premised on dubious metaphysical assumptions regarding the nature of free will and the nature of the relation between the mental and the physical.

Thirty years after its inception, the lively debate started by Libet's studies still hasn't lost steam. On the one hand, additional empirical findings have further highlighted the role of unconscious processes in action production and thus tended to further fuel what Bayne and Levy (2006) aptly termed 'will-skepticism'. On the other hand, both philosophers and cognitive scientists have contributed to the development of more sophisticated models of the structure of human agency, emphasizing both the hierarchical nature of intentions and action representations and the crucial importance of control and monitoring processes in action production.

In this paper, I shall concentrate on the role of conscious agency in action production. I shall argue that taking into account the hierarchical nature of intentions and control processes can lead to a reassessment of the relation between intentions and action and of the role of conscious agency in action production. Apart from a brief discussion in the next section of why Libet thought his results should undermine beliefs in free will and why others have resisted this conclusion, I will set aside in what follows issues regarding free will. Importantly, I will also keep away from issues in the metaphysics of consciousness. If we accept Chalmers' distinction between easy and hard problems of consciousness, the problem I am concerned with falls squarely on the easy side of the divide. As Chalmers (1995) explains, the easy problems of consciousness are easy because they concern the explanation of cognitive abilities and functions, explanations that can be given by specifying

a mechanism that can perform the function, a task to which the methods of cognitive science are well suited.

In my view, the specific challenge posed by Libet's results and other recent empirical results is not, as is often assumed, a metaphysical challenge. Rather, these results seem to cast doubt on traditional functional characterizations of conscious intentions by providing evidence that the roles these accounts assign to conscious intentions can be fulfilled by unconscious states and processes. The risk then is not so much that conscious intentions be cast aside as being metaphysically untenable than as being useless.

Some have tried to save conscious intentions from functional obsolescence by arguing that their primary function is not pragmatic but epistemic (i.e., that their role is to contribute to self-awareness rather than to action production).¹ While I certainly agree that conscious intentions can serve epistemic functions, I don't think we should give up on intentions' pragmatic functions too easily. But to see conscious agency in action, one should look for it in the right places and these are not always the most obvious ones or those most brightly lit up by the lampposts of either philosophy or empirical research. In search of a function for conscious intentions, philosophers and cognitive scientists have largely concentrated on action initiation. If, as causal theories of actions have claimed, the role of intentions is to cause action, it was indeed not unreasonable to expect this to be where to look to see conscious intentions at work. These expectations seem to have met with disappointment. What I propose to do here is to look at what happens both before action initiation and after action initiation.

I will start in section 2 with a brief review of Libet's experiments, the ways in which their results have been interpreted and the debates these interpretations have provoked. In section 3, I shall concentrate on what happens after action initiation. I will discuss Frankfurt's early critique of the standard causal theory of action and his proposal that the essential function of intentions is not so much action initiation as action guidance. I will also consider how empirical work on motor control and the forms it can take can help us both flesh out

¹ This possibility is discussed by Velleman (2007), who does not really endorse it but rather uses it as an antidote to the assumption that the *raison d'être* of intentions is to support pragmatic functions. Dennett (1992, 2003) and Wegner (2002) seem closer to endorsing this view.

this proposal and assess more precisely the respective roles of automatic and conscious processes of action control and their complementary. In section 4, I will be concerned with what happens before action initiation, when prospective intentions are formed and advance planning intervenes in order to flesh out these prospective intentions to the point where they can eventuate into corresponding actions. In particular, I will distinguish two main ways in which advance planning can proceed. I will also discuss the implications these different modes of action planning can have for the role of conscious intentions at the moment of action initiation.

2. Action initiation

In Libet's famous studies on the 'readiness potential' (Libet *et al.* 1983; Libet 1985), 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 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.

Libet and others have claimed that these results provide evidence in favor of a skeptical attitude towards freely willed actions. While Libet himself fell short of endorsing full-blown will-skepticism and suggested that in the interval separating the agent's conscious decision from the onset of the action a kind of free will could still intervene in the form of conscious veto of the unconsciously initiated action, many have been more radical and have taken his work as a death blow to free will.²

It is far from completely obvious, however, why the fact, if it is one, that conscious intentions are reliably preceded by unconscious brain activity in the form of a readiness

² Examples of free will skepticism that appeal to Libet's work include Banks and Isham (2011), Hallett (2007), Pockett (2004), Roediger et al. (2008), Spence (2009) and Wegner (2002).

potential should lead us to conclude that our actions are not freely willed. It seems that one or several further premises are needed to reach this conclusion. One of the main lines of objection to the will-skeptic interpretations of Libet's results concern the conception(s) of free will these interpretations must work with. Libet and others sometimes make it sound as though the fact that the volitional process is initiated by unconscious brain activity is enough to undermine free will. This suggests a conception of free will requiring that the conscious decision or intention to act be the ultimate cause of the action, that is, a cause that is itself uncaused. However, as several critiques have pointed out (Bayne 2011; Mele 2007), the notion of an uncaused cause is, to put it euphemistically, hard to square with a naturalistic stance. Only hardcore mind-body dualists would contend that conscious intentions do not arise as a result of brain activity.

It can also be argued that what makes Libet's data a threat to free willed action is not the mere fact that unconscious brain activity contributes to the arising of conscious intentions, but the fact that these conscious intentions are fully determined by this unconscious brain activity. Here the argument would rest on a conception of free will according to which free will is incompatible with determinism. However, incompatibilism isn't the only option. Indeed, the majority of contemporary philosophers are compatibilists,³ that is, they believe that even if determinism were true, we would still be able to act freely.

Another route to will-skepticism involves interpreting Libet's data as showing that the RP is causing two independent effects: on the one hand it causes the action itself; on the other it causes the conscious experience of intending. On this reading, the conscious intention isn't an intermediary step in the causal path leading from unconscious brain activity to action but is actually part of a different causal path. On this interpretation of Libet's data, conscious intentions do not causally contribute to action initiation and in thinking that they do, we fall prey to an illusion.⁴ This route to will-skepticism is

³ Contemporary defenses of compatibilism include, among many others, Frankfurt (1988), McKenna (2005), Scanlon (2008) and Smith (2003).

⁴ This is the interpretation of Libet's data Wegner appears to favor: "The position of conscious will in the time line suggests perhaps that the experience of will is a link in a causal chain leading to action, but in fact it might not even be that. It might just be a loose end—one of those things, like the action, that is caused by prior brain and mental events" (Wegner 2002: 55). On Wegner's account of how the experience of conscious will is

metaphysically less loaded than the previous ones, since it does not require one to subscribe to substance dualism, to accept the truth of determinism or to take an incompatibilist stance on free will. It simply takes conscious intentions to be epiphenomenal with respect to action initiation, because they lie on a different causal pathway.⁵ However, one may wonder what warrants this interpretation of Libet's data. These data simply give us a time line with unconscious brain activity (RP) coming first, then conscious intentions (W) and finally action onset (A). This time line is equally compatible with two causal stories, one on which unconscious brain activity causes conscious intentions that in turn more proximally cause action onset and the other on which unconscious brain activity causes both conscious intentions and action initiation as independent effects. To follow this route to will skepticism, we would need evidence that the second story is the correct one, but Libet's data themselves simply do not provide this evidence.

However, what many philosophers have taken to be the central worry raised by Libet's data is that if consciousness lags behind decision making, then our deliberative activity is merely epiphenomenal. If we take the exercise of deliberation to be the hallmark of freely willed actions, the worry raised by Libet's data is that the conscious decision to act is arrived at through unconscious brain processes rather than through conscious deliberative processes. How serious this threat is depends, first, on how robust Libet's data are and,

generated, what he calls the theory of apparent mental causation, conscious will is experienced when we infer that our thought has caused our action and that we draw such an inference when we have thoughts that occur just before the actions, when these thoughts are consistent with the actions, and when other potential causes of the actions are not present. Actually, however, our actions spring from unconscious causal processes and the conscious ideas that we mistakenly experience as their causes are themselves caused by unconscious processes whose links to the unconscious processes causing the action, when they exist, are often at best indirect.

⁵ This "local" epiphenomenalism should be distinguished from global epiphenomenalist claims stemming from the causal exclusion problem, where the argument is that the causal efficacy of mental properties is excluded by the causal efficacy of the physical properties on which they supervene (Kim 1993). This challenge applies to mental causation in general and is much broader than the challenge raised by Libet's data. The causal exclusion problem has given rise to a huge literature and there is no clear consensus as to how it can be solved. However, I won't engage with this debate here; rather my focus here will be on Libet's more specific challenge.

second, on whether the processes he studies and the time-window he focuses on are the only relevant ones.

One group of objections raised against Libet's experiments have questioned the empirical validity of their results. Do his data really show that the actions he studied were initiated not by conscious decisions but by the RP instead? Some of these objections have targeted the methodology of the experiment (see, e.g. the commentaries on Libet 1985). One line of objection emphasizes potential inaccuracies and biases in the measurement of the onset of conscious experience and of the onset of movement (Glynn 1990; Gomes 1998, 1999; Joordens, van Duijn, & Spalek 2002; Miller *et al.* 2010; van de Grind 2002). For instance, in a recent study, Miller and colleagues (2011) found that clock monitoring had an effect on electroencephalographic activity. The participants' task was to make spontaneous key-press movements and in one condition they also had to report the position of the dot on a clock when they decided to move. Average EEG activity preceding key presses was substantially different when participants had to monitor the clock than when they did not. The authors suggest that the effects of clock monitoring on EEG activity could be responsible for Libet's findings that movement-related brain activity begins before participants have consciously decided to move.

Another important line of objection, raised for instance by Mele (2009) and Roskies (2011), points to a fundamental limitation of the back-averaging techniques used to extract signals from EEG recordings. Because EEG recordings are typically noisy, to extract signals from noise, one has to average 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 that can be identified across trials. In Libet's experiment, to extract the RP data, EEG recordings had to be aligned using the onset of muscular activity as a fixed point. As a result, RPs that are not followed by an action would not be measured. This means that we don't know whether the correlation between the RP and Libet-actions is sufficiently robust to make it plausible that the RP is what initiate the action.⁶

⁶ In addition, even limiting ourselves to cases where RPs are followed by actions, RPs and W-judgments do not appear to be sufficiently coupled to warrant a causal claim. Haggard and Eimer (1999) found no correlation between the onset of the RP and W-judgments, but did find a positive correlation between the onset of a later phase of the readiness potential, the lateralized readiness potential (LRP), and W-judgments, suggesting that the RP does not

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, as Libet proposed, with activity that coincides with, or is subsequent to, the agent's decision. Schurger and colleagues proceeded on the assumption 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 certain 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 Schurger and colleagues predict the same premotor activation buildup that Libet does when a movement is produced. However, 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 their 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

reflect processes causal of W but that the LRP might. However, a recent study by Schlegel et al. (2013) failed to replicate their results 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.

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 *Libetus 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.

These considerations have given heart to philosophers hoping to preserve a role for conscious intentions in the initiation of action. They suggest that the correlations between RPs and actions are not robust enough to warrant the claim that the apparition of a RP 550 ms before action onset corresponds to the brain decision to initiate an action. According to Mele (2007), it might be more credible to associate RPs with desires, wants, wishes or urges than with intentions or decisions. This would leave it open that at -550 ms, the person instead acquires an unconscious urge or desire and that what happens at -200 ms is that the person becomes conscious of this desire or urge. If so, no one should be surprised to find that desires or urges precede conscious intentions. In addition, finding that we have such desires does

not commit us to acting upon them. For all Libet has shown, it might still be that a further conscious decision is necessary for action initiation.⁷

Philosophers have also questioned whether the kind of cases Libet's experiments concentrate on are really the best suited to an investigation of the role conscious agency could play in action production. For largely methodological reasons, Libet's and other neuroscientific work on intentions and voluntary actions has tended to focus on very short time scales, typically limited to seconds before action onset, and thus on immediate intentions or intentions to act "now", as opposed to prospective intentions or intentions to do something in the future.

One line of criticism, voiced notably by Flanagan (1996) and Gallagher (2006), is that it is misguided to attempt to frame the question of free will at the time scale and in terms of the short term motor intentions and control processes Libet considers. Rather, free will involves temporally extended deliberative processes and applies to intentional actions considered at levels of descriptions typically higher and more abstract than descriptions in terms of motor processes and bodily movements. Leaving to the side the issue of free will, a number of philosophers (Nahmias 2002; Pacherie 2006; Searle 2001; Zhu 2003) have argued that the exercise of conscious agency isn't in principle limited to action initiation. Even if it were true that the subjects in Libet's experiment initiated their movements when they did as a result of unconscious decisions, it is quite implausible that the participants in his studies would have produced these movements unless they had intended to participate in the

⁷ One can offer a similar deflationary interpretation of a more recent experiment (Soon, Brass, Heinze, & Haynes 2008) that found that while subjects who had to decide between two actions reported having made a conscious decision on average 1000 ms before action onset, the outcome of their decision could be predicted from brain activity in prefrontal and parietal cortex up to 10 s before it entered awareness. These results have been claimed to show that decisions about what to do and not just about when to act are made unconsciously. However, in view of the fact that the accuracy with which the conscious decisions could be predicted from prior unconscious brain activity in these areas was less than 60% (50% corresponding to chance), a more reasonable conclusion might be that the decision was causally influenced to some degree by these unconscious brain processes rather than determined by them (see, for instance, section 3 in Mele (2012) for further discussion of this point).

experiment and to comply with the experimenter's instructions in the first place. This leaves plenty of room for conscious agency to play a role. A discussion of the role of conscious agency should thus consider not just the contribution of conscious intentions to the final stage of action initiation, but also their contribution to earlier stages of action production.

In addition, in experiments investigating voluntary action, subjects are typically asked to choose between very simple actions they are indifferent to (when to flex the wrist, whether to press the right or the left button) and to perform them in unproblematic contexts. They have no reason to veto an action, to act at time t rather than at time t' , to choose one action over another or to alter the course of the action once started, apart from whim. If conscious intentions are crucial to acting for reasons, to steering actions in difficult contexts, and to keeping them on course, then their role will be obscured in such experimental settings.

3. Action control

The objections reviewed in the previous section suggest that neither Libet's data nor more recent empirical results from psychology and neuroscience provide incontrovertible evidence that conscious intentions play no role in the initiation of action. But even if they did, would we thereby be forced to conclude that conscious agency is an illusion? The temptation to think this conclusion inevitable depends on our taking action initiation to be the touchstone of conscious agency. While this assumption appears widespread among both scientists and philosophers, should we really accept it?

The idea that action initiation is the touchstone of agency certainly seemed to play a central role in early versions of the causal theory of action. The causal theory of action, first made popular by Davidson (1980, Essay 1) and Goldman (1970), holds that behavior qualifies as action just in case it has psychological antecedents of a certain kind. While different philosophers offered different accounts of what these psychological antecedents must be, disagreeing among other things on whether intentions are reducible to belief and desire complexes, they all agreed that the role of these psychological antecedents is to cause the ensuing action. In an early and powerful critique of these causal theories, Harry Frankfurt (1978) claimed that their main flaw was to direct attention away from the actions themselves and away from the time at which they occur. As a result, Frankfurt argued,

causal theories fail to acknowledge that: "a person must be in some particular relation to the movements of his body *during* the period of time in which he is presumed to be performing an action" (1978: 157). Frankfurt characterized this relation as one of guidance and pointed out that while guidance must appeal to causal mechanisms, a guidance theory of action differs in two crucial ways from a causal theory of action.⁸ First, the activity of these causal mechanisms is not prior to but concurrent with the movements they guide. Second, and perhaps even more importantly, it is not essential to the purposefulness of these movements that the mechanism under whose guidance they proceed causally affect them. Frankfurt illustrates his point with the following analogy:

A driver whose automobile is coasting downhill in virtue of gravitational forces alone may be entirely satisfied with its speed and direction, and so he may never intervene to adjust its movement in any way. This would not show that the movement of the automobile did not occur under his guidance. What counts is that he was prepared to intervene if necessary and that he was in a position to do so more or less effectively. Similarly, the causal mechanisms which stand ready to affect the course of a bodily movement may never have occasion to do so; for no negative feedback of the sort that would trigger their compensatory activity may occur. The behavior is purposive not because it results from causes of a certain kind, but because it could be affected by certain causes if the accomplishment of its course were to be jeopardized. (1978: 160).

The central role assigned by Frankfurt to guidance mechanisms in the characterization of agency finds a strong echo in the literature on motor cognition with its emphasis on monitoring and control mechanisms.⁹ According to the very influential internal model theory of motor control, motor control strategies are based on the coupling of two types of internal

⁸ While I am focusing here on Frankfurt's guidance theory of action, it is important to note that other philosophers also criticized the over-simplistic view of the relation between intention and action found in early causal theories of action. For instance, Brand (1984) Bishop (1989) and Mele (1992) all insist that a full-blown causal theory should incorporate the guiding and monitoring roles of intentions in the production of intentional action. The notion of an intention-in-action proposed by Searle (1983) was also aimed at capturing the close and continuous connection between intention and ongoing action (see, for instance, the discussion in Pacherie (2000)).

⁹ Useful entry points into the literature on motor cognition are provided by Jeannerod (1997, 2006).

models: inverse models and forward models (Frith, Blakemore, & Wolpert 2000; Jordan & Wolpert 1999; Wolpert 1997; Wolpert & Ghahramani 2000; Wolpert, Ghahramani, & Jordan 1995). Inverse models compute the motor commands needed for achieving a desired state given the current state of the system and of the environment. An efference copy of these commands is fed to forward models, whose role is to make predictions about the consequences of the execution of these commands. The control of action is thought to depend on the coupling of inverse and forward models through a series of comparators: error signals arising from the comparison of desired, predicted, and actual states (monitoring) are used for various kinds of regulation (control). In particular, they can be used to correct and adjust the ongoing action in the face of perturbations, as well as to update both inverse and forward models to improve their future functioning.

While both Frankfurt's critique of causal theories and recent work in motor cognition stress the crucial role of action control and not just action initiation in agency, one may still wonder what this has to do with conscious agency. After all, it seems that action control could well proceed independently of conscious awareness. The real issue, however, isn't whether some or much of action control can operate automatically and outside of conscious awareness, but whether action control always takes this form.

While the internal model theory of motor control was initially introduced to account for fine-grained aspects of motor control (such as joint angles, torque, limb positions and trajectories), more recent versions of the theory emphasize the hierarchical nature of motor control (Jeannerod 1997; Fletcher & Frith 2009). They propose that internal inverse and forward models are arranged in a hierarchy and that error signals generated at one level of the hierarchy can propagate to the next level when correction mechanisms at this level are not able to make the necessary compensations. As a first approximation, one can distinguish three main levels in this action specification and control hierarchy (Pacherie, 2008). At the highest level, action representations represent the whole action as a unit, in terms of its overarching goal and of the sequence of steps or subgoals needed to achieve that goal. At this level, the action may still be represented in a rather abstract format. The second level is concerned with the implementation of each step in the action plan and involves selecting an appropriate motor program given the immediate goal and contextual information about the current state of the agent and the current state of its environment. In other words,

processes at this level are in charge of anchoring the successive steps of the action plan in the current situation and of selecting appropriate motor programs. Finally, once a motor program has been selected, the exact values of its parameters must still be set. This is done at the third level, where incoming sensory information about external constraints is used to specify these values.

This distinction of three levels is an oversimplification and should be qualified in several ways. First, the organization within each level can itself be decomposed into hierarchical stages. Second, the distinction between levels is not always sharp. In particular, a given action may be planned to a greater or a lesser extent, depending on the agent's expertise. Third, talk of a hierarchical organization and of a series of levels may give the impression that the processing steps must be ordered serially; that planning must be over before programming starts, and that programming in turn must be over before the execution starts. As Jeannerod (1997) points out, however, activation in the cortical areas thought to correspond to the various levels of organization occurs simultaneously and the existence of a sequence can only be detected statistically. Yet, the existence of parallel processing in the motor system does not contradict the idea of hierarchy of levels. A hierarchy between levels implies degrees of specialization for each level but it does not imply a sequential order of activation.

Acknowledging the existence of different levels of action control may allow us to accommodate both unconscious and conscious action control processes. Indeed, direct evidence for this duality of control processes comes from a set of experiments (Fournieret and Jeannerod 1998; Fournieret *et al.* 2001, 2002) in which participants were instructed to move a stylus on a graphic tablet along a straight line to a visual target. Participants could not see their drawing hand, but its trajectory was visible as a line on a computer screen. On some trials, the experimenter introduced a directional bias electronically so that the visible trajectory no longer corresponded to that of the hand. When the bias was small ($< 7^\circ$) participants made automatic adjustments of their hand movements to reach the target but remained unaware that they were making these corrections. It is only with larger biases that participants became aware of a discrepancy and began to use conscious correction strategies to compensate for the bias and reach the target. For instance, in trials with a 15° perturbation to the right, participants explicitly reported a strategy of moving their hand to

the left so as to be able to join the starting point and the target with a straight line. These results suggest that although small discrepancies between predicted and actual sensory feedback are detected at some level since they are used to make appropriate corrections of the hand movement, they are not normally consciously monitored and the corrections are automatic. However, when discrepancies become too large to be automatically compensated, we become aware of them and shift to a conscious compensation strategy.

Thus far, I have suggested in this section that we should heed Frankfurt's insight and take agency to have more to do with the guidance and control than with the initiation of movement. Importantly, and following Frankfurt again, it is not essential for movements to be controlled that control mechanisms actually affect their course, but that these control mechanisms would intervene to adjust the movements if the need arose. In addition, there is evidence that action control operates at several levels in parallel and that while low-level motor control may be automatic and unconscious, control at higher-level may be conscious. In other words, an action may be consciously controlled even though unconscious rather than conscious control mechanisms intervene to adjust movements, provided conscious control mechanisms would have kicked in, had automatic movement corrections proved insufficient.

Finally, I would like to draw two further points from this discussion of action control. The first point is that Libet's suggestion that conscious intentions can veto actions even though they don't initiate them can be easily accommodated within this control framework.¹⁰ The same control processes that are used to adjust and correct actions on the basis of the output of a predictive model could also be used to cancel them completely. The instructions Libet gave his subjects in his veto experiments (Libet 2005), asking them both to prepare to flex their wrist at a prearranged time and then to veto the action, may appear too artificial to

¹⁰ Note that Libet actually made two claims about conscious vetoes. The first claim was that the interval between conscious intention and movement onset is sufficient to allow a process of conscious veto, which would inhibit an impending action before execution. The second, much more contentious claim, reflecting his dualist attitude, was that these decisions to inhibit action involved a purely conscious form of "free won't" and did not depend on unconscious brain processes. I am concerned here only with the first of these two claims. For empirical evidence against the second, see for instance Brass & Haggard (2007) or Filevich, Kühn and Haggard (2013).

warrant the claim that the subjects actually formed the intention to flex their wrist and then vetoed the action (see, for instance, Mele (2007) for a critical discussion of this experiment). However, as pointed out by Haggard (2008), it is easy to think of everyday situations where we find ourselves about to perform a certain action (e.g., sending an angry email) but refrain at the last minute.

The second point is that if action control is an essential function of intentions, then we should perhaps stop thinking of intentions as simply mental representations of goals somehow triggering motor processes that, if everything goes well, will yield the desired outcomes. Rather, we should think of monitoring and control processes as intrinsic to intentions, that is, of intentions as encompassing not just representations of goals but also a specific set of monitoring and control processes organizing and structuring the motor processes that themselves generate movements. On this view, representations of goals should be seen as convenient indexes to the much richer set of representations and processes that constitute an intention rather than as constituting the intentions all by themselves. In construing intentions as simply representations of goals, we may be concentrating on the finger and thus missing the moon.

4. Advance Action Planning

At the end of section 2, I suggested that even if we had overwhelming evidence that actions are initiated unconsciously, we should look at what happens both before and after action initiation before concluding that conscious agency is an illusion. In section 3, I discussed the role conscious agency could play in the control of action after its initiation. I now turn to what happens before action initiation and to the role of prospective intentions.

Bratman's theory of human planning agency (Bratman 1987) probably offers the most detailed and influential philosophical account of prospective intentions, or as he calls them, future-directed intentions. I start with a brief review of his account of prospective intentions, of what functions they serve and what makes it useful to have them. I then turn to the issue what kind of cognitive processes are engaged in advance planning and how these relate to the processes involved in immediate intentions.

We are, in Michael Bratman's words, planning agents regularly making more or less complex plans for the future and guiding our later conduct by these plans. This planning ability appears to be if not unique to humans at least uniquely developed in the human species. People can, and frequently do, form intentions concerning actions not just in the near but also in the distant future (e.g., to make a six-month trip to Asia after retirement). Why should we bother forming today an intention to do something tomorrow, or next week or ten years from now? What purposes can forming prospective intentions serve? What benefits does it bring us?

Bratman offers two complementary answers to that challenge. The first stems from the fact that we are epistemically limited creatures: our cognitive resources for use in attending to problems, gathering information, deliberating about options and determining likely consequences are limited and these processes are time consuming. As a result, if our actions were influenced by deliberation only at the time of action, this influence would be minimal as time pressure isn't conducive to careful deliberation. Forming prospective intentions makes advance planning possible, freeing us from that time pressure and allowing us to deploy the cognitive resources needed for successful deliberation. Second, intentions once formed commit us to future courses of action, thus making the future more predictable and making it possible for agents to coordinate their activities over time and to coordinate them with the activities of other agents. Making deliberation and coordination possible are thus the two main benefits that accrue from a capacity to form prospective intentions.

What makes it possible for prospective intentions to yield these benefits is, according to Bratman, the fact that intentions essentially involve commitments to action. Bratman distinguishes two dimensions of commitments: a volitional dimension and a reasoning-centered dimension. The volitional dimension concerns the relation of intention to action and can be characterized by saying that intentions are "*conduct-controlling pro-attitudes*" (1987: 16). In other words, unless something unexpected happens that forces me to revise my intention, my intention today to go shopping tomorrow will control my conduct tomorrow. The reasoning-centered dimension of commitment is a commitment to norms of practical rationality and is most directly linked to planning. What is at stake here are the roles played by intentions in the period between their initial formation and their eventual execution. First, intentions have what Bratman calls a characteristic *stability* or inertia: once

we have formed an intention to *A*, we will not normally continue to deliberate whether to *A* or not. In the absence of relevant new information, the intention is rationally required to resist reconsideration: we will see the matter as settled and continue to so intend until the time of action. Intentions are thus *terminators of practical reasoning* about ends or goals. Second, during this period between the formation of an intention and action, we will frequently reason from such an intention to further intentions, reasoning from instance from intended ends to intended means or preliminary steps. When we first form an intention, our plans are typically only partial, but if they are to eventuate into action, they will need to be filled in. Thus intentions are also *prompters of practical reasoning* about means.

In the remainder of this section, I will concentrate on this last function of prospective intentions.¹¹ What kind of cognitive processes are engaged in advance planning? To what extent can we and do we actually plan our actions in advance? How is the gap bridged between prospective intentions that may initially remain unspecific in many ways and immediate intentions specific enough to ensure that the action is actually produced?

Let me consider first the cognitive processes that can support advance planning. On a traditional philosophical picture, action planning is often, implicitly if not explicitly, taken to involve *instrumental reasoning* exploiting semantic knowledge. We have instrumental beliefs about means-ends relations as well as various beliefs about how the world is and we use a relevant subset of these beliefs as premises in our reasoning from intended ends to intended means. Another way in which advance planning could be carried out is by using what has now become known as Mental Time Travel, the faculty that allows a person to mentally project herself backward in time to relive past events or forwards to pre-live events (Klein 2013; Klein *et al.* 2010; Suddendorf *et al.* 2009; Suddendorf & Corballis 1997, 2007; Suddendorf & Busby 2005; Szpunar 2010; Wheeler 1997). Mental time travel exploits episodic information and involves processes of mental simulation. Mental time travel in the

¹¹ For reasons of space, I decided to limit my discussion to action planning after the formation of an initial prospective intention. I am not assuming that the formation of prospective intentions always involves explicit conscious deliberation. When it does, however, deliberation could involve classical forms of practical reasoning, but also what I describe below as mental time travel processes. For instance, we may mentally simulate various potential future situations and use our emotional responses to these imagined scenarios as guides to our decisions (Boyer 2008).

past, the conscious reliving of past events known as episodic memory, has been intensively studied (e.g., Tulving 1983, 2005). Mental travel into the future, in contrast, has only recently begun to draw attention. However, recent work indicates that mental travel into the past and mental travel into the future are closely related, involving similar cognitive processes and recruiting strongly overlapping neural systems (D'Argembeau & Van der Linden 2006; Hassabis *et al.* 2007; Klein 2002).¹²

Planning using instrumental reasoning (IR planning, for short) and planning using mental time travel (MTT planning) should not be seen as mutually exclusive conceptions of what action planning is about, rather they may often function as complementary processes. Nor should we think that instrumental reasoning exploits only semantic information while mental time travel is concerned only with episodic information. For one thing, semantic and episodic information are not without links. The encoding of episodic information often relies on semantic information we already have and, conversely, semantic information is often extracted from episodic information.

A further important communality between IR planning and MTT planning is that conscious thought appears to be needed for both. In their discussion of the functions of conscious thought, Baumeister and Masicampo (2010) argue that human consciousness enables, among other things, the construction of meaningful, sequential thought, as in narratives and logical reasoning, the simulation of hypothetical scenarios and the exploration of options in complex decisions. As they point out, evidence has accumulated that automatic, non-conscious systems do not properly engage in logical reasoning (Lieberman *et al.* 2002; Smith & DeCoster 1999, 2000). There is also direct evidence that logical reasoning is the province of conscious thought. In particular, studies have shown that manipulations of conscious processes (e.g., hampering conscious processes by cognitive load or increasing conscious involvement by setting up conscious goals) affects performance in logical reasoning tasks but that corresponding manipulations of unconscious processes have no effect on reasoning performance (De Neys 2006; DeWall *et al.* 2008). Similarly, while there is evidence that mental time travel into the past or into the future can often take place in an involuntary or

¹² Indeed, several researchers have argued that mental time travel into the future is a crucial cognitive adaptation, enhancing planning and deliberation by allowing a subject to mentally simulate and evaluate contingencies, and thus enhancing fitness, and that mental time travel into the past is subsidiary to our ability to imagine future scenarios (Dudai & Carruthers 2005; Suddendorf & Corballis 2007).

spontaneous mode (Berntsen & Jacobsen 2008), theorists have argued that mental time travel in the service of planning is a goal-directed and intentional process that requires the engagement of higher order executive mental functions (e.g., Suddendorf & Corballis 1997, 2007; Wheeler 1997). This claim is supported by neuroimaging evidence indicating that MTT planning engages the neural substrates of working memory, including the dorsolateral prefrontal cortex (Gerlach *et al.* 2011; Spreng *et al.* 2010). As the engagement of working memory resources is generally considered as the hallmark of consciously controlled processes as opposed to automatic processes, that are independent of working memory (e.g., Evans 2008; Baars & Franklin 2003), this strongly suggests that MTT planning is a form of conscious thought, involving the cooperation of the autobiographical memory system and of executive control processes.

While IR planning and MTT planning share important properties, what is distinctive of MTT planning is that the representations it yields inherit some of the central characteristics of episodic memories. First, MTT planning involves imagining or mentally constructing specific future events, including at least some of their particularities. Second, these events are not just thought about, they are experienced. This experiential quality is manifested first by the multimodal content of these simulations that bind information from diverse systems (perceptual, motor, emotional) into a unitary episode representation and second by their auto-noetic dimension, that is, these events are mentally experienced from a first-person perspective.

The construction of plans for future actions depends in part on semantic memory since it is crucial to their success that the plans we come up with be consistent with our general knowledge about the world. Yet, filling in the details of a plan may depend on our ability to imagine future episodes, since they provide the particularities that will help fine-tune the plan to the particular occasion. However, a number of factors may modulate the extent to which we rely on advance planning and the extent to which this planning takes the form of IR as opposed to MTT planning. As an illustration, let me contrast three cases. As a first case, suppose that I form the prospective intention to go to my office tomorrow as I normally do on weekdays. In such a case, it seems that I don't need to engage in explicit planning as to how to get to my office. I already have a well worked out plan stored in memory and the route is familiar enough that I can trust myself to do the right thing when the time comes.

Now, consider instead the case where I have my first appointment tomorrow with a new dentist. I have never been to this particular dental office before but I have been living in Paris for many years and I know my way around the city. In that case, it may be worthwhile for me not just to rely on semantic information (e.g., public transportation maps) but also to engage in the mental simulation of different ways of getting there, using stored episodic information I have. For instance, I may remember that changing lines at this station takes forever and involves walking along endless, badly lit, corridors or I may remember getting stuck in heavy traffic on a given bus line. Mentally simulating these scenarios may not just help me decide among various options but also facilitate the future implementation of the chosen scenario through the anticipation of relevant situational cues. Finally, consider a third case where I am setting foot in a given city, say Budapest, for the first time in my life. In deciding how to get from one place to another, I have no relevant episodic information I could rely on. All I have to get by is semantic information in the form of maps or instructions from guidebooks. What these examples suggest is that the extent to which we rely on MTT or on IR in planning our actions depends on the degree of novelty of the prospective action and on whether the information at our disposal is mostly semantic or episodic in form. In addition, whether we engage in MTT planning may also depend whether we think we need to reinforce our motivation. For instance, while on a diet, I may be aware that it will be difficult for me to resist ordering chocolate cake tonight at the restaurant; rehearsing a scenario where I virtuously order fruit salad instead may help me abide by my resolution to limit my calorie intake.¹³

The distinction between IR planning and MTT planning bears an interesting relation to the distinction Gollwitzer (1999) draws between goal-intentions ("I intend to do x") and implementation-intentions ("when situation x arises, I will initiate the goal-directed response y"). While IR reasoning mainly involves deriving sub-goals from goals and thus moving from more general to more specific goal-intentions, MTT planning may be strongly involved in the formation of 'implementation intentions', which specify the "when, where, and how of

¹³ In addition, there appear to be important individual differences in the ability to project oneself into possible future events. A recent study (D'Argembeau & Van der Linden 2005) provides evidence that the individual differences in dimensions known to affect memory for past events similarly influence the experience of projecting oneself into the future. People less adept at recalling in vivid detail past episodes of their life, are also less able to simulate specific future events.

responses leading to goal attainment [...] and thus link anticipated opportunities with goal-directed responses" (Goldwitzer 1999: 494).

A number of studies of implementation intentions in applied psychology (for a review and meta-analysis, see Goldwitzer & Sheraan 2006) provide evidence that forming implementation intentions furthers goal attainment. For instance, goal-intentions such as taking vitamins everyday are more likely to be acted on if people link their implementation to a specific external event (Sheeran & Orbell 1999). According to Gollwitzer (1999), what explains the efficacy of implementation intentions is the fact that their formation triggers two sets of processes. First, when an implementation intention is formed, mental representations of the relevant situational cues become highly activated, leading to heightened accessibility and thus a better detection of these cues when they are encountered (Aarts *et al.* 1999; Gollwitzer 1999; Webb and Sheeran 2004). Second, implementation intention formation not only enhances the accessibility of the specified situational cue, but also forges an association between that cue and a response that is instrumental for obtaining one's goal, thus making action initiation more immediate and efficient.¹⁴

Importantly, Gollwitzer points out that experiments on implementation intentions suggest that they automatize action initiation: when the specified situational cue is present, action is initiated without requiring the formation of a conscious intention to act now. This doesn't mean that the agent is not conscious of what he intends to do now, but that his intention is not arrived at via a conscious decision to act now. The distinction between intention formation and intention acquisition proposed by Mele (2002) can help clarify this point. Mele reserves the term intention formation for intentions arrived at via a conscious decision and speaks of intention acquisition for intentions arrived at by other means. What Goldwitzer suggests is that in forming implementation intentions we dispense with the need to form intentions to act now, instead the intention to act now is automatically triggered by the situational cue. Goldwitzer describes this phenomenon as one of strategic automaticity. Like habitual automaticity, strategic automaticity leads to faster and more efficient action

¹⁴ Such implementation intentions may take advantage of the fact that externally-cued intentions are normally more strongly held, in the sense of being harder to overturn, than internally-generated intentions (Fleming *et al.* 2009).

initiation, the main difference between them lies in how they originate. While habitual automaticity depends on repeated and consistent practice, strategic automaticity stems from the single act of will implied in forming implementation intentions.

So far, I have concentrated on the role of conscious agency prior to action initiation and more specifically on the advance planning processes that take place once one has formed a prospective intention to pursue a certain goal. However, understanding what happens prior to action initiation may also help us understand what happens at the time of action initiation. In particular, it could help us answer the following question: what additional contribution, if any, could conscious immediate intentions make to the production of actions that are already preceded by prospective intentions?

In this section, I distinguished three types of scenarios involving prospective intentions. The answer to the present question could depend, I suggest, on what scenario is exemplified. When we form a prospective intention to perform an action for which we have a well-established routine, we don't need to engage in advance planning. We can rely on situational cues to trigger the appropriate pattern of behavior at the appropriate time. Alternatively, if the prospective action is not habitual but we have relevant episodic information we can exploit, we can use MTT planning to form relevant implementation intentions and thus strategically delegate the initiation of goal-directed responses to anticipated situational cues. In both types of scenarios, the conscious formation of immediate intentions does not appear to be needed to ensure action initiation. This contrasts with the third type of scenario where either for lack of relevant episodic information or for lack of sufficient motivation to engage in MTT planning, the action plan we have formed consists solely or mainly in a hierarchical structure of goal-intentions. Goal-intentions even if they specify goals in a detailed manner do not specify how the intended action is to be weaved into the ongoing flow of behavior and experience. To bridge this gap between her goal intentions and action initiation, the agent must retain some endogenous control over action initiation. She cannot delegate it to automatic responses to environmental triggers, but must form immediate intentions specifying how the intended action is to be contextually modulated. In such cases, I surmise, action initiation would depend on the formation of conscious immediate intentions.

5. Concluding remarks

In this paper, I have tried to argue that the debate on the role of conscious agency in action production started by Libet's experiments was too narrowly focused on action initiation. I suggested that if we widened our temporal horizons and considered the potential role of conscious agency both prior to and after action initiation, we might be less tempted to succumb to will-skepticism. In section 3, I argued that conscious action control during action execution constitutes an important facet of conscious agency. In particular, I emphasized the fact that conscious control does not necessarily mean causal intervention: action may be consciously controlled even though unconscious rather than conscious control mechanisms intervene to adjust movements, provided conscious control mechanisms would have kicked in, had automatic control proved insufficient. In section 4, I argued that conscious agency could also be manifested in the form of advance action planning and that such planning can, when it takes a particular form, induce strategic automaticity. While, for the sake of exposition, I distinguished three routes that may lead from an initially underspecified prospective intentions to action initiation – reliance on action routines, IR planning and MTT planning, it should be expected that in most cases we will use mixed strategies.

My parting message is that we shouldn't see automaticity as an unbearable affront to our vanity as conscious agents. Rather than picturing automaticity and conscious agency as fiends vying for the same territory, one army being supported by cognitive scientists in an all-conquering mood, the other by philosophers fighting a rear-guard action, it may be more fruitful to see them as complementary forces, even if their cooperation isn't always easy and straightforward. If anything, what makes human agency unique is the richness and the complexity of their interactions and the flexible ways in which control can be passed up to conscious processes or down to automatic processes as the need arises.

Acknowledgments

Preliminary versions of this paper were presented at the Topoi Conference "Intentions: Philosophical and Empirical Issues" in Rome in November 2012 and at the Colloquium series of the Department of Philosophy at the Central European University in Budapest in February 2013. I would like to thank both audiences for their questions and comments. I am also grateful to two anonymous reviewers for this journal for their insightful comments and

suggestions. I completed this paper while a Fellow at the Institute of Advanced Studies at the Central European University in Budapest and am grateful to this institution for its support.

- Aarts, H, Dijksterhuis, A P, & Midden, C (1999) To plan or not to plan? Goal achievement or interrupting the performance of mundane behaviors. *European Journal of Social Psychology*, 29(8): 971-979.
- Baars, B J, & Franklin, S (2003) How conscious experience and working memory interact. *Trends in Cognitive Sciences*, 7(4): 166-172.
- Banks, WP and Isham, E A (2011) Do we really know what we are doing? Implications of reported time of decision for theories of volition. In W Sinnott-Armstrong and L Nadel (eds) *Conscious will and responsibility*. New York: Oxford University Press, pp. 47-60.
- Baumeister, R F, & Masicampo, E J (2010) Conscious thought is for facilitating social and cultural interactions: How mental simulations serve the animal–culture interface. *Psychological Review*, 117(3): 945-971.
- Bayne, T J & Levy, N (2006) The feeling of doing: Deconstructing the Phenomenology of Agency. In N Sebanz & W Prinz (eds), *Disorders of volition*. Cambridge, Mass: MIT Press, pp. 49-68.
- 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.
- Berntsen, D, & Jacobsen, A S (2008) Involuntary (spontaneous) mental time travel into the past and future. *Consciousness and Cognition*, 17(4): 1093-1104.
- Bishop, J C (1989) *Natural agency: an essay on the causal theory of action*. Cambridge University Press.
- Boyer, P (2008) Evolutionary economics of mental time travel? *Trends in Cognitive Sciences*, 12(6): 219-224.
- Brand, M (1984) *Intending and acting: toward a naturalized action theory*. Cambridge, Mass: MIT Press.
- Brass, M & Haggard, P (2007) To do or not to do: the neural signature of self-control. *Journal of Neuroscience*, 27: 9141–9145.
- Bratman, M E (1987) *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.
- Chalmers, D (1995) Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2, 3: 200-19.

- D'Argembeau, A & Van der Linden, M (2006) Individual differences in the phenomenology of mental time travel: The effect of vivid visual imagery and emotion regulation strategies. *Consciousness and Cognition* 15: 342-350.
- Davidson, D (1980) *Essays on actions and events*. Oxford: Oxford University Press.
- Dennett, D C (1991) *Consciousness explained*. Boston, MA: Little Brown.
- Dennett, D C (2003) *Freedom evolves*. New York: Viking.
- De Neys, W (2006) Dual processing in reasoning: two systems but one reasoner. *Psychological Science*, 17: 428–433.
- DeWall, C N, Baumeister, R F, & Masicampo, E J (2008) Evidence that logical reasoning depends on conscious processing. *Consciousness and Cognition*, 17: 628–645.
- Dudai, Y & Carruthers, M (2005) The Janus face of Mnemosyne. *Nature*, 434: 567.
- Evans, J S B (2008) Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59: 255-278.
- Filevich E, Kühn S, Haggard P (2013) There Is No Free Won't: Antecedent Brain Activity Predicts Decisions to Inhibit. *PLoS One* 8(2): e53053. doi:10.1371/journal.pone.0053053
- Fischer, J M (1994) *The metaphysics of free will*. Oxford: Blackwell Publishers.
- Fourneret, P, & Jeannerod, M (1998) Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, 36, 11: 1133-1140.
- Fourneret, P, Franck, N, Slachevsky, A, & Jeannerod, M (2001) Self-monitoring in schizophrenia revisited. *Neuroreport*, 12(6): 1203-1208.
- Fourneret, P, Vignemont, F D, Franck, N, Slachevsky, A, Dubois, B, & Jeannerod, M (2002) Perception of self-generated action in schizophrenia. *Cognitive Neuropsychiatry*, 7(2): 139-156.
- Flanagan, O (1996) Neuroscience, agency, and the meaning of life. In *Self-Expressions*. Oxford: Oxford University Press, pp. 53-64.
- Fleming, S M, Mars, R J, Gladwin, T E, & Haggard, P (2009) When the brain changes its mind: Flexibility of action selection in instructed and free choices. *Cerebral Cortex*, 19(10): 2352–2360.
- Frankfurt, H (1978) The problem of action. *American Philosophical Quarterly*, 15, 2: 157-162.

- Frankfurt, H (1988) *The importance of what we care about*. Cambridge: Cambridge University Press.
- Fletcher, P C, & Frith, C D (2009) Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Neuroscience*, 16: 48-58.
- Frith, C D, Blakemore, S-J, & Wolpert, D M (2000) Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London B*, 355: 1771–1788.
- Gallagher, S (2006) Where's the action? Epiphenomenalism and the problem of free will. In W Banks, S Pockett, & S Gallagher (eds), *Does consciousness cause behavior? An investigation of the nature of volition*, Cambridge, MA : MIT Press, pp. 109-124.
- Gerlach, K D, Spreng, R N, Gilmore, A W, & Schacter, D L (2011) Solving future problems: Default network and executive activity associated with goal directed mental simulations. *Neuroimage*, 55(4): 1816–1824.
- Glynn, I M (1990) Consciousness and time. *Nature*, 348, 6301: 477-479.
- Goldman, A (1970) *A theory of human action*. Englewood Cliffs, NJ: Prentice-Hall.
- Gollwitzer, P M (1999) Implementation intentions: Strong effects of simple plans. *American Psychologist*, 54: 493–503.
- Gollwitzer, P M, & Sheeran, P (2006) Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology*, 38: 69-119.
- Gomes, G (1998) The Timing of Conscious Experience: A Critical Review and Reinterpretation of Libet's Research. *Consciousness and Cognition*, 7, 4: 559-595.
- Gomes, G (1999) Volition and the readiness potential. *Journal of Consciousness Studies*, 6/8-9: 59-76.
- Haggard, P (2006) Conscious intention and the sense of agency. In N Sebanz and W Prinz (eds) *Disorders of volition*. Cambridge, MA: MIT Press, 69-86.
- Haggard, P (2008) Human volition: Towards a neuroscience of will. *Nature Reviews Neuroscience*, 9: 934-46.
- Haggard, P and Eimer, M (1999) On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research* 126: 128–133.

- Hallett, M (2007) Volitional control of movement: the physiology of free will, *Clinical Neurophysiology* 118: 1179-92.
- Hassabis, D K D, Vann, S D, & Maguire, E A (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, 104: 1726–1731.
- Jeannerod, M (1997) *The cognitive neuroscience of action*. Oxford: Blackwell.
- Jeannerod, M (2006) *Motor cognition*. Oxford: Oxford University Press.
- Joordens, S, van Duijn, M & Spalek, T M (2002) When timing the mind one should also mind the timing: Biases in the measurement of voluntary actions. *Consciousness and Cognition*, 11, 2: 231-240.
- Jordan, M I, & Wolpert, D M (1999) Computational motor control. In M Gazzaniga (ed), *The cognitive neurosciences*. Cambridge, MA: MIT Press.
- Keller, I & Heckhausen, H (1990) Readiness Potentials preceding spontaneous motor acts: voluntary vs. involuntary control. *Electroencephalography and Clinical Neurophysiology*, 76: 351-61.
- Kim, J (1993) *Supervenience and mind*. Cambridge: Cambridge University Press.
- Klein, S B (2002) Memory and temporal experience: The effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Social Cognition*, 20: 353–379.
- Klein, S B (2013) The complex act of projecting oneself into the future. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(1): 63-79.
- Klein, S B, Robertson, T E, & Delton, A W (2010) Facing the future: Memory as an evolved system for planning future acts. *Memory and Cognition*, 38 (1): 13-22.
- 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, & 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: 623–42.
- Lieberman, M D, Gaunt, R, Gilbert, D T, & Trope, Y (2002) Reflexion and reflection: A social cognitive neuroscience approach to attributional inference. In M P Zanna (ed), *Advances in experimental social psychology*. San Diego, CA: Academic Press, pp. 199–249.

- McKenna, M (2005) Reasons Reactivity & Incompatibilist Intuitions. *Philosophical Explorations*, 8 (2): 131–43.
- Mele, A R (1992) *Springs of action: Understanding intentional behavior*. Oxford University Press.
- Mele, A R (2007) Free will: action theory meets neuroscience. In C Lumer & S Nannini (eds), *Intentionality, deliberation and autonomy*, Burlington, VT: Ashgate, pp. 257-272.
- Mele, A R (2009) *Effective intentions: the power of conscious will*. New York: Oxford University Press.
- Mele, A R (2012) Another Scientific Threat to Free Will? *The Monist*, 95(3): 422-440.
- Miller, J, Shepherdson, P, & Trevena, J (2011) Effects of Clock Monitoring on Electroencephalographic Activity Is Unconscious Movement Initiation an Artifact of the Clock? *Psychological Science*, 22(1): 103-109.
- Miller, J O, Vieweg, P, Kruize, N, & McLea, B (2010) Subjective reports of stimulus, response, and decision times in speeded tasks: How accurate are decision time reports? *Consciousness and Cognition*, 19: 1013-36.
- Nahmias, E (2002) When consciousness matters: A critical review of Daniel Wegner's The illusion of conscious will. *Philosophical Psychology*, 15: 527-541.
- Pacherie, E (2000) The content of intentions. *Mind and Language*, 15, 4: 400-432.
- Pacherie, E (2006) Towards a dynamic theory of intentions. In S Pockett, WP Banks & S Gallagher (eds), *Does consciousness cause behavior? An investigation of the nature of volition*. Cambridge, MA: MIT Press, pp. 145-167.
- Pacherie, E (2008) The phenomenology of action: A conceptual framework. *Cognition*, 107, 1:179-217.
- Pockett, S (2004) Does consciousness cause behaviour? *Journal of Consciousness Studies*, 11/2: 23-40.
- 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.
- Roediger, H K, Goode, MK, and Zaroomb, FM (2008) Free will and the control of action. In J Baer, JC Kaufman and RF Baumeister (eds) *Are we free?* Oxford: Oxford University Press, pp. 205-225.

- 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.
- Scanlon, TM (2008) *Moral dimensions: permissibility, meaning, blame*. Cambridge, Mass: Belknap Harvard Press.
- Schlegel, A, Alexander, P, Sinnott-Armstrong, W, Roskies, A, Tse, P & Wheatley, T (2013) Barking up the wrong free: readiness potentials reflect processes independent of conscious will. *Experimental Brain Research*. DOI 10.1007/s00221-013-3479-3
- Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences*, 109(42): E2904-E2913
- Searle, J R (1983) *Intentionality: an essay in the philosophy of mind*. Cambridge University Press.
- Searle, J R (2001) *Rationality in action*. Cambridge, Mass: MIT Press.
- Sheeran, P, & Orbell, S (1999) Implementation intentions and repeated behaviour: Augmenting the predictive validity of the theory of planned behaviour. *European Journal of Social Psychology*, 29(2-3): 349-369.
- Smith, M (2003) Rational Capacities. In S Stroud and C Tappolet, (eds), *Weakness of will and practical irrationality*. Oxford: Clarendon Press, pp. 17-38.
- Smith, E R, & DeCoster, J (1999) Associative and rule-based processing: A connectionist interpretation of dual-process models. In S Chaiken & Y Trope (eds), *Dual-process theories in social psychology* (pp. 323–336). New York, NY: Guilford Press.
- Smith, E R, & DeCoster, J (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4: 108–131.
- Soon, C S, Brass, M, Heinze, H-J, Haynes, J-D (2008) Unconscious determinants of free decisions in the human brain, *Nature Neuroscience*, 11/5: 543-5.
- Spence, S (2009) *the actor's brain: exploring the cognitive neuroscience of free will*. New York: Oxford University Press.
- Spreng, R N, Stevens, W D, Chamberlain, J P, Gilmore, A W, & Schacter, D L (2010) Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *Neuroimage*, 53(1): 303–317.

- Suddendorf, T, Addis, D R, & Corballis, M C (2009). Mental time travel and the shaping of the human mind. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521): 1317-1324.
- Suddendorf, T, & Busby, J (2005) Making decisions with the future in mind: Developmental and comparative identification of mental time travel. *Learning and Motivation*, 3: 110–125.
- Suddendorf, T, & Corballis, M C (1997) Mental time travel and the evolution of the human mind. *Genetic, Social, and General Psychology Monographs*, 123: 133–167.
- Suddendorf, T, & Corballis, M C (2007) The evolution of foresight: What is mental time travel, and is it unique to human? *Behavioral and Brain Sciences*, 30: 299–351.
- Szpunar, K K (2010) Episodic future thought an emerging concept. *Perspectives on Psychological Science*, 5(2): 142-162.
- Tulving, E (1983) *Elements of episodic memory*. Oxford: Clarendon Press.
- Tulving, E (2005) Episodic memory and auto-noesis: Uniquely human? In H S Terrace & J Metcalfe (eds), *The missing link in cognition: origins of self-reflective consciousness*, Oxford: Oxford University Press, pp. 3-56.
- Webb, T L, & Sheeran, P (2004) Identifying good opportunities to act: Implementation intentions and cue discrimination. *European Journal of Social Psychology*, 34(4): 407-419.
- Webb, T L, & Sheeran, P (2007) How do implementation intentions promote goal attainment? A test of component processes. *Journal of Experimental Social Psychology*, 43: 295–302.
- Wheeler, M A, Stuss, D T, & Tulving, E (1997) Toward a theory of episodic memory: The frontal lobes and auto-noetic consciousness. *Psychological Bulletin*, 121: 331–354.
- van de Grind, W (2002) Physical, neural, and mental timing. *Consciousness and Cognition*, 11: 241–264.
- Velleman, D (2007) What good is a will? In A Leist & H Baumann (eds), *Action in context*, Berlin: de Gruyter, pp. 193-215.
- Wegner, DM (2002) *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Wolpert, D M (1997) Computational approaches to motor control. *Trends in Cognitive Sciences*, 1(6): 209–216.

Wolpert, D M, & Ghahramani, Z (2000) Computational principles of movement neuroscience. *Nature Neuroscience Supplement*, 3: 1212–1217

Wolpert, D M, Ghahramani, Z, & Jordan, M I (1995) An internal model for sensorimotor integration. *Science*, 269: 1880–1882.

Zhu, J (2003) Reclaiming Volition. *Consciousness and Cognition*, 10: 61-77.