# ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

# Decoding and predicting intentions

## John-Dylan Haynes

Bernstein Center for Computational Neuroscience Berlin, Charité–Universitätsmedizin Berlin, Germany

Address for correspondence: John-Dylan Haynes, Bernstein Center for Computational Neuroscience, Charité–Universitätsmedizin Berlin, Haus 6, Philippstrasse 13, 10115 Berlin, Germany. haynes@bccn-berlin.de

There has been a long debate on the existence of brain signals that precede the outcome of decisions, even before subjects believe they are consciously making up their mind. The framework of multivariate decoding provides a novel tool for investigating such choice-predictive information contained in neural signals leading up to a decision. New results show that the specific outcome of free choices between different plans can be interpreted from brain activity, not only after a decision has been made, but even several seconds before it is made. This suggests that a causal chain of events can occur outside subjective awareness even before a subject makes up his/her mind. An important future line of research would be to develop paradigms that allow feedback of real-time predictions of future decisions to reveal whether such decisions can still be reverted. This would shed light on how tight the causal link is between early predictive brain signals and subsequent decisions.

Keywords: intention; prediction; multivariate decoding; free will; decision

## Introduction

In a seminal experiment, Benjamin Libet and colleagues presented a fundamental challenge to our intuitions about how we make decisions.[1,2] They investigated the temporal relationship between brain activity and a conscious intention to perform a simple voluntary movement.[1,2] Subjects viewed a "clock" that consisted of a light point moving on a circular path rotating once every 2.56 seconds. They were asked to flex a finger at a freely chosen point in time and to remember and report the position of the moving light point when they first felt the urge to move. The reported position of the light could then be used to determine the time when the person consciously formed their intention, a time subsequently called "W," shorthand for the conscious experience of "wanting" or "will." Libet recorded encephalography signals (electroencephalogram (EEG)) from movement-related brain regions while subjects were performing this task (Fig. 1A). It had previously been known that negative deflections of the EEG signal can be observed immediately preceding voluntary movements[3] (Fig. 1B). These so-called readiness potentials (RPs) originate from brain regions involved in motor preparation, primarily sup-

plementary motor cortex (SMA) and premotor cortex, although preparatory signals can also be observed across wider cortical and subcortical regions[4–7] (Fig. 1C). Libet and colleagues were interested in whether the RP might begin to arise even before the person had made up their mind to move. Indeed, they found that the RP already began to rise a few hundred milliseconds before the "feeling of wanting" entered awareness (Fig. 1A). This systematic temporal precedence of brain activity before a freely timed decision was taken as evidence that the brain had made the decision to move *before* this decision entered awareness. It was proposed that the RP reflects the primary cortical site where the decision to move is made.[8]

Due to the far-reaching implications that unconscious brain processes might shape the outcome of seemingly free choices, Libet's groundbreaking experiments immediately met severe criticism.[9–14] According to the philosopher Hume,[15] two empirical criteria are required to argue for a causal relationship between two events, for example, event B (brain) causing event W (will). First, there has to be a *temporal precedence* of B before W, and second there has to be a *constant connection* between events B and W. It has been debated
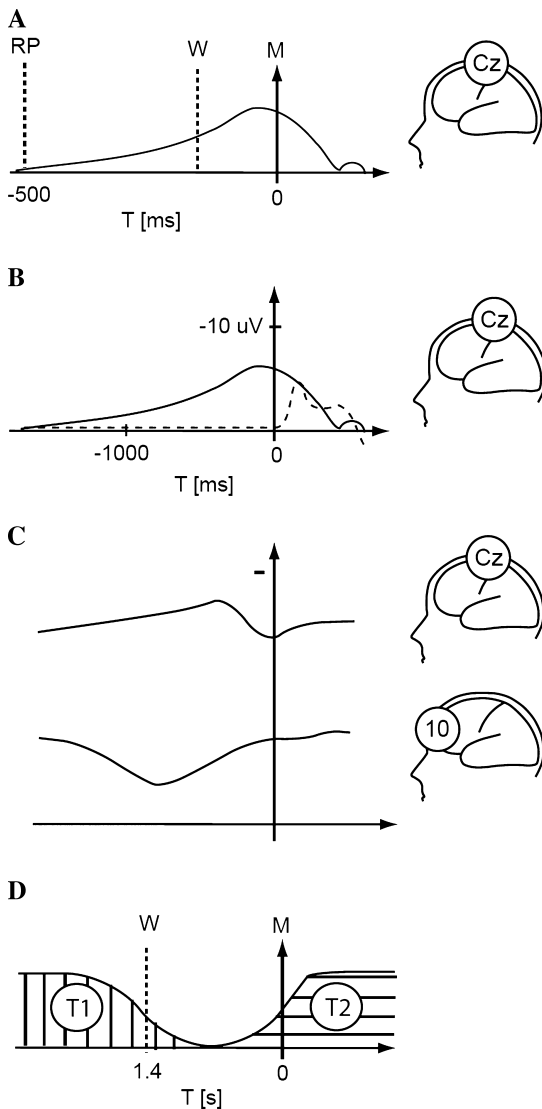
**Figure 1.** Anticipatory EEG potentials predict voluntary movements. (A) The Libet experiment. The onset of the RP precedes the subjective time of decision (W), which in turn precedes the onset of the movement (M) (schematically redrawn from Ref. 1). (B) Recordings over movement-related brain regions (Cz) before a voluntary movement (solid line) versus a stimulus-triggered movement (dashed line). A slow negative potential is visible up to one second only before the voluntary movement, but not before the cued movement (schematically redrawn from Ref. 28). (C) Recordings over Cz and over Brodmann area 10 in frontopolar cortex before a voluntary movement. The RP is preceded by a potential over frontopolar cortex (schematically redrawn from Ref. 5). (D) Based on a novel stop-signal paradigm, it has been recently proposed that W might precede the movement by up to 1.4 seconds. Subjects are exposed to a series of tones. If they had a movement intention when they heard a tone, they were asked to cancel the movement. This reveals a time phase during which the tone is too early and the subject

whether Libet's experiments fulfill either of these criteria.

### Timing

Several authors have questioned whether there is indeed a temporal precedence between RP and intention, in particular, by arguing that the timing judgments are unreliable.[9,13] It has long been known that there are substantial inaccuracies in determining the timing and position of moving objects.[13,16–18] Thus, the choice of a moving light point to report the timing is far from optimal. Similarly, the task to report the onset time of W requires that subjects direct their attention to the onset of their intention, which can potentially provide additional distortions of the true timing compared to a task that requires subjects to report the perceived timing of their movement.[19] By providing incorrect feedback on movement timing, it is possible to further bias the timing judgments by up to 50 m/sec.[20] Based on results from a novel stop-signal paradigm, much longer delay times between conscious will and movement of up to 1.4 seconds have been proposed[21] (see Fig. 1D).

### Constant connection

A different line of arguments addresses the constant connection between B and W. Libet reports data averaged across a number of trials. Although this shows that on average there is an RP before the urge to move, it doesn't show whether this holds for every single trial, which would be necessary to provide evidence for a constant connection. For example, the early onset of the RP might be an artifact of temporal smearing and might reflect only the onset of the earliest urges to move.[22] Alternatively, the early onset RP could be present only in a subset of trials, in which case it could not be considered a *cause* of the subsequent choice. This could only be assessed by measuring the onset time of individual RPs, which is a particularly challenging signal processing problem.[23] An alternative would be to see how well the choice can be predicted from brain signals.

### Other brain regions

A further important shortcoming of Libet's experiment is that it only investigates RPs, which means it

has not yet formed an intention (T1) and a time phase when the tone is too late and the subject has already performed the movement (leftmost section of region T2). In between is a time phase that may correspond to the subjective extension of the experienced will (schematically redrawn from Ref. 21).

is restricted to signals originating from movement-related brain regions. This leaves unclear how other areas might contribute to the build up of a decision. This is particularly important because several other regions of prefrontal cortex have frequently been shown to be involved in free choice situations,[24] although it remains unclear to what degree they are involved in preparing a decision. One shortcoming of RPs as a marker of preparation is that they only emerge in a narrow time window immediately preceding a movement and leave unclear whether they do indeed reflect the earliest stage when a decision is cortically prepared. In fact, it has been argued that the close temporal proximity of RPs and conscious awareness of the urge to move indicates that these two processes are scientifically indistinguishable.[25] Interestingly, even before the original Libet experiments, it has been demonstrated that prefrontal cortex prepares voluntary movements across longer periods than are visible from the RP alone[5] (Fig. 1C). Activity levels and connectivity in an extended network of parietal, medial, and lateral prefrontal regions are increased in free choices,[26–28] although the level of change preceding the conscious decision is not always clear.

*Content specificity*

Another limitation of Libet's experiment is that it involves decisions with an extremely reduced number of degrees of freedom. The subjects only choose the timing of the movement, but they cannot choose between various different movements to make. In a seminal follow-up to Libet's experiment, Haggard and Eimer used a multichoice version of the original experiment.[29] They instructed subjects to choose between making two different movements, one with the left and one with the right hand. Then they examined a content-selective brain signal, the lateralized readiness potential (LRP), a signal recorded over motor cortex that indicates a lateralized hemispheric preparation of movements. Interestingly, they found that it was the LRP, not the classical RP that was consistently related to the timing of subjects' choices.[29] This highlights the importance of separating *specific* from *unspecific* processes in preparing choices.

**The modified Libet experiment**

In order to advance beyond the shortcomings of the Libet experiment, it can be useful to first consider how intentions are coded in the human brain. There is a long history of research on the cortical processing and encoding of intentions. In humans and nonhuman primates it has been repeatedly shown from recordings of single cells, populations of cells, local field potentials, blood-oxygenation levels, and intracranial EEG signals that extended regions of cortex prepare for upcoming movements. These areas include primary motor cortex,[30] SMA and pre-SMA,[3,31,32] dorsolateral prefrontal cortex,[33] posterior parietal cortex,[34,35] and even frontopolar cortex (FPC)[5] (see Fig. 1C). Such preparatory signals can be predictive for more than one movement ahead.[32] It has been proposed based on findings in patients with brain lesions[36] and on brain stimulation[37] that the parietal cortex (PC) might be the site where conscious action plans are generated[38] and where predictive control of movements takes place.[39] Importantly, it has been demonstrated that movements and movement intentions can be decoded with high accuracy from neural population signals[40–42] in primates as well as multichannel intracranial EEG[43] and surface EEG[23] signals in humans. This not only opens up new perspectives on brain–computer interfaces,[41,44–46] but these techniques can also be used to reinvestigate the buildup of intentions using more sensitive methods than were available in the days of Libet. Importantly, such decoding techniques can also be applied to functional magnetic resonance imaging (fMRI) signals that allow for the noninvasive investigation of content-selective processes across large regions in the human brain.[47] This analysis may reveal the prefrontal storage of cued intentions across delay periods[48] (for details, see Fig. 2).

We performed a novel variant of the Libet task[49] using fMRI instead of EEG. The hemodynamic latency of fMRI signals means that it is suitable only for assessing decision-related brain activity across longer timespans. Our focus on longer timespans and the low temporal sampling rate of the fMRI signal enabled us to relax our requirement on temporal precision of the timing judgment, thus overcoming a severe limitation of Libet's original experiments. We replaced the rotating clock with a randomized stream of letters that updated every 500 m/sec. Subjects had to report the letter visible on the screen when they made their conscious decision. This mode of report has the additional advantage of being unpredictable, which minimizes systematic preferences for specific clock positions.
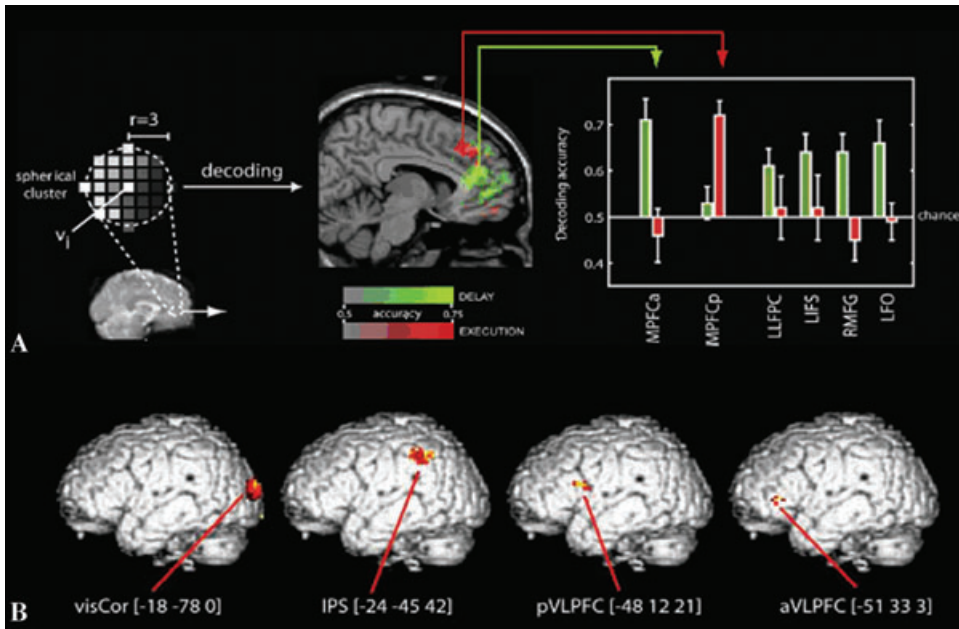
**Figure 2.** Decoding intentions using multivariate pattern recognition. (A) Decoding *self-chosen* intentions:[48] On each trial, subjects freely chose to add or subtract two numbers they were about to see. After a delay, subjects were shown the numbers and could perform the calculation. We used multivariate pattern classifiers to decode which specific intention the subjects were holding. A searchlight classifier[48,52,57] (left) was used to assess how much intention-related information was coded in local spherical clusters of brain activity for each position in the brain. This can be used to plot a map of local information (center). The regions plotted in green code the *prospective intention* during the preparatory period (i.e., after the choice had been made but before the numbers had been shown). The regions plotted in red coded the *ongoing intention* while the calculation was being performed. (B) Decoding *cued* intentions.[57] Subjects were cued to perform one of two possible visual–motor mapping tasks. The sequence from left to right shows the temporal buildup of information about the task cue, starting in visual cortex and then proceeding to parietal and lateral prefrontal cortex.

Subjects were asked to freely decide between two response buttons while lying in an MRI scanner (Fig. 3). They fixated on the center of the screen where the stream of letters was presented. While viewing the letter stream they were asked to relax and freely decide at some point in time to press either the left or right button. In parallel, they were asked to remember and report the letter presented when their decision to move reached their awareness. Importantly, in order to facilitate spontaneous behavior, we did not ask subjects to balance the left and right button selections in successive trials. This would require keeping track of the distribution of button selections in memory and would also encourage preplanning of choices. Instead, we selected subjects that spontaneously chose a balanced number of left and right button presses without prior instruction based on a behavioral selection test before scanning.

We then used a multivariate decoder[47,50–53] to predict how a subject would decide based on his/her brain activity (see Fig. 3). For each time point, we examined the activation preceding the intention and whether a given brain region carried information related to the specific outcome of a decision, that is, the urge to press *either* a left *or* a right button. To understand the advantage of decoding, it can help to review the standard analysis techniques in fMRI studies. Most conventional neuroimaging analyses perform statistical analyses on one position in the brain at a time and then proceed to the next position.[54] This yields a map of statistical parameters that plots how strong a certain effect is expressed at each individual position in the brain and has been used in most previous studies on switching between different intentions.[55,56] However, such analyses do not reveal the coding of *specific* intentions that might be represented in distributed spatial
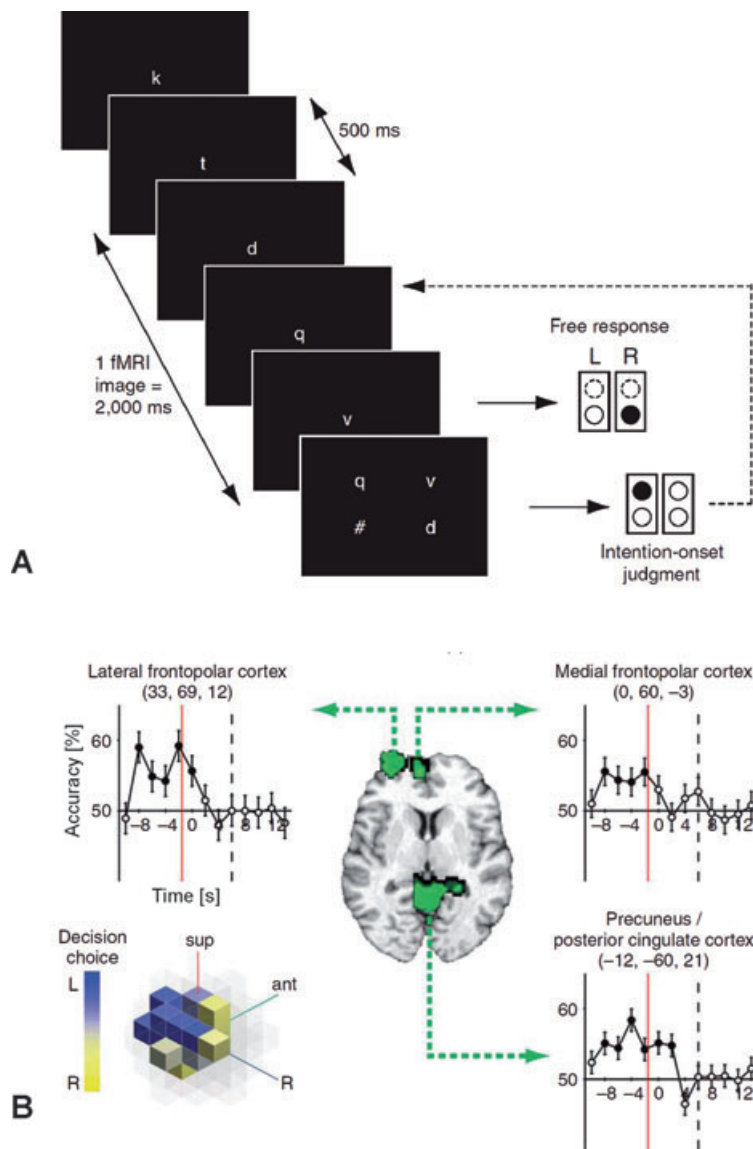
**Figure 3.** (A) The revised Libet task. Subjects are given two response buttons, one for the left and one for the right hand, and view a stream of letters on a screen that changes every 500 m/sec. Subjects are asked to spontaneously press either the left or the right button. Once the button is pressed, they are asked to report which letter was on the screen when they made up their mind. (B) Pattern-based decoding and prediction of decisions ahead of time. Using a searchlight technique,[48,49,52] we assessed for each brain region and each time point preceding the decision whether it is possible to decode the choice before it occurs. Decoding is based on small local spherical clusters of voxels that form three-dimensional spatial patterns. This allowed us to systematically investigate which brain regions had predictive information at each time point preceding the decision. We assessed which brain regions had predictive information about a subject's decision even before the subject knew how they were going to decide. This yielded regions of the frontopolar cortex and precuneus/posterior cingulate cortex, which coded predictive information seven seconds before the decision was made.

patterns of fMRI signals.[48,57,58] It has recently emerged that such fine-grained fMRI patterns contain information that is predictive of the detailed contents of a person's thoughts,[47,50–53] even if the detailed nature of such fine-grained information is still under debate.[59–61] Therefore, we used pattern-based decoding analyses to extract a maximal amount of predictive information contained in the

fine-grained spatial pattern of activity. This information allows one to predict the *specific* choice a subject is going to make on each trial. Such decoding techniques have been successfully used for decoding motor preparation and motor intentions from electrophysiological signals,[23,40–43,46,62] although, in most cases it remains unclear how the onset of such information was related to the awareness of the decision.

To first validate our method, we investigated from which brain regions the specific decision could be decoded *after* it had been made and when the subject was already executing the motor response (Fig. 3). As would be expected, we found that the outcome of the decision was encoded in motor cortex. It is important to note that, as expected, the informative fMRI signals are delayed by several seconds relative to the decision because of the delay of the hemodynamic response. Next, we addressed whether any brain region encoded the subject's decision *ahead* of time. We found that, indeed, two brain regions partially predicted whether the subject was about to choose the left or right response prior to the conscious decision, and even though the subject did not know yet which way he/she was about to decide (Fig. 3). The first region to predict the decision was in the FPC/ Brodmann area 10 (BA10). The predictive information in the fMRI signals from this brain region was present already seven seconds prior to the subject's decision. This period of seven seconds is a conservative estimate that does not yet take into account the delay of the fMRI response with respect to neural activity. Because this delay is several seconds, the predictive neural information will have preceded the conscious decision by up to 10 seconds. A second predictive region was located in PC, stretching from the precuneus into posterior cingulate cortex. Notably, the predictive accuracy in FPC/ BA10 and in PC, though statistically significant, only reached a maximum of 10% above the level of chance, whereas decoding accuracy in motor cortex after the choice reached 25% above chance level.

It is also important to note that there is no *overall* signal increase in the frontopolar and precuneus/posterior cingulate during the preparation period. Rather, the predictive information is encoded in the spatial pattern of fMRI responses, which is presumably why it has only rarely been noticed. Notably, due to the temporal delay of the hemodynamic response the small lead times in

SMA/pre-SMA of up to several hundred milliseconds reported in previous studies[1,29] are below the temporal resolution of our method. Hence, one cannot exclude the possibility that other regions contain predictive information in the short period *immediately preceding* the intention.

### The role of BA10
The finding of unconscious, predictive brain activity patterns in BA10 is interesting because this area is not normally discussed in connection with free choices, although nearby regions of the medial wall have been observed for other free choices.[63] This is presumably due to the fact that conventional analyses will only pick up regions with changes in activity *overall*, but not regions where only the *patterning* of the signal changes in a choice-specific fashion. However, it has been repeatedly demonstrated using other tasks that BA10 plays an important role in the encoding and storage of intentions. It has long been known that lesions to BA10 lead to a loss of prospective memory, thus disrupting the ability to hold action plans in memory for later execution.[64] In a previous study from our group, we have found that BA10 also stores intentions across delay periods after they have reached consciousness, especially if there is a delay between decision and execution[48] (Fig. 2). Although BA10 has only rarely been implicated in preparation of voluntary actions, a direct comparison across different brain regions has revealed that the cortical region exhibiting the earliest preparatory signals before voluntary movements is the FPC.[5] BA10 is also cytoarchitectonically special because it has a very low cell density, but each cell forms a large number of synapses, indicating that it is a highly associative brain region.[65] One could speculate that this would allow for locally recurrent processing that could support the storage of action plans in working memory. Furthermore, BA10 is believed to be the area that has most disproportionately grown in size in humans compared to nonhuman primates.[65]

### Two preparatory circuits: "what" versus "when"
Predictive brain signals are not unusual; instead, they have been demonstrated in a variety of domains and across multiple timescales, ranging from predictive coding in the visual system;[66] movement preparation;[4,7,31] anticipation of future events, locations, and intentions;[48,58,67,68] or the prediction of errors from slow brain signals.[69] Predictive fMRI

signals have to be interpreted carefully,[70] but there is evidence suggesting that they may reflect high-frequency local field potentials.[71,72] The long-term anticipatory dynamics of such free and self-paced decisions seem to be in contrast with the speed of neural processes that accumulate evidence for other types of decisions.[73] Why does it take so long for the decision to be made if it starts to build up several seconds before? Please note that self-paced paradigms such as this one leave the time at which a decision is made up to the subject. This is in contrast to responsive paradigms, where subjects make a speeded decision in response to an external event, as when stopping at the red traffic lights. In the absence of any time pressure for making the decision, it is inappropriate to think of the delay between onset of predictive neural information and the conscious decision as a "reaction time" to the neural information. Thus, multichoice versions of the Libet experiment involve not one, but two decisions to be made.[29,49] On one hand, a decision needs to be made as to *when* to decide; on the other hand, a decision has to be made as to *which* button to choose. Brass and Haggard[74] have referred to this as "when" and "what" decisions. We also conducted a further decoding analysis where we assessed the degree to which the timing of the decision (as opposed to its outcome) can be decoded. The time of conscious intention could be significantly predicted from supplementary motor area (SMA) and pre-SMA.[49] The earliest decodable information for timing was available five seconds before a decision. This might suggest that the brain begins to prepare self-paced decisions through two independent networks that only converge at later stages of processing. The classical Libet experiments, which were primarily concerned with "when" decisions, found short-term predictive information in the SMA. This is compatible with our prediction of the timing from pre-SMA and SMA. In contrast, as our results show, a "what" decision is prepared much earlier and by a much more extended network in the brain.

## Sanity checks

Our findings point toward long-leading brