Why neuroscience does not disprove free will

Marcel Brass[1], Ariel Furstenberg[2] & Alfred R. Mele [3]

[1]Department of Experimental Psychology, Ghent University

Henri Dunantlaan 2, 9000 Ghent, Belgium

marcel.brass@ugent.be

http://users.ugent.be/~mbrass

[2]Racah Institute of Physics,

Edmond and Lily Safra Center for Brain Sciences, The Hebrew University Jerusalem

Edmond J. Safra Campus

Jerusalem 9190401, Israel

ariel.furstenberg@gmail.com

[3]Department of Philosophy, Florida State University

151 Dood Hall, Tallahassee 32306-1500, USA

amele@fsu.edu

Abstract

While the question whether free will exists or not has concerned philosophers for centuries, empirical research on this question is relatively young. About 35 years ago Benjamin Libet designed an experiment that challenged the common intuition of free will, namely that conscious intentions are causally efficacious. Libet demonstrated that conscious intentions are preceded by a specific pattern of brain activation, suggesting that unconscious processes determine our decisions and we are only retrospectively informed about these decisions. Libet-style experiments have ever since dominated the discourse about the existence of free will and have found their way into the public media. Here we review the most important challenges to the common interpretation of Libet-style tasks and argue that the common interpretation is questionable. Brain activity preceding conscious decisions reflects the decision process rather than its outcome. Furthermore, the decision process is configured by conditional intentions that participants form at the beginning of the experiment. We conclude that Libet-style tasks do not provide a serious challenge to our intuition of free will.

1. Introduction

Controversy about whether free will exists has a lengthy history (Bobzien, 1998; Dilman, 1999). But the issue may be more widely debated now than ever. The alleged news that free will is an illusion has found its way into the public media, and some researchers have presented evidence that this news has a negative influence on people's behaviour (Vohs and Schooler, 2008). Recent research suggests that whether people believe in free will or not has an effect on their lives. Belief and disbelief in free will have been argued to have consequences for our social behaviour (Baumeister et al., 2009; Vohs and Schooler, 2008), the way we evaluate the behaviour of others (Genschow et al., 2017) and even basic neurocognitive processes (Rigoni et al., 2011; Rigoni et al., 2013b). If claims about the existence or nonexistence of free will have such consequences, it is important to know how well-supported these claims are.

For centuries, the free will debate was restricted to philosophy and theology. Only relatively recently has free will been tackled empirically using neuroscience methodology. The first serious attempt to address the issue with an experimental approach was made by Benjamin Libet (Frith and Haggard, 2018; Libet, 1985; Libet et al., 1983; for a review see Saigle et al., 2018). In his classical experiment, Libet tested whether the conscious intention to act is preceded by unconscious brain processes. Libet's experiment has been viewed as a challenge to a strong and common intuition about free will (Nichols and Knobe, 2007; Sarkissian et al., 2010), namely, that it is at least sometimes up to us what we decide and our conscious decisions and intentions are causally relevant to what we do. Most people have the strong conviction that they can decide between different options, for example to order pizza rather

than pasta in an Italian restaurant. While people may admit that their decisions are partly influenced by factors beyond their control (e.g. hunger), they intuitively reject the idea that their decisions are products only of things they have no control over (Nahmias et al., 2014).

The main contribution of neuroscience to the free will debate has revolved around the question whether conscious decisions can be predicted from brain activation preceding such decisions (Bode et al., 2014; Brass et al., 2013). This contribution has attracted a lot of attention, including many articles in the popular press. Perhaps this is because many people find neuroscientific evidence that questions free will more convincing than metaphysical arguments (Monterosso et al., 2005). Furthermore, the formulation of a kind of neural determinism is much easier to grasp than traditional philosophical arguments regarding determinism and free will. The abstract philosophical discussion about whether beings with free will can inhabit a deterministic universe is replaced by the more concrete question whether our conscious decisions can be predicted from preceding neural activity or not and whether we have any control over what we decide.

Has neuroscience shown that free will is an illusion? The aim of this paper is to address this important question. We believe that the contribution of neuroscience to the free will debate can best be evaluated by a joint effort from philosophy and neuroscience. Our intention is not to address the question whether free will exists or not. Furthermore, we will not discuss the merits of different philosophical positions on free will. The main aim of this review is to evaluate the validity of the neuroscientific challenge to the existence of free will that has been presented by Libet-style tasks. Furthermore, we think that after 35 years of discussion on the

implications of the Libet task on the free will debate, it is time to integrate the findings and develop a more timely interpretation of the results.

In this paper, we will first discuss some conceptual issues before outlining the experimental evidence for the claim that our conscious decisions to act are preceded by unconscious brain processes that predetermine them. We will discuss the classical experiment by Libet (Libet, 1985; Libet et al., 1983) as well as more recent intracranial recordings in humans (Fried et al., 2011). We will also briefly discuss an influential fMRI study investigating the prediction of conscious choice from brain activity (Soon et al., 2008). Then we turn to research that questions both the validity of the experimental procedure and the interpretation of the results of Libet-style tasks. We will also discuss methodological and conceptual challenges to the task. Finally, based on the discussion of the literature and very recent decision models of Libet-style experiments, we will develop an alternative interpretation of Libet's results.

1.1. Decisions and intentions: a conceptual map

There are decisions to do things and decisions not to do things. There are also decisions about what is the case, as in "On the basis of a careful review of the evidence, the detective decided that Smith was lying about his whereabouts last weekend." Our primary concern in the sphere of decisions is decisions to do things.

We conceive of decisions to act as momentary actions of intention formation that are responses to uncertainty or unsettledness about what to do. For example, if Joe thinks about whether to accept or reject a job offer and decides to accept it, his thinking is not part of his

deciding to accept it. Instead, his deciding to do that is preceded by his thinking about what to do. And what it is for him to decide to accept the offer is for him to actively form an intention to accept it.

We hasten to add that we leave room for spontaneous decisions – decisions that are not products of thought about what to do. We also leave plenty of room for intentions that are not associated with corresponding decisions. When Joe gets to his office door in the morning on a normal day, he is not at all unsettled or uncertain about what to do. He unlocks the door. Joe intended to do that, and he had no need to decide to do it. Thus, we distinguish among thinking about what to do, deciding what to do, and intending to do something. In some cases, they are all part of the same process.

There are proximal and distal intentions (Mele 1992, pp. 143-44, 158) – intentions to do something *now* and intentions to do something in the (non-immediate) future. When Joe arrived at his office door, he had a proximal intention to unlock it. Shortly after he decided to accept that job offer, he had a distal intention to move to London six months from then. Proximal and distal decisions are distinguished from each other in the same way. The studies at issue here are specifically about *proximal* intentions and decisions.

Some intentions have an if-then or when-then structure. They are conditional intentions. Joe has an intention with the following content: When I say "now!" to myself, I will flex my right wrist at once. That is a conditional intention. In the cognitive literature, such conditional intentions have been termed 'prepared reflexes' (Hommel, 2001), because under some circumstances the 'if' component of the if-then rule reflexively triggers the specified action. In

the social psychological literature, conditional intentions are known as implementation intentions (Gollwitzer 1999, Gollwitzer and Sheeran 2006). Their function is to assist in the achievement of less specific goals. For example, while he is eating lunch, Joe recalls his intention to write a letter of recommendation for a colleague today. Shortly thereafter, he forms an implementation intention with the following content: When I return to my office from lunch, I will immediately sit down at my computer and start writing that letter. Some other conditional intentions are back-up plans. For example, Joe has an intention with the following content: If I fail to fix my car today, I will ask Jeff to do it tomorrow. What is common to all conditional intentions is the conditional (i.e., if-then or when-then) structure of their content. With these conceptual clarifications in place, we will now outline the classical experiment by Libet (Libet et al., 1983) and what has been the common interpretation of the results.

2. ***The Libet experiment***

The experiment by Libet and colleagues (Libet, 1985; Libet et al., 1983) is one of the most iconic experiments of experimental psychology. In this experiment, Libet combined two approaches to study cognitive function, creating an experimental setup that inspired generations of psychologists, neuroscientists and philosophers. The two elements are electrophysiological measurement of brain activity preceding intentional action and subjective timing of conscious intentions. In the experiment, participants have to carry out a simple action such as flexing their wrist or pressing a key. They can choose when to execute this action. While forming the intention to act, participants observe a revolving spot on a clock face. They are instructed to remember the location of the revolving spot when they first

experience the decision or urge to act. After they execute the action, they have to indicate on a clock face when exactly they became aware of the intention to act (defined as the “time of conscious intention to act” in the title of Libet et al., 1983), the so called will judgement or W. Libet’s first observation was that W precedes the actual motor response (M) by about 200 ms. This indicates that participants can distinguish the intention to act from the act itself. More important, during the experiment brain activity was recorded with electroencephalography (EEG). Previous research had indicated that voluntary action is preceded by a specific brain wave, the so called ‘Bereitschaftspotential’ or readiness potential (RP, Kornhuber and Deecke, 1964). Libet’s ingenious idea was to relate the onset of the RP to W, finding that the onset of the RP precedes W by a few hundred milliseconds. Interestingly, Libet et al. (1983) distinguished two types of RPs. First, RPs that arise from the spontaneous decision to act without any preplanning of the action (type II RP). Second, RPs that arise in situations where some preplanning occurred (type I RP). The type I RP has an earlier onset than the type II RP. But even in cases where no preplanning was reported by participants, the onset of the RP preceded W by about 500 ms. Based on these findings, Libet argued that the fact that a brain wave precedes the conscious intention by a few hundred milliseconds shows that the conscious intention cannot be the cause of the action but rather is the consequence of brain processes preceding it. In the words of Libet (1985, page: 536): ‘the brain "decides" to initiate or, at least, to prepare to initiate the act before there is any reportable subjective awareness that such a decision has taken place’. Thus, our intuition that we have conscious control over our actions seems to be wrong. Unconscious brain processes apparently determine what we do and we only learn about this immediately before we execute the action.

Why then do we form conscious intentions at all? According to Libet, the conscious intention allows us to be able to stop the action at the last moment. In the 200 ms between the conscious intention and the execution of the action we can consciously decide to veto the action. We will return to the veto idea later and discuss it in more detail.

2.1. Extension of Libet's findings using intracranial recording and functional MRI

Libet's basic findings have been replicated numerous times (Haggard and Eimer, 1999; Keller and Heckhausen, 1990; Rigoni et al., 2013b), demonstrating that they are reliable and can be reproduced in different labs. Importantly, there is one conceptual replication of the Libet task that merits further discussion. In a seminal study, Fried et al. (2011) carried out the Libet experiment in a group of patients with intracranial recordings. The electrodes were placed in different parts of the medial frontal cortex, including the SMA/preSMA and the anterior cingulate cortex (ACC). Such intracranial recordings have a few advantages over classical EEG recordings. First, it is possible to determine the exact location from which the activity is originating. Second, the signal is reliable enough to carry out single trial analyses. Third, one can investigate the firing rate of single neurons as well as the number of neurons that are recruited at a specific point in time. Fried and colleagues (2011) replicated Libet's basic result, showing that activity in units in the medial frontal cortex increased firing a few hundred milliseconds before W. Interestingly, they also demonstrated that the number of units that were recruited increased before W. In addition to this conceptual replication of Libet's findings, they reported some unique results. A variety of response patterns was observed for different neurons. Some neurons continuously increased the firing rate prior to W, very similar to what is observed in the RP in EEG recordings. However, some neurons also showed a sharp

increase of the neural firing prior to W, indicating that the RP might reflect an integration of different activation patterns. Most important, recording of multiple units and the high reliability of the recordings also allowed the use of multivariate pattern analysis (MVPA) to decode the onset of neural activity indexed as a departure from baseline. In contrast to the classical RP approach, which is based on trial averages, this is an approach based on single trial activation. By looking at the pattern of activity across different units it was possible to decode departure from baseline activity with an accuracy of 70 % as early as 500 ms before W. By pooling units across participants, the prediction accuracy could even be increased. To summarize, Fried et al. (2011) not only replicated Libet's results using intracranial recordings, they also provided a quantification of how well neural activity can predict the subjective awareness of intention prior to W.

Finally, there is a replication of a Libet-style experiment using functional magnetic resonance imaging (Soon et al., 2008). In an influential study, Soon et al. (2008) asked participants to choose (at a moment participants selected) between two response alternatives while observing a stream of letters on the computer screen. After they chose and executed the action, participants were asked to indicate the letter that was presented on the screen at the moment they became aware of their intention to act. Soon et al. (2008) used multivariate analysis techniques to predict the upcoming choice from brain activation. They found that it was possible to predict choices more than 8 seconds before participants reported that they consciously chose. However, the accuracy with which the researchers could predict behaviour was about 6-10 % above chance. Because the time scale of the prediction was so fundamentally different than in previous EEG experiments (i.e. several seconds compared to only a few hundred milliseconds in the Libet task), the paper attracted a lot of attention in the

scientific literature (the paper received enough citations to place it in the top 1% of its domain) and the public media. In subsequent work the original findings of Soon and colleagues were directly replicated (Bode et al., 2011). Furthermore, it was demonstrated that the findings generalize beyond simple decisions between response alternatives and can be conceptually replicated with choices between abstract intentions related to tasks rather than responses (Soon et al., 2013).

3. Challenges to the Libet experiment

The conclusions that Libet drew from his experiment have been strongly challenged on different grounds (e.g. Mele, 2014; Pockett and Purdy, 2010; Roskies, 2010b; Zhong, 2016). First, the question arises whether the experimental setup of the Libet experiment resembles anything that can be called a voluntary action as it occurs in a natural setting. This critique challenges the ecological validity and generalizability of the experimental setup (Mele, 2014). Second, the internal validity of the task has been challenged as well. Here, one major criticism is related to the subjective time of awareness (Banks and Isham, 2009; Matsuhashi and Hallett, 2008; Rigoni et al., 2010). Furthermore, it has been debated whether the method of reporting the conscious intention might alter the process under investigation (Lau et al., 2004; Rigoni et al., 2013a). Another criticism is related to the interpretation of the readiness potential. Does it really reflect unconscious preparation of the upcoming action (Schurger et al., 2012)? Finally, alternative interpretations of the Libet task locate the general conscious intention to act at the beginning of the experiment rather than at W (Keller and Heckhausen, 1990). In the following, we will systematically discuss all these challenges to the Libet experiment. Then, we will integrate the existing findings and provide an alternative view of the Libet task.

3.1. How ecologically valid is the Libet experiment?

One basic implicit assumption of the Libet task is that it can be generalized to other situations in which people act voluntarily. After all, the conclusions that have been drawn from this experiment often refer to human intentional behaviour in general. However, there are a number of challenges to the generalizability of the experiment that relate to different aspects of the task (Mele, 2014). *First*, the type of volitional act that is investigated in the Libet experiment is very narrow. Participants can choose when to execute a predefined action. Volition here refers only to the timing of the action (when to do something) but not to the content (for an overview of the different aspects of intentional action see Brass and Haggard, 2008). However, there are variants of the Libet task that also investigate content decisions (Haggard and Eimer, 1999; Trevena and Miller, 2002). Another problematic aspect of the task is the instruction that is given, namely to carry out a specific action at a chosen moment in time. Unlike in 'natural choices', in the Libet task participants are instructed to execute a specific action and there are also a number of constraints on when to execute the action (i.e. after a specific interval and within a specific time window). So, the participants' options are very limited (but see Soon et al., 2013 for choices between more abstract intentions). Again, there are very good experimental reasons to strongly limit participants' flexibility in choosing. Computing event related potentials like the RP requires a number of repetitions. Furthermore, asking participants to indicate the time of conscious intention necessarily restricts the degree of spontaneity of the action. Interestingly, there are attempts to address these concerns to which we will return later (e.g. Keller and Heckhausen, 1990).

Second, the choice is completely arbitrary, because there is no plausible reason to act a bit earlier or a bit later. Following the distinction set forth by Ullmann-Margalit and Morgenbesser (1977), this type of selection between moments of action or between options that make no difference to the subject is termed "picking" and is distinguished from "choosing," in which there is a reason for the selection of one of the alternatives. Thus, the situation that is investigated in Libet-style tasks is very different from a 'natural choice' situation, where people have a reason to execute an action at a particular time or a reason to choose one option over another. Moreover, we tend to think of meaningful choices as a more authentic expression of our own agency and free will. Now, one could argue that introducing strong reasons or strong motivations makes it extremely difficult to investigate choice (e.g. Wisniewski et al., 2018; Wisniewski et al., 2015). If participants have a strong motivation to act in a specific way and little or no opposing motivation, does it even make sense to say that they *choose* to act that way? An alternative is to give the options some value, while avoiding making the value of one obviously greater than the value of the other(s) (for such a method see Maoz et al., 2018). The resulting situation may be more similar to real-life situations in which we make choices than is the situation in which participants in Libet-style experiments find themselves. Furthermore, it may be argued that no real decision process is required in Libet-style *picking* situations because there is nothing to decide on. If no real decision process is required, it is not very surprising that the decision unfolds unconsciously and does not involve conscious deliberation (we will return to this point later). In short, the results of the Libet task might strongly depend on the specific requirements of the task, which are not very natural. It is perfectly possible that for many choices that are more meaningful, much conscious deliberation takes place.

The final problem of the Libet task concerns the subjective report or W, which is a crucial element of the task. In everyday life, people usually do not have to indicate when they choose to do something. Very often people have no awareness of exactly when they make their choices. This raises the question to what degree the subjective report influences the choice itself (Keller and Heckhausen, 1990; Lau et al., 2004; Rigoni et al., 2010). This point will be further discussed in the next section.

Taken together, the Libet task investigates intentional action in a very restricted setup that can hardly be generalized to 'natural choices'. And although there might be good reasons for these experimental constraints, tight experimental control strongly reduces the validity of the paradigm and might even determine the type of results one can expect from such a task. In any case, it is crucial to keep these limitations in mind when discussing the implications of the Libet experiment.

3.2. Introspection and timing

One fundamental innovation of the Libet experiment was to provide a way to determine the moment in time when participants become aware of their intention to act. However, there are a number of problems and questions related to this time of intention or W. A first question is whether the fact that participants have to report the time when they become aware of their intention alters the action-generating process itself? Second, can participants accurately report the timing of their conscious intention and is this report really a valid measure of awareness of the intention? Third, this leads to the related question, whether different ways to measure W lead to different results?

3.2.1. Reporting conscious intentions alters processes involved in intentional action

A first question that needs to be addressed is whether the fact that participants are asked to report W alters the processes that are involved in intentional action. There are two ways to address this question. The first approach is to investigate whether the neural signature of intentional action differs in situations where participants act with and without reporting W. Keller and Heckhausen (1990) were the first to address this question. They compared spontaneous movements that occurred without participants consciously intending to move with intentional movements in a Libet-style setup. While they observed an RP for both types of movements, the RP in the Libet task had a larger amplitude and a slightly different scalp distribution. The onset of the RPs did not differ reliably between these situations. One limitation of this experiment was that participants had to answer questions about the intentional nature of their spontaneous actions after each movement. Hence, while no online monitoring was required, the fact that participants had to give subjective reports after each movement, presumably increased awareness of the movements. Miller et al. (2011) directly compared the RP for intentional action with and without reporting W. They found that the amplitude of the RP was largely reduced when participants were not required to report W. In trials without the requirement of reporting W, they even failed to detect a reliable RP. Schlegel et al. (2015) investigated whether the RP differs for actions under hypnotic induction that removed conscious awareness of the action. They found that whether participants were aware of the action or not did not influence the RP. However, in the hypnosis case participants still acted in accordance with the instruction even though they could not report the instruction afterwards. Taken together, the results from these studies suggest that the requirement of actively reporting W changes the brain processes involved in intentional action.

A second approach investigating the influence of directing attention to the intention to act was proposed by Lau et al. (2004). In an fMRI experiment they asked participants to either attend to the intention or to the motor act itself. They found that activation in the SMA/preSMA, a region that is assumed to be involved in the generation of the RP, was amplified when participants attended to the intention rather than the action. Rigoni et al. (2013a) replicated the experiment with EEG. They found that attending to the intention leads to larger RPs than attending to the action itself. Taken together the results suggest that the requirement of the Libet task to report W strongly influences the processes that are involved in intentional action (Keller and Heckhausen, 1990; Lau et al., 2004; Miller et al., 2011; Rigoni et al., 2013a).

3.2.2. The validity of W and how it is measured

So far, we have discussed the influence of reporting W on brain correlates of intentional motor control. In this paragraph, we will discuss the validity of W itself. The basic idea of W is that it reflects the moment in time when we become aware of our intention or urge to act. Participants have to remember the time when they feel the urge and report it after they have implemented their intention. If W indeed reflects the time of conscious intention, this implies that it is solely driven by events that precede our conscious intentions and should be unaffected by the action itself and what follows it. However, this basic assumption has been challenged by a series of behavioural experiments (Banks and Isham, 2009). In the first experiment, participants had to carry out the Libet task. The only difference with the classical manipulation was that after participants had responded, a tone was presented with a variable

delay. As in the original experiment, participants had to indicate W after each trial. In accordance with the predictions of the authors, W was reliably delayed when a tone was presented with a delay. In a second experiment, delayed visual feedback of the response implemented by videotaping the hand also led to a shift of W. These findings strongly suggest that W does not only reflect processes preceding the reported intention but also processes following the action. This however, strongly challenges the validity of W as an indicator of the time when participants first experience the intention to act.

In a follow up study Rigoni et al. (2010) investigated the delayed tone manipulation in an EEG experiment. First, they replicated the findings that delayed feedback led to a shift of W. Furthermore, they demonstrated that delayed feedback resulted in a modulation of the feedback related negativity (Nae), an ERP component that is locked to the effect of the response. The amplitude of the Nae correlated with the shift in W. These findings indicate that processes related to processing the action affect W. Another study suggesting that W is influenced by action or post-action related processes was carried out by Lau et al. (2007). They showed that stimulating the SMA/preSMA simultaneously with or briefly after the action execution leads to a small shift in W.

Taken together, these findings seem to indicate that W is influenced by events that are related to action execution or even feedback processing (Banks and Isham, 2009; Rigoni et al., 2010). This indicates that W is not only determined by intention formation but also by a retrospective reconstruction of events. However, one also has to mention that the shifts of W that are induced in these experiments are rather small. Furthermore, there is evidence that W is

related to intention formation (e.g. Schurger, 2018). It is therefore doubtful whether such influences could explain the temporal gap between the onset of the RP and W.

A different way to address the validity of W is to investigate whether the method by which the time of conscious intention is obtained has an influence on the results. Banks and Isham (2011) implemented a small change in the measurement of W by introducing a digital clock rather than an analogue one. Interestingly, they observed a substantial change in W by using a digital clock. A more fundamental change in the measurement of W was introduced by Matsuhashi and Hallett (2008). Rather than ask participants to indicate the moment in time when they became aware of the intention to act after each trial, they randomly probed whether participants were consciously thinking about the movement or not at a given moment in time. If participants indicated that they were thinking about the movement when they were probed, they had to cancel the action. Matsuhashi and Hallett (2008) used the distribution of the interval between the probe and the action to infer the time of thought (T). If the probe is presented before participants think about the action, participants wait for the intention to arise and then act. If, however, the probe is presented while they are thinking about the action, this should not lead to action and thus these intervals do not occur in the distribution. If, however, the probe is presented too close to the action, although participants are aware of thinking about their next action, nevertheless they are not able to stop the action and will press the key. Matsuhashi and Hallett (2008) argue that the so-called time of thought (T), which indicates the dip in the time distribution when participants already thought about the action and therefore inhibited the action, provides an alternative measure for W. The results revealed a T that was about 1.4 sec before the response, hence much earlier than W in the classical Libet experiment. Given that the RP started about 2 sec before the response, T

was still significantly later than the onset of the RP. However, in a large proportion of participants (almost 30%) the RP did not start earlier than T. The results of Matsuhashi and Hallett (2008) leave open the possibility that T (their measure of the time of conscious intention) is much closer to the onset of the RP than W. While the averaged data still indicate that T comes later than the RP, in a substantial number of participants T preceded the onset of the RP. Therefore, one could argue that this observation strongly compromises the general interpretation of the Libet experiment by suggesting that at least some people consciously decide first and then show an RP. However, one crucial question is whether T really reflects the time of conscious intention or something that precedes it. Matsuhashi and Hallett (2008) did not instruct participants to attend to the intention to act but rather to indicate when they start to think about the next movement. Such thinking does not necessarily imply that participants formed the intention to move. In any case, if such small differences in the instruction can substantially change the timing of the conscious intention, this certainly compromises the validity of the timing procedure.

Here it is important to briefly refer to the study by Soon et al. (2008) which demonstrated that with fMRI measurements and multivariate analyses techniques it is possible to predict intentional action about 8 seconds before participants become aware of their intention to act. Such a temporal delay between the prediction and the conscious awareness of the intention can hardly be explained by the objections against W that we have outlined above. However, there are a number of aspects of this study that need to be considered. First, the very long prediction interval was only possible to achieve by asking participants to implement a very long decision interval. Participants were only allowed to respond about every 14 seconds. This, however, makes the already artificial Libet task even more artificial. If participants are not

explicitly instructed, they will presumably decide within a few hundred milliseconds. Second, as mentioned above, the prediction accuracy is relatively low (less than 10 % above chance). While one could argue that such a low prediction accuracy has a methodological explanation and might be substantially improved in the future, it seems to point more to a decision bias that was decoded rather than a prediction of the decision (Bode et al., 2014; Brass et al., 2013). This point is further supported by a behavioural study which achieved similar prediction accuracies by simply basing their predictions on the response sequence (Lages and Jaworska, 2012). While it is very impressive that a reliable prediction is possible seconds before participants become aware of their decision, the study of Soon et al. (2008) can hardly be used to support the type of argumentation against the common intuition of free will that we discuss in the current review. While the results might show that the decision is already biased seconds before the time of conscious intention, it cannot support the view that the decision is fully determined seconds in advance.

3.3. What does the readiness potential reflect?

After discussing W as a measure of the time of conscious intention, we will now turn to the second crucial element of the Libet experiment, namely the RP as a measure of preconscious decision and motor preparation. Ever since the discovery of the RP by Kornhuber and Deecke (1964) more than 50 years ago, there has been an ongoing debate about the functional role of the RP in motor preparation (Jahanshahi and Hallett, 2003; Shibasaki and Hallett, 2006). While it is outside the scope of the current article to review this literature, we will briefly discss findings that are relevant for the interpretation of the Libet experiment.

3.3.1. The RP as a neural marker of voluntary motor preparation

Kornhuber and Deecke (1964) developed a method to average EEG activity preceding voluntary action. They found that voluntary action is preceded by a 'slowly increasing surface-negative potential' reflecting the readiness (Bereitschaft) to act. Therefore, they called the ERP component the 'Bereitschaftspotential' (readiness potential, RP). They argue that the RP increases with attention and intentional engagement and is reduced by 'mental indifference'. The *RP* has an onset of about 1000 to 2000 ms before movement onset. In addition to the RP there is a later movement-related component, the so calle lateralized readiness potential (Eimer, 1998, LRP) which has an onset of about 500 to 300 ms before movement onset (Shibasaki and Hallett, 2006). While the RP has a fronto-central scalp distribution, the LRP shows a clear lateralization contralateral to the moving hand. It has been demonstrated that the RP and the LRP have different neural generators and also different functional properties (Lang, 2003). A number of factors have been found to affect the RP, such as attention to the movement, the mode of movement selection (self-paced versus externally triggered) and motivational factors. The classical interpretation of the RP is that it reflects the readiness to act indicating a form of abstract motor preparation (Kornhuber and Deecke, 1964). However, the RP has also been related to the intention to act, to resource mobilization, effort and timing of movements (Jahanshahi and Hallett, 2003).

As outlined above, Libet's basic assumption was that the RP reflects the result of an unconscious decision process leading to motor preparedness (Libet et al., 1983). At the moment of W, participants become aware of these unconscious processes. From this

perspective, one would assume that there is a correlative relationship between the RP and W. As we will outline below, this assumption has been challenged empirically (Haggard and Eimer, 1999; Schlegel et al., 2015). A second implicit assumption is that such an unconscious preparation process would necessarily lead to action if not prevented by a conscious act of vetoing. The question whether an RP necessarily leads to action if not consciously vetoed has been addressed by Furstenberg et al. (2015a). They discuss the possibility of a non-conscious, non-executed proximal intention (Furstenberg, 2014). Finally, the ultimate question is whether the RP reflects an unconscious decision or motor preparation process at all or whether it is related to random fluctuations in cortical excitability (Schurger et al., 2012).

3.3.2. How does the RP relate to the conscious intention to act?

One basic assumption of the classical interpretation of Libet's experiment is that the RP reflects an unconscious decision or motor preparation process of which participants become aware at W. If this simple assumption holds true, one would expect that the onset and the amplitude of the RP should affect W: the earlier the onset of the RP, the earlier the onset of W relative to the RP. Haggard and Eimer (1999) developed a variant of the Libet task in which participants were instructed to carry out a response of the left or right hand whenever they decided to do so. In one condition, participants could choose to use the left or right hand and in the other condition they were cued in advance which hand to use. Furthermore, participants were instructed to indicate the point in time when they consciously decided to carry out the response (W). This experimental design combines the classical decision when to act that is required in the Libet task with a decision which action to execute. With this manipulation, Haggard and Eimer (1999) could compute two types of preparation-related

potentials that presumably reflect the two decision components, namely the RP and the lateralized RP (LRP). As briefly outlined above, the LRP occurs later than the RP (i.e. about 300 ms before the response) and has a scalp distribution contralateral to the response hand (Eimer, 1999). The LRP has been related to specific programming of the response. First, the authors tested whether choosing between a left or a right response affected the RP and/or W. This was not the case. Then they split the chosen responses into those with an early and a late W based on the W-response interval. They reasoned that if W reflects the event of becoming aware of the unconscious motor preparation process that is indexed by the RP, trials in which participants indicate their intention earlier (relative to the motor response) should show an earlier onset or a stronger increase of the RP. This, however, was not the case. Instead, the RP was highly similar for early and late W trials (see also a replication of this null finding by Schlegel et al., 2013; but see Schurger, 2018 for a different reasoning). Interestingly, however, they found that the LRP showed a correlation with W. For early W trials, the LRP showed an earlier onset than for late W trials. This finding, however, could not be replicated (Schlegel et al., 2013). The RP data seem to suggest that W does not reflect the event of becoming aware of an unconscious preparation or decision process as indexed by the RP. This conclusion has severe implications for the interpretation of the Libet experiment. While Libet argued that our conscious intentions are a consequence of unconscious processes that are reflected in the RP, the results of Haggard and Eimer (1999) suggest that our conscious intentions are independent of the RP and therefore it is not very likely that W reflects the event of becoming aware of a formerly unconscious decision, at least as measured in the RP.

Another way to investigate whether the RP is related to W is to test whether variables that modulate the RP also lead to a change of W. If there is a strong relationship between the RP

and W, any variable that leads to a modulation of the onset or amplitude of the RP should also affect W. A study by Rigoni et al. (2011) manipulated free will beliefs and found a modulation of the RP by this high-level belief manipulation. Participants who were primed with disbelief in free will showed a reduced RP compared with participants who were primed with a text that was not related to free will. Interestingly, however, this manipulation did not affect W, again questioning the idea that W is directly related to the RP. Taken together, these studies call into question the assumption that W is directly related to the RP. However, as we will outline below, more complex models of the relationship of W and the RP might provide evidence for such a relationship.

3.3.3. Does the RP reflect averaging of stochastic fluctuations of neural activity?

The most serious challenge to the classical interpretation of the Libet experiment stems from the hypothesis that the RP is related to stochastic neural fluctuations rather than to motor preparation (Schurger et al., 2012). A seminal study by Schurger et al. (2012) investigated this old idea using more recent modelling techniques of decision processes. A basic assumption of this research is that the Libet task does not differ substantially from other decision-making tasks. Decision-making in other domains, such as the perceptual domain, has been formalized in models that assume evidence accumulation to a threshold or integration-to-bound (ITB) processes (e.g. Gold and Shadlen, 2007). Such models are considered to be a strategy for overcoming the unavoidable noise that accompanies the evidence signal (references in Miller & Katz 2013; “Introduction”). When participants have to decide, for instance, whether a dynamic random dot motion pattern moves to the left or to the right, the decision system accumulates evidence for these options until a specific threshold is reached and a decision is

made. While in perceptual decision-making, perceptual evidence can be accumulated, in the Libet task participants are explicitly instructed not to base their spontaneous action on any external event. This raises the question what type of evidence is accumulated in such paradigms.

One possible solution to this problem is to treat stochastic noise in the motor system as evidence for the accumulation process (Schurger et al., 2012). In other words, when the evidence for a certain decision is weak or ambiguous, the threshold crossing is mainly determined by subthreshold neuronal noise. In order to make such a model work, the decision threshold has to be set very low to ensure that it is sometimes crossed by random neural fluctuations. When this happens, participants press the key. Schurger et al. (2012) investigated whether such a simple ITB model can explain the shape of the RP and the distribution of decision times (the time from the start of the trial until participants execute the action). First, the authors fitted a drift-diffusion model to the decision times. A drift-diffusion model provides a mathematical formalization of an evidence accumulation process towards a threshold by considering the stochastic nature of the accumulation (Ratcliff, 1978). Then they used the parameters derived from this model to fit the shape of the RP. This led to an excellent fit, suggesting that the drift-diffusion model derived from the behavioural decision times nicely predicts the RP.

In a second experiment, Schurger at al. (2012) wanted to directly test the hypothesis that the decision process is related to stochastic neural fluctuations. They argued that if participants get an interrupting signal to act at a specific point in time, they will act slowly if this go signal occurs when the random fluctuation is far from the decision threshold and quickly when it is

close to the decision threshold. One way to test this hypothesis is by sorting decision times into fast and slow and then plot the RP for these two types of trials. Fast decision times should show a large RP, because neural activity close to the decision threshold is averaged, whereas slow decision times should result in a small RP because neural activity that is averaged is far away from the threshold. This is what they observed. Schurger et al. (2012) interpret the crossing of the threshold as the decision to move now, which then triggers a cascade of motor related processes. This interpretation differs substantially from the classical interpretation of the Libet task which assumes that the RP indexes the preparation to move and is the consequence of a decision process of which participants become aware at W (Schurger et al., 2016). Thus, while an ITB interpretation of the results assumes that the decision happens close to the moment when we become aware of the decision (at W), Libet assumed that the decision happens 500 ms before W.

Interestingly, a recent study used a similar logic to explain neural activity in the secondary motor cortex (M2) of rats using an inter-temporal choice task (Murakami et al., 2014). Rats could either get a small reward after a short waiting-interval or a large reward after a long waiting-interval. Rats could skip waiting for the reward by pressing a lever. Waiting times showed a large variability. Neural activity in rat M2 cortex (which is assumed to be the rat homologue to human premotor cortex) showed ramping up activity very similar to the RP. A simple ITB model could explain the correlation of neural activity and waiting times. Thus, there are interesting parallels to the Libet task. Like in the Libet task, behaviour is not determined by an external signal but rather by internal states. Similar to the study by Schurger et al. (2012), noise in the motor system plays a crucial role in explaining inter-trial variability in neural activity. Hence, decision making without perceptual evidence seems to be highly similar in

non-human animals and humans (see also Khalighinejad et al. (2018) for a similar study in humans).

To investigate how an ITB model relates to subjective timing of consciousness, Kang et al. (2017) used W in a perceptual choice task. They hypothesized that the subjective experience of making up one's mind occurs when the accumulation of evidence reaches a termination threshold. From this perspective, W reflects the moment of the decision rather than a post hoc inference or arbitrary report. In their study, they asked participants to make perceptual decisions about the net direction of dynamic random dot motion. Participants had to use the Libet clock to indicate the time of conscious decision after each trial. The authors fitted a drift-diffusion model to the decision times (here the decision time is the interval between the start of the trial and W). They argued that if W reflects the time when the evidence accumulation process reaches the threshold, the choice proportion should be accurately predicted from such a model. The model that used W as the time when the decision threshold was crossed accurately predicted the choice data. The idea of W corresponding to the point of threshold crossing is also supported by Schurger (2018).

To summarize, simple ITB models of decision making can explain decision times and neural activity in Libet-style tasks. These models assume that decision time in the Libet task is based on a process of accumulation of evidence to a threshold, just like in other decision-tasks. Because the decision is not based on perceptual or other external evidence, this accumulation of evidence might operate primarily on stochastic neural fluctuations in the motor system (Schurger et al., 2012). The moment in which participants become aware of their decision might reflect the point in time when the accumulator crosses the decision threshold (Kang et al., 2017; Schurger, 2018). This conceptualization of the Libet task has important implications

for the interpretation of the RP and W. It would suggest that W does not reflect the moment when participants become aware of a decision that has been made hundreds of milliseconds earlier but rather reflects the moment when a decision process reaches the decision threshold, in other words when the decision is taken. This is consistent with the assertion of Mele (2009) that the onset of the RP does not reflect the proximal decision or proximal intention but rather a part of the causal process that leads to the proximal decision or proximal intention (perhaps the onset of a decision process) which is only formed at W. Such an interpretation of W, however, would be much closer to our intuition of conscious will, namely that we decide at the moment we become aware of our decision (whether such an interpretation is compatible with the idea of conscious free will, will be discussed below). Furthermore, the difference between Libet-style experimental situations and classical perceptual decision tasks would primarily lie in the fact that evidence accumulation in such arbitrary situations is based on stochastic fluctuations of neural activity rather than on accumulation of evidence based on perceptual information. Finally, such unconscious neural activity as the RP would not necessarily lead to a specific behavior, but processes that influence the accumulation of evidence can lead to a change of the behavior until the last moment.

3.3.4. Evidence for preconscious changes of an unconscious decision process

This last point was recently demonstrated by showing that unconscious changes of a preconscious decision process can occur (Furstenberg, 2014). Furstenberg et al. (2015a)

designed an EEG paradigm in which participants were instructed to either choose an action or to execute a cued action. Importantly, an action tendency was induced by subliminal primes that were presented unbeknownst to participants. Previous research has demonstrated that subliminal primes can induce an action tendency in the observer, as indicated by an LRP (Eimer and Schlaghecken, 2003). Furstenberg et al. (2015a) investigated whether such an unconscious action tendency necessarily leads to the primed action or whether participants can choose the response alternative. In principle, one would expect that the primed response tendency would be executed. However, Furstenberg et al. (2015a) demonstrated that in a substantial number of trials, participants switched to the non-primed response even though the primed response tendency was clearly visible in the LRP. In other words, even though the subliminal prime induced a response tendency as indicated by the LRP, participants sometimes overrode this tendency and carried out another response. Importantly, participants were not aware of the primed response tendency in the first place and therefore also not aware that they overrode this tendency. Furstenberg (2014) interpreted these results as an indication of a non-conscious, non-executed intention. With respect to the Libet task, the results suggest that an RP and even an LRP does not necessarily lead to the prepared action but can be unconsciously modified. This interpretation is compatible with an ITB model of choice as outlined above. Any influence that affects the accumulation of evidence before the decision threshold is reached can tip the decision in another direction. Such an influence can be part of stochastic fluctuations (Furstenberg et al., 2015b; Schurger et al., 2012), or an internal process such as the retrieval of an action sequence that biases participants to execute a specific response (Bode et al., 2014), or an external cue that primes the alternative response.

Taken together, these studies place the Libet task in a broader context of decision making and motor control tasks (see Bode et al., 2014 for a similar reasoning; Roskies, 2010a; Schurger et al., 2016) and challenge the basic assumptions underlying the classical interpretation of the task. From this perspective, intentional action in Libet-style experiments is decision making without clear evidence for the decision. Like action models that explain how we respond to our environment, such decision processes might be based on processing different types of internal and external information such as stochastic noise, response biases induced by environmental stimuli or by specific action sequences.

3.4. Locating the conscious intention in the Libet task

We have so far discussed the formation of a proximal conscious intention either as a post-hoc realization of a decision process that happened before or as crossing a decision threshold. However, nearly 30 years ago, Keller and Heckhausen (1990) had discussed the possibility that a conscious intention is formed at the beginning of the experiment when the instructions are given (Mele, 2009). Here the basic idea is that the instruction that is given in the beginning of the experiment leads participants to form intentions how to carry out the task. In other words, the instructions induce a kind of metacognitive strategy. Such conditional conscious intentions might strongly influence how participants perform the task. Consider someone, Al, who hits upon the following strategy as a participant in Libet's study. He will say "now!" silently to himself from time to time, flex right away in response to the silent speech act, and then report where the hand was on the clock when the time for reporting arrives. During the experiment, Al has a conditional intention with the following content: If I silently say "now!", I flex my right wrist at once. We are not saying that this is the only intention he has during the

experiment, of course. Nor are we saying what determines when Al says “now!”. Based on the instructions, participants focus their attention on internal events leading them to detect a normally unconscious process. This detection of an internal signal serves as a triggering event for the implementation of the previously formed intention. Such a mechanism is strongly reminiscent of the concept of implementation intentions (Gollwitzer and Schaal, 1998) or the concept of the ‘prepared reflex’ (Exner, 1879; Hommel, 2000).

Here the basic idea is that participants consciously set themselves in a state of preparedness in the beginning of each experimental trial. Once this state of preparedness is reached, the action can be automatically triggered by a specified stimulus and unfolds with minimal intentional effort (Brass et al., 2017; Cohen-Kdoshay and Meiran, 2009; Hommel, 2000). In Libet-style situations, the triggering event is not an external stimulus but an internal event that is detected through introspection (e.g. an introspectable consequence of stochastic noise reaching a threshold in the motor cortex). This internal event then triggers the proximal intention to act. Combined with an ITB model of choice, the experimental instruction configures a decision process that unfolds outside the awareness of participants. When a specific threshold is crossed, the proximal decision is made and the proximal intention formed in that act of decision making leads to action if not vetoed. Such an interpretation would assume two forms of intentions: A first intention to respond when a specific internal signal occurs. This is a conditional distal intention (Mele, 1992; Pacherie, 2008), and it is a conscious intention. The second, more proximal intention, occurs when the predefined conditions of the implementation intention are met. This second, proximal intention can reach consciousness when required by the instruction but can also arise without consciousness.

4. The veto idea

The research that we have discussed so far strongly questions the classical interpretation of the Libet task. But what about Libet's veto idea? Is there enough time between W and action execution to veto actions that we have decided on? This question is relevant independent of whether one follows the interpretation of Libet with respect to the RP and W. Rather, the same question arises in the context of ITB models of voluntary action. If W reflects the moment in time when an accumulation process crosses the decision threshold which then kicks off the execution of the motor response (Kang et al., 2017), the question arises whether at this point in time it is still possible to stop the action.

For Libet, the veto process was central to maintain the idea of free will. He assumed that one crucial function of becoming aware of formerly unconscious decisions is to be able to veto these decisions at the last moment. From an ITB perspective, the crucial question is whether the motor act can still be prevented after the decision threshold is crossed. Libet's evidence for this veto idea, however, was relatively weak (Mele, 2009). He tested his idea by asking participants to respond at a specific position of the clock hand but inhibit this response on some trials. His results showed a build-up of the RP which then flattens out. Interestingly, despite the extensive research that his classical experiment generated, there is not much research on the veto process. Only recently, have researchers started to address this question systematically (Brass and Haggard, 2007; Kuhn et al., 2009; Schultze-Kraft et al., 2016; Walsh et al., 2010).

In an fMRI study, Brass and Haggard (2007) asked participants to carry out a Libet-style task. However, they instructed participants to sometimes veto the action at the last moment. Hence, participants could decide whether to execute the action or veto it at the last moment. Brass and Haggard (2007) found that a specific brain region in the medial prefrontal cortex was more active in trials where participants vetoed the action at the last moment compared to trials where they simply executed the action. They interpreted their findings as evidence for intentional inhibition which is conceptually similar to Libet's veto idea: participants stop the action after they become aware of the proximal intention to act. Interestingly, the brain regions they observed for intentional inhibition was different from brain areas that were usually found in classical stop signal task (Aron et al., 2014) assuming that intentional inhibition was more about distancing oneself from an intention to act rather than simply stopping the motor action (Ridderinkhof et al., 2014). One crucial criticism of this experiment relates to the question whether participants really prepared an action before they veto it (for a detailed citicism of the experiment see Mele, 2009).

In a follow-up study, a different design was used where a strong action tendency was induced that participants then had to veto (Kuhn et al., 2009). A similar pattern of brain activation was observed (but see Schel et al., 2014 for problems to replicate these findings). However, the inferior temporal resolution of fMRI does not allow one to conclude that the veto process really took place after participants became conscious of their intention to act. In another follow-up study, Walsh et al. (2010) used EEG to investigate intentional inhibition with a method that has a better temporal resolution, using the design of Brass and Haggard (2007). The results of this study indicate that participants indeed initiate the inhibition process around the time of W. However, the question still remains whether the intention to veto the ongoing

action is only formed after participants become aware of the proximal intention or whether the possibility to veto the action tendency influences the formation of the proximal intention to act.

A completely different approach to investigating Libet's veto idea is based on an online detection of neural activity preceding voluntary action (Bai et al., 2011; Schneider et al., 2013; Schultze-Kraft et al., 2016). Advanced methods of data analysis and increased computational power make it possible to online predict voluntary behaviour based on EEG measures (Bai et al., 2011). If one can detect an upcoming action online, the question is until which point in time participants are able to veto this action tendency?

Recently, based on a similar idea, Schultze-Kraft et al. (2016) used a classification algorithm to online predict voluntary action from EEG recordings to directly test until which moment in time participants are able to veto the predicted action. In this study, participants were playing a game against a brain-computer interface (BCI) that tried to predict their actions. Participants had to press a key after a green light was turned on and before a red light was turned on. The BCI system predicted their key press and turned on the red light. If participants managed to press the key before the red light went on, they won. If they pressed the key after the red light was turned on they lost. In all other cases, nobody won. The experiment was divided into three phases. In the first phase, the red light was turned on at random moments after the green light. This phase was used to train the BCI system to predict upcoming actions. In the second phase, the BCI system online predicted the action and turned on the red light when it predicted an action. While the prediction was not extremely accurate, it was sufficient for the purpose of the present experiment. By comparing the second phase with the first phase one

can test whether the BCI system was accurately predicting the action and until which moment in time participants were able to inhibit the action. The third phase was identical to the second phase, but this time participants were explicitly told that the BCI system was predicting their behaviour. By comparing the second phase with the third phase one can test whether participants can use strategies to mislead the BCI system when they are explicitly aware of being predicted. The first result of this study was that the BCI system was able to predict the upcoming action and turn on the red light before participants executed their action in a substantial number of trials. Furthermore, the BCI system prevented participants from carrying out a button press before the red light was turned on in a large number of trials (a reduction of almost 40% compared to the random prediction). Interestingly, however, there was also a substantial number of trials where participants cancelled their button press at the last moment, when muscular activity was already detected by EMG, indicating that cancelling an action is possible until the last moment. Furthermore, there were hardly any cases where participants moved later than 200 ms after the BCI had predicted their behaviour, suggesting that when participants are made aware of an action tendency they need less than 200 ms to cancel it (about 130 ms in Matsuhashi and Hallett, 2008). A control condition, where predictions were made but no red light was turned on demonstrated that the BCI could predict the button press until 1000 ms before the response. However, such early predictions never lead to action when revealed to participants.

Where does that leave us with respect to Libet'sveto idea? Are 200 ms enough to veto an action after becoming aware of the intention to act? Furthermore, how does the veto decision originate? The data of Schultze-Kraft et al. (2016) suggest that the time of conscious intention comes too late to allow participants to veto the action after conscious intention, given that

the point of no return is about 200 ms before action, which corresponds to W. One has to consider, however, that the veto decision in the Libet task is not based on an external signal that needs to be processed but rather on an internal process. If one assumes that stop-signal tasks usually overestimate the stop-signal reaction time (Skippen et al., 2019), there might be still about 100 ms for such an internal decision. This corresponds to the point-of-no return in Matsuhashi and Hallett (2008). Of course, the question is whether this is sufficient to initiate a veto decision. There is, however, an alternative way to conceptualize such a veto decision within the context of the ITB model of intentional action. This alternative interpration assumes that the veto decision might be influenced by post-decisional evidence accumulation that has been demonstrated to influence decision certainty (Kiani and Shadlen, 2009; Pleskac and Busemeyer, 2010) and might reflect a kind of change-of-mind bound (Resulaj et al., 2009). Here the idea is that the accumulation process does not stop when the decision threshold is reached but rather continues for a restricted amount of time. Such a post-decisional evidence accumulation process might form the basis for a veto decision. Alternatively, one can also assume two thresholds (Schurger, 2018). The first threshold is crossed at W and is an advanced warning threshold. The second threshold is the main activation threshold and indicates the point of no return.

5. Where are we now?

When Libet carried out his experiments on free will about 35 years ago, his research kicked off a debate that continues to be very lively. His observation of an RP that precedes the time of conscious intention by a few hundred milliseconds has motivated philosophers, neuroscientists and psychologists to question the existence of free will on empirical grounds.

Interestingly, however, these findings have not only inspired scientists but also science journalists and the general public. In the current article, more than three decades of empirical research on free will are reviewed. This review indicates that Libet's findings are reliable and can be replicated. However, this review also indicates that the interpretation and conclusions that Libet drew from his results are on very shaky ground. First, the literature shows that W is not a very reliable measure of the time of conscious intention. Furthermore, the relationship between the RP and W is not very well understood. Most importantly, however, recent findings suggest that the RP most likely does not reflect an unconscious preparation process following an unconscious decision that necessarily leads to action if not vetoed. Rather, recent findings indicate that the RP might reflect a decision process primarily based on stochastic fluctuations in the motor system that leads to action when an arbitrary threshold is crossed. The 'decision' is not made unconsciously hundreds of milliseconds before the action but only very shortly before we act. In the final part of this review we will re-conceptualize Libet's results based on the recent decision-making literature and the idea that conditional intentions configure such a process. Furthermore, we will integrate the findings we reported above with such a decision-making model of intentional action, and compare them to cases in which intentional action might be impaired, such as in psychiatric and neurological disorders.

5.1. Summarizing and re-conceptualizing Libet's results as a decision process without perceptual evidence

An updated interpretation of the Libet task needs to accommodate three major elements. First, the idea that decisions in Libet-style tasks are based on a general decision process as formalized in ITB models of decision making. Second, that such decision processes are

configured by the task instruction and conditional intentions that participants form at the beginning of the experiment. And finally, the decision to act can be vetoed in a short time window after the decision threshold is reached and before the point of no return.

As outlined above, the basic assumption of such an ITB model of intentional action is the idea that choices about when and what to do are based on decision-making processes that are not fundamentally different from other decision processes in an interesting respect (see Bode et al., 2014; Roskies, 2010a; Schurger et al., 2016 for a similar approach). Similar to perceptual decision-making, information for different options is accumulated until a specific threshold is crossed. In contrast to perceptual decision making, however, the accumulation of evidence is not based on perceptual information but on internal information and stochastic neural activity. In such a model, the RP is a neural index of the continuous integration of information and stochastic neural activity. Importantly, the integration process is not the consequence of a decision but the basis for the decision. The decision is only made when the threshold is crossed. Furthermore, the decision process might integrate other kinds of information that prime a specific response option such as sequential information (Bode et al., 2012) or unconscious perceptual primes (Mattler and Palmer, 2012). Such information could be integrated in the model by assuming a bias for a specific response option at the beginning of the trial or by influencing the evidence accumulation process. Importantly, in principle any kind of information can influence the decision before the threshold is reached. This means that the RP and the LRP do not reflect a ballistic process that necessarily leads to action but rather a gathering of evidence. While the crossing of the decision threshold can in principle occur outside the awareness of participants, the Libet task requires participants to indicate

when they become aware of their intention. In such a context, W might indicate the crossing of the decision threshold (Kang et al., 2017).

Importantly, the decision process is configured by conscious conditional intentions that participants form at the beginning of the experiment as a result of the task instructions. These conditional intentions determine which factors influence the decision process. First, they determine which type of information is accumulated. In the Libet task, evidence is accumulated over internal signals that serve as triggering conditions for the proximal intention. Second, they determine the threshold setting. Depending on the instruction, a fixed threshold can be implemented or a collapsing decision threshold (Mattler and Palmer, 2012) which would ensure that a decision based on stochastic fluctuations can be reached within a given time window. Finally, there is a short time window after the decision threshold is reached where participants can still veto the action. This veto process is presumably based on a change-of-mind bound that is crossed based on post-decisional evidence accumulation. For simplicity, we will refer to this model as the **co**nditional **in**tention and integration **to b**ound (COINTOB) model (see Figure 1).

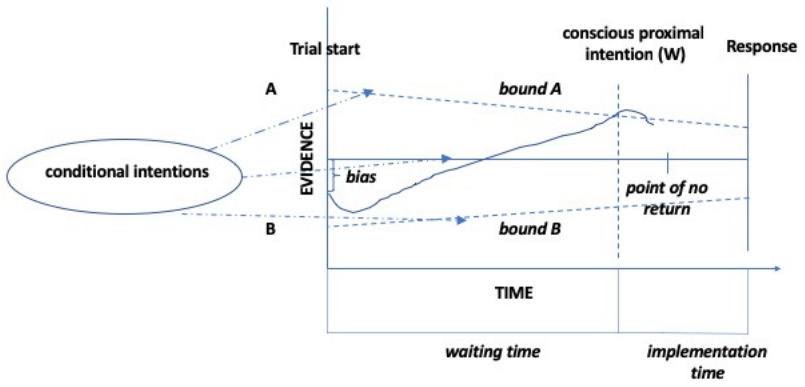

Figure 1: Schematic drawing of the COINTOB model for a choice between two response alternatives (A and B). First a conditional intention is formed that reflects participants' understanding of the task instruction. This conditional intention configures the integration to bound (ITB) process that leads to the choice. The bias reflects an initial preference for one response alternative. The ITB process is furthermore determined by the accumulation rate and the bound. In the example above, evidence first accumulates towards the non choosen response option. The choice is made when the accumulation process crosses the bound. This is also around the moment when participants become aware of their choice (W). The time between the beginning of the trial and W is the waiting time. The time between W and the response is the implementation time. A veto process can be implemented after participants become aware of their intention and before the point of no return. Such a veto process might be driven by a change-of-mind bound that is crossed based on post-decisional evidence accumulation.

Below we will first discuss how consistent this model is with existing research on the Libet task and whether it can account for important empirical findings. Then we will discuss how

consistent the COINTOB model is with the intuition of free will that we outlined in the beginning of this article.

5.1.1. Ecological validity of the Libet task

According to the COINTOB model of intentional action, the only difference between an arbitrary picking decision, as in the Libet task, and a meaningful decision, such as when to hit the brakes if you see pedestrians crossing a street, is related to the kind of information that is accumulated. In the case of the Libet task, it is primarily stochastic fluctuations in the motor system. In the traffic example, it is mainly perceptual evidence. Hence, such a model blurs the alleged distinction between an endogenous action and a stimulus-guided action (Krieghoff et al., 2011), because the only difference between both is whether internal information or perceptual information is integrated. Furthermore, it also blurs the distinction between arbitrary *picking* decisions and meaningful *choosing* decisions since they both express an interaction between *bias-signals* (signals resulting from perceptual evidence, values etc.) and stochastic fluctuations (Furstenberg et al., 2015b). When the bias-signal is weak or ambiguous, the threshold crossing is mainly determined by subthreshold neuronal noise, thus creating a continuum between these possibilities. From this perspective, the criticism that the Libet task is not a good model of choice is not valid. The COINTOB model of choice would assume that while the information on which the decision is based differs between ecologically more or less valid situations, the decision mechanism itself is the same. However, as we will outline below, it is relevant for our intuition of free will which type of information is accumulated. Finally, the COINTOB model of intentional action can easily deal with different decision situations, namely

the decision when to execute an action, the decision which action to execute and the decision whether to execute an action at all (Brass and Haggard, 2008).

5.1.2. The validity of W

As we have outlined above, there are a number of observations regarding W that are difficult to explain when interpreting W as the point in time when participants become aware of an unconscious decision. But how well can such findings be accommodated by an ITB model of the task? The first question is how the COINTOB model of intentional action accounts for the influence of attention to intention on the readiness potential (Keller and Heckhausen, 1990; Rigoni et al., 2013a). A plausible explanation is that attending to the intention to act either changes the accumulation rate or the setting of the threshold or both. One potential interpretation of the larger RP for trials in which participants attend to the intention (Rigoni et al., 2013) might be that this changes the accumulation rate, which would result in faster crossing of the threshold. However, given that participants are instructed not always to act immediately, they need to increase the threshold to prevent premature responding. In any case, the parameters of an accumulation model strongly depend on task demands which are specified in the task instruction. If participants have to decide fast, they will apply a lower threshold than if they have time to decide. Hence, it is very plausible that whether participants attend to the decision or not influences the readiness potential.

Another finding that has to be accounted for is the observation that events happening after the response influence W (Banks and Isham, 2009). This observation compromises the classical interpretation of W as the time of conscious intention. This finding is also difficult to

accommodate from the COINTOB perspective. Of course, conceptually, W is different from the moment of 'conscious intention', as the latter denotes the time of the onset of the subject's consciousness or awareness of a proximal intention to act and W denotes the time the subject *believes* to be the 'conscious intention' time when answering the experimenter's question (Mele, 2009). The aim of the experimenter is to create a design in which these two coincide. As we have explained above, in the COINTOB model of the Libet task W might reflect the moment in time when the decision threshold is crossed (Kang et al., 2017). However, our model does not depend on this assumption. Phenomena such as intentional binding (Haggard and Clark, 2003) suggest that temporal judgements about objective events such as motor responses or stimuli can be influenced by events that follow the to be judged event. Therefore, it is perfectly possible that a temporal judgement about an internal event (the crossing of the decision threshold) is influenced in such a way as well.

In addition, the COINTOB model of choice has to account for the lack of a direct relationship between W and the RP. As we have explained when discussing Kang et al. (2017), the COINTOB model can account for the relationship between electrophysiological antecedents of a decision process and the subjective timing of this decision process. Crucially, the model does not try to explain the relationship between RP and the W-response interval (Haggard and Eimer, 1999) but rather the relationship between the RP and the decision time (Kang et al., 2017; Schurger et al., 2012). While the interval between W and the response might be primarily determined by the implementation of the proximal intention rather than the decision process, waiting times (the interval between the beginning of the trial and the W) are presumably related to the decision process (but see Schurger, 2018 for a different conceptualization).

Finally, the COINTOB model of intentional action has to explain why different methods of probing W lead to different results regarding the time of conscious intention. Most importantly, Matsuhashi and Hallett (2008) demonstrated that when they explicitly tested whether participants were consciously thinking about executing the action (thinking time, T), T was shifted to 1.4 seconds before the response. One potential explanation for such a dramatic shift from the perspective of the COINTOB model would be that probing participants directly leads to reports of T before the response threshold is reached. Here it is crucial to keep in mind that in contrast to W, participants do not respond after T. Furthermore, (Libet et al., 1983) reported that on some trials participants were consciously thinking about the action before deciding to act. It might be interesting to test whether the confidence with which participants report T is similar to the confidence with which they report W.

5.1.3. Psychopathological results and interpretation

It is insightful to compare intentional choice as described in this review with cases in which intentional choice is impaired, such as in psychiatric and neurological disorders. The timing of W has been reported to be affected by various neurological and psychiatric conditions. In such cases it is delayed and appears to be close to the movement itself. For instance, parietal patients estimated W later than healthy subjects and closer to the movement (but compare to Lafargue and Duffau, 2008, a failure to replicate; Sirigu et al., 2004). Similar results were found in patients with psychogenic tremor (Edwards et al., 2011), schizophrenic patients (Caspar and Cleeremans, 2015), Gilles de la Tourette syndrome (Ganos et al., 2015; Moretto et al., 2011) and Parkinson's disease (Tabu et al., 2015). Furthermore, studies outside the realm of psychiatric and neurological diseases, such as in healthy subjects with varied impulsivity traits, support the idea that subjects with higher impulsivity scores estimate W

closer to the time of movement and often after the point of no return, thus executing a non-optimal choice (Caspar & Cleeremans, 2015, Rossi et al., 2018). From the COINTOB model perspective these behavioural results can be understood as an inability to self-monitor and detect the crossing of the threshold. What does the RP look like in these impaired W-judgement cases? While the RP is normal in parietal patients judging the time of their movement (M-judgement) in a Libet task, it is poorly detectable in W-judgement trials (Sirigu et al., 2004). Similar results were obtained in patients with lesioned cerebellum (Kitamura et al., 1999). In several pathological conditions abnormal RP was measured: Parkinson, hemiparesis, dystonia, mirror movement (for detailed description see Shibasaki & Hallett, 2006). Various factors can influence the RP by changing the accumulation rate or the setting of the threshold according to the COINTOB model. Thus, lack of flexibility in adjusting the threshold in accordance to a certain context might be the source of such RP abnormalities. Interestingly, the RP shows a greater amplitude in subjects with higher impulsivity scores during W-judgement trials (Caspar and Cleeremans, 2015; Rossi et al., 2018). As explained above (in 5.1.2.) given that participants are instructed not always to act immediately, they need to increase the threshold to prevent premature responding. However, studies show that impulsive individuals perceive time differently and overestimate time intervals (Caspar and Cleeremans, 2015; Wittmann and Paulus, 2008), which in turn may result in an overincreased threshold. In order to fully understand the mechanisms that underlie changes in RP psychopathology it is important, however, not only to look at changes in W but also investigate the waiting time (Schurger, 2018).

Future research needs to investigate more closely how different aspects of the COINTOB model such as the formation of conditional intentions, the setting of the threshold, the

accumulation rate or response biases can explain pathology-related abnormalities in the perceived time of conscious intention and the RP.

5.1.4. Flexibility of unconscious processes

Another challenge to the classical interpretation of the Libet task is the observation that the readiness potential does not seem to reflect a ballistic process that necessarily leads to action if not vetoed (Furstenberg, 2014; Furstenberg et al., 2015a). Rather, motor-related potentials seem to be malleable by internal processes and external events. If the RP is seen as reflecting the result of an internal decision process, this is difficult to accommodate. Whenever the unconscious decision is made, it will unfold until it reaches consciousness and only then the subject has a possibility to veto it. However, if it is seen as reflecting the internal decision processes itself, this malleability is easily explained. As long as a decision threshold is not reached, the process of accumulating evidence can be easily modified. Which factors influence the decision process depends on the task instructions. Such a view is highly compatible with recent models of motor control where different action tendencies compete until one finally wins the competition (Cisek and Kalaska, 2010).

5.2. Worries about free will

But where does the COINTOB model leave us with respect to our intuition that we can freely and consciously decide when to act and which action to execute? First, the main conclusion of Libet that our conscious experience of deciding is illusory because the decision is made unconsciously hundreds of milliseconds before seems to be unfounded. Rather, according to

the ITB model our conscious experience of the decision and the time when the decision is taken seem to be closely related.

Even so, it may be claimed that the COINTOB model of Libet-style tasks also precludes free decision making. We have suggested that, in Libet-style studies, the threshold crossing for proximal decisions or intentions is mainly a product of subthreshold neuronal noise. And, it may be claimed, if noise is doing this work, the decisions are not made freely. However, such a conclusion would be based on the assumption that all decision situations are based on neural noise. How plausible this inference is, depends on how similar decisions to act made by participants in Libet-style studies are to decisions to act of all other kinds. We have pointed out a similarity, namely that the decision process is similar. But a difference is noteworthy too. In the experiments at issue, participants select from options with respect to which they are indifferent (a *picking* selection) – for example, an array of nearby moments for the beginning of pressing the button on the left rather than the button on the right when they do not think that they have been favoring one button over the other lately. But in many cases of real-world decision making, people are far from indifferent about their leading options. In typical cases, when people make a decision about whether to accept or reject a job offer, whether or not to make a bid on a certain house, and so on, their leading options differ from one another in ways that are important to them. In such cases, people often have a wealth of information to mull over. There is no need to depend primarily on subthreshold neuronal noise. Our basic point about the argument under consideration now is that the circumstances surrounding the arbitrary decisions at issue in the studies we have discussed are so different from the circumstances surrounding many decisions that one cannot properly generalize from the proposition that the former decisions are products primarily of subthreshold neuronal noise to the conclusion that all decisions are products primarily of such noise. Indeed, according to

our ITB model of intentional choice, the decision making process usually involves substantive information, and it relies on noise only when substantive information is not available (see Schurger et al., 2012 for a similar reasoning).

6. Conclusions

More than 35 years ago Benjamin Libet published his seminal work on the relationship of conscious intentions and preceding neural activity. He concluded that the brain decides on our behaviour and we are only informed about this decision in retrospect. His findings have been taken as evidence for a kind of neural determinism and have strongly influenced the free will debate in philosophy, psychology and neuroscience. However, a careful review of more than three decades of research on the Libet task shows that this interpretation is highly problematic. Recent research suggests that neural activity preceding conscious reports of intention presumably reflects the decision process itself rather than the results of an unconscious decision. The decision is only taken when a decision threshold is crossed, at which moment participants become aware of their decision. This interpretation, however, is very close to the common idea or feeling that we decide at the moment of conscious intention. Rather than showing that free will does not exist, the Libet task demonstrates that decisions are the result of conscious conditional intentions that configure a decision process that partly unfolds unconsciously. Whether free will exists or not remains an open question.

References

Aron, A.R., Robbins, T.W., Poldrack, R.A., 2014. Inhibition and the right inferior frontal cortex: one decade on. Trends Cogn Sci 18, 177-185.

Bai, O., Rathi, V., Lin, P., Huang, D., Battapady, H., Fei, D.Y., Schneider, L., Houdayer, E., Chen, X., Hallett, M., 2011. Prediction of human voluntary movement before it occurs. Clin Neurophysiol 122, 364-372.

Banks, W.P., Isham, E.A., 2009. We infer rather than perceive the moment we decided to act. Psychol Sci 20, 17-21.

Banks, W.P., Isham, E.A., 2011. Do we really know what we are doing? Implications of reported time of decision for theories of volition, in: Sinnott-Amstrong, W.S., Nadel, L. (Eds.), Conscious will and responsibilty. Oxford University Press, Oxford, UK, pp. 47-60.

Baumeister, R.F., Masicampo, E.J., Dewall, C.N., 2009. Prosocial benefits of feeling free: disbelief in free will increases aggression and reduces helpfulness. Pers Soc Psychol Bull 35, 260-268.

Bobzien, S., 1998. Determinsm and freedom in stoic philosophy. Claredon Press, Oxford.

Bode, S., He, A.H., Soon, C.S., Trampel, R., Turner, R., Haynes, J.D., 2011. Tracking the unconscious generation of free decisions using ultra-high field fMRI. PLoS One 6, e21612.

Bode, S., Murawski, C., Soon, C.S., Bode, P., Stahl, J., Smith, P.L., 2014. Demystifying "free will": the role of contextual information and evidence accumulation for predictive brain activity. Neurosci Biobehav Rev 47, 636-645.

Bode, S., Sewell, D.K., Lilburn, S., Forte, J.D., Smith, P.L., Stahl, J., 2012. Predicting perceptual decision biases from early brain activity. J Neurosci 32, 12488-12498.

Brass, M., Haggard, P., 2007. To do or not to do: the neural signature of self-control. J Neurosci 27, 9141-9145.

Brass, M., Haggard, P., 2008. The what, when, whether model of intentional action. Neuroscientist 14, 319-325.

Brass, M., Liefooghe, B., Braem, S., De Houwer, J., 2017. Following new task instructions: Evidence for a dissociation between knowing and doing. Neurosci Biobehav Rev 81, 16-28.

Brass, M., Lynn, M.T., Demanet, J., Rigoni, D., 2013. Imaging volition: what the brain can tell us about the will. Exp Brain Res 229, 301-312.

Caspar, E.A., Cleeremans, A., 2015. "Free will": are we all equal? A dynamical perspective of the conscious intention to move. Neurosci Conscious 2015, niv009.

Cisek, P., Kalaska, J.F., 2010. Neural mechanisms for interacting with a world full of action choices. Annu Rev Neurosci 33, 269-298.

Cohen-Kdoshay, O., Meiran, N., 2009. The representation of instructions operates like a prepared reflex: flanker compatibility effects found in first trial following S-R instructions. Exp Psychol 56, 128-133.

Dilman, I., 1999. Free will: An historical and philosophical introduction. Routledge, London.

Edwards, M.J., Moretto, G., Schwingenschuh, P., Katschnig, P., Bhatia, K.P., Haggard, P., 2011. Abnormal sense of intention preceding voluntary movement in patients with psychogenic tremor. Neuropsychologia 49, 2791-2793.

Eimer, M., 1998. The lateralized readiness potential as an on-line measure of central response activation processes. Behav Res Meth Ins C 30, 146-156.
Eimer, M., 1999. Facilitatory and inhibitory effects of masked prime stimuli on motor activation and behavioural performance. Acta Psychol (Amst) 101, 293-313.
Eimer, M., Schlaghecken, F., 2003. Response facilitation and inhibition in subliminal priming. Biol Psychol 64, 7-26.
Exner, S., 1879. Physiologie der Grosshirnrinde. Handbuch der physiologie 2, 189-350.
Fried, I., Mukamel, R., Kreiman, G., 2011. Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. Neuron 69, 548-562.
Frith, C.D., Haggard, P., 2018. Volition and the Brain – Revisiting a Classic Experimental Study. Trends in Neurosciences 41, 405-407.
Furstenberg, A., 2014. Proximal intentions, non-executed proximal intentions and Change of intentions. Topoi 33, 13-22.
Furstenberg, A., Breska, A., Sompolinsky, H., Deouell, L.Y., 2015a. Evidence of Change of Intention in Picking Situations. J Cogn Neurosci 27, 2133-2146.
Furstenberg, A., Deouell, L.Y., Sompolinsky, H., 2015b. Change of intention in “picking” situations, in: Mele, A.R. (Ed.), Surrounding free will. Oxford University Press, Oxford, UK, pp. 165-183.
Ganos, C., Asmuss, L., Bongert, J., Brandt, V., Munchau, A., Haggard, P., 2015. Volitional action as perceptual detection: predictors of conscious intention in adolescents with tic disorders. Cortex 64, 47-54.
Genschow, O., Rigoni, D., Brass, M., 2017. Belief in free will affects causal attributions when judging others' behavior. Proc Natl Acad Sci U S A 114, 10071-10076.
Gold, J.I., Shadlen, M.N., 2007. The neural basis of decision making. Annu Rev Neurosci 30, 535-574.
Gollwitzer, P.M., Schaal, B., 1998. Metacognition in action: the importance of implementation intentions. Pers Soc Psychol Rev 2, 124-136.
Haggard, P., Clark, S., 2003. Intentional action: conscious experience and neural prediction. Conscious Cogn 12, 695-707.
Haggard, P., Eimer, M., 1999. On the relation between brain potentials and the awareness of voluntary movements. Exp Brain Res 126, 128-133.
Hommel, B., 2000. The prepared reflex: Automaticity and control in stimulus-response translation., in: Monsell, S., Driver, J. (Eds.), Control of cognitive processes: Attention and performance XVIII. MIT Press, Cambridge, MA, pp. 247-273.
Jahanshahi, M., Hallett, M., 2003. The Bereitschaftspotential. Springer, Boston, MA, pp. 19-34.
Kang, Y.H.R., Petzschner, F.H., Wolpert, D.M., Shadlen, M.N., 2017. Piercing of Consciousness as a Threshold-Crossing Operation. Curr Biol 27, 2285-2295 e2286.
Keller, I., Heckhausen, H., 1990. Readiness potentials preceding spontaneous motor acts: voluntary vs. involuntary control. Electroencephalogr Clin Neurophysiol 76, 351-361.
Khalighinejad, N., Schurger, A., Desantis, A., Zmigrod, L., Haggard, P., 2018. Precursor processes of human self-initiated action. Neuroimage 165, 35-47.
Kiani, R., Shadlen, M.N., 2009. Representation of confidence associated with a decision by neurons in the parietal cortex. Science 324, 759-764.
Kitamura, J., Shabasaki, H., Terashi, A., Tashima, K., 1999. Cortical potentials preceding voluntary finger movement in patients with focal cerebellar lesion. Clin Neurophysiol 110, 126-132.

Kornhuber, H.H., Deecke, L., 1964. Hirnpotentialänderungen beim Menschen vor und nach Willkürbewegungen, dargestellt mit Magnetbandspeicherung und Rückwärtsanalyse. Pflügers Archiev 281, 52.
Krieghoff, V., Waszak, F., Prinz, W., Brass, M., 2011. Neural and behavioral correlates of intentional actions. Neuropsychologia 49, 767-776.
Kuhn, S., Haggard, P., Brass, M., 2009. Intentional inhibition: how the "veto-area" exerts control. Hum Brain Mapp 30, 2834-2843.
Lafargue, G., Duffau, H., 2008. Awareness of intending to act following parietal cortex resection. Neuropsychologia 46, 2662-2667.
Lages, M., Jaworska, K., 2012. How Predictable are “Spontaneous Decisions” and “Hidden Intentions”? Comparing Classification Results Based on Previous Responses with Multivariate Pattern Analysis of fMRI BOLD Signals. Frontiers in Psychology 3.
Lang, W., 2003. Surface recodings of the Bereitschaftpotential in normals, in: M., J., M., H. (Eds.), The Bereitschaftspotetial. Springer, Boston, MA, pp. 19-34.
Lau, H.C., Rogers, R.D., Haggard, P., Passingham, R.E., 2004. Attention to intention. Science 303, 1208-1210.
Lau, H.C., Rogers, R.D., Passingham, R.E., 2007. Manipulating the experienced onset of intention after action execution. J Cogn Neurosci 19, 81-90.
Libet, B., 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. Behavioral and Brain Sciences 8, 529-539.
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 (Pt 3), 623-642.
Maoz, U., Yaffe, G., Koch, C., Mudrik, L., 2018. Neural precursors of decisions that matter — an ERP study of deliberate and arbitrary choice. bioRxiv.
Matsuhashi, M., Hallett, M., 2008. The timing of the conscious intention to move. Eur J Neurosci 28, 2344-2351.
Mattler, U., Palmer, S., 2012. Time course of free-choice priming effects explained by a simple accumulator model. Cognition 123, 347-360.
Mele, A.R., 1992. Springs of action. Oxford University Press, Oxford.
Mele, A.R., 2009. Effective intentions. Oxford University Press, Oxford.
Mele, A.R., 2014. Free: Why science hasn't disproved free will. Oxford University Press, Oxford, UK.
Miller, J., Shepherdson, P., Trevena, J., 2011. Effects of clock monitoring on electroencephalographic activity: is unconscious movement initiation an artifact of the clock? Psychol Sci 22, 103-109.
Monterosso, J., Royzman, E.B., Schwartz, B., 2005. Explaining away responsibility: Effects of scientific explanation on perceived culpability. Ethics Behav 15, 139-158.
Moretto, G., Schwingenschuh, P., Katschnig, P., Bhatia, K.P., Haggard, P., 2011. Delayed experience of volition in Gilles de la Tourette syndrome. J Neurol Neurosurg Psychiatry 82, 1324-1327.
Murakami, M., Vicente, M.I., Costa, G.M., Mainen, Z.F., 2014. Neural antecedents of self-initiated actions in secondary motor cortex. Nat Neurosci 17, 1574-1582.
Nahmias, E., Shepard, J., Reuter, S., 2014. It's OK if 'my brain made me do it': people's intuitions about free will and neuroscientific prediction. Cognition 133, 502-516.
Nichols, S., Knobe, J., 2007. Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions. Noûs 41, 663-685.

Pacherie, E., 2008. The phenomenology of action: a conceptual framework. Cognition 107, 179-217.
Pleskac, T.J., Busemeyer, J.R., 2010. Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. Psychol Rev 117, 864-901.
Pockett, S., Purdy, S.C., 2010. Are voluntary movements initiated preconsciously? The relationships
between readiness potentials, urges, and decisions., in: Sinnott-Armstrong, W., Nadel, L. (Eds.), Conscious Will and Responsibility: A Tribute to Benjamin Libet. Oxford University Press, New York, pp. 34-46.
Ratcliff, R., 1978. A theory of memory retrieval. Psychological Review 85, 59–108.
Resulaj, A., Kiani, R., Wolpert, D.M., Shadlen, M.N., 2009. Changes of mind in decision-making. Nature 461, 263-266.
Ridderinkhof, K.R., van den Wildenberg, W.P., Brass, M., 2014. "Dont" versus "wont": principles, mechanisms, and intention in action inhibition. Neuropsychologia 65, 255-262.
Rigoni, D., Brass, M., Roger, C., Vidal, F., Sartori, G., 2013a. Top-down modulation of brain activity underlying intentional action and its relationship with awareness of intention: an ERP/Laplacian analysis. Exp Brain Res 229, 347-357.
Rigoni, D., Brass, M., Sartori, G., 2010. Post-action determinants of the reported time of conscious intentions. Front Hum Neurosci 4, 38.
Rigoni, D., Kuhn, 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. Psychol Sci 22, 613-618.
Rigoni, D., Wilquin, H., Brass, M., Burle, B., 2013b. When errors do not matter: weakening belief in intentional control impairs cognitive reaction to errors. Cognition 127, 264-269.
Roskies, A.L., 2010a. How does neuroscience affect our conception of volition? Annu Rev Neurosci 33, 109-130.
Roskies, A.L., 2010b. Why Libet's studies don't pose a threat to free will, in: Sinnott-Armstrong, W., Nadel, L. (Eds.), Conscious Will and Responsibility: A Tribute to Benjamin Libet. Oxford University Press, New York, pp. 11-22.
Rossi, A., Giovannelli, F., Gavazzi, G., Righi, S., Cincotta, M., Viggiano, M.P., 2018. Electrophysiological Activity Prior to Self-initiated Movements is Related to Impulsive Personality Traits. Neuroscience 372, 266-272.
Saigle, V., Dubljević, V., Racine, E., 2018. The Impact of a Landmark Neuroscience Study on Free Will: A Qualitative Analysis of Articles Using Libet and Colleagues' Methods AU - Saigle, Victoria. AJOB Neuroscience 9, 29-41.
Sarkissian, H., Amita, C., Felipe, D.B., Joshua, K., Shaun, N., Smita, S., 2010. Is Belief in Free Will a Cultural Universal? Mind & Language 25, 346-358.
Schel, M.A., Ridderinkhof, K.R., Crone, E.A., 2014. Choosing not to act: neural bases of the development of intentional inhibition. Dev Cogn Neurosci 10, 93-103.
Schlegel, A., Alexander, P., Sinnott-Armstrong, W., Roskies, A., Tse, P.U., Wheatley, T., 2013. Barking up the wrong free: readiness potentials reflect processes independent of conscious will. Exp Brain Res 229, 329-335.
Schlegel, A., Alexander, P., Sinnott-Armstrong, W., Roskies, A., Tse, P.U., Wheatley, T., 2015. Hypnotizing Libet: Readiness potentials with non-conscious volition. Conscious Cogn 33, 196-203.
Schneider, L., Houdayer, E., Bai, O., Hallett, M., 2013. What we think before a voluntary movement. J Cogn Neurosci 25, 822-829.

Schultze-Kraft, M., Birman, D., Rusconi, M., Allefeld, C., Gorgen, K., Dahne, S., Blankertz, B., Haynes, J.D., 2016. The point of no return in vetoing self-initiated movements. Proc Natl Acad Sci U S A 113, 1080-1085.

Schurger, A., 2018. Specific Relationship between the Shape of the Readiness Potential, Subjective Decision Time, and Waiting Time Predicted by an Accumulator Model with Temporally Autocorrelated Input Noise. eNeuro 5.

Schurger, A., Mylopoulos, M., Rosenthal, D., 2016. Neural Antecedents of Spontaneous Voluntary Movement: A New Perspective. Trends Cogn Sci 20, 77-79.

Schurger, A., Sitt, J.D., Dehaene, S., 2012. An accumulator model for spontaneous neural activity prior to self-initiated movement. Proc Natl Acad Sci U S A 109, E2904-2913.

Shibasaki, H., Hallett, M., 2006. What is the Bereitschaftspotential? Clin Neurophysiol 117, 2341-2356.

Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., Posada, A., Haggard, P., 2004. Altered awareness of voluntary action after damage to the parietal cortex. Nat Neurosci 7, 80-84.

Skippen, P., Matzke, D., Heathcote, A., Fulham, W.R., Michie, P., Karayanidis, F., 2019. Reliability of triggering inhibitory process is a better predictor of impulsivity than SSRT. Acta Psychol (Amst) 192, 104-117.

Soon, C.S., Brass, M., Heinze, H.J., Haynes, J.D., 2008. Unconscious determinants of free decisions in the human brain. Nat Neurosci 11, 543-545.

Soon, C.S., He, A.H., Bode, S., Haynes, J.D., 2013. Predicting free choices for abstract intentions. Proc Natl Acad Sci U S A 110, 6217-6222.

Tabu, H., Aso, T., Matsuhashi, M., Ueki, Y., Takahashi, R., Fukuyama, H., Shibasaki, H., Mima, T., 2015. Parkinson's disease patients showed delayed awareness of motor intention. Neurosci Res 95, 74-77.

Trevena, J.A., Miller, J., 2002. Cortical movement preparation before and after a conscious decision to move. Conscious Cogn 11, 162-190; discussion 314-125.

Ullmann-Margalit, E., Morgenbesser, S., 1977. Picking and choosing. Social research 44, 757-785.

Vohs, K.D., Schooler, J.W., 2008. The value of believing in free will: encouraging a belief in determinism increases cheating. Psychol Sci 19, 49-54.

Walsh, E., Kuhn, S., Brass, M., Wenke, D., Haggard, P., 2010. EEG activations during intentional inhibition of voluntary action: an electrophysiological correlate of self-control? Neuropsychologia 48, 619-626.

Wisniewski, D., Forstmann, B., Brass, M., 2018. How exerting control over outcomes affects the neural coding of tasks and outcomes. bioRxiv.

Wisniewski, D., Reverberi, C., Tusche, A., Haynes, J.D., 2015. The Neural Representation of Voluntary Task-Set Selection in Dynamic Environments. Cereb Cortex 25, 4715-4726.

Wittmann, M., Paulus, M.P., 2008. Decision making, impulsivity and time perception. Trends Cogn Sci 12, 7-12.

Zhong, J.Y., 2016. What does neuroscience research tell us about about human consciousness? An overview of Benjamin Libet's legacy. Journal of Mind & Behavior 37, 287-310.

Acknowledgements: We would like to thank David Wisniewski, Kobe Desender and Tom Verguts for commenting on a previous version of the manuscript. AF wishes to thank Haim Sompolinsky for his support and for meaningful discussions.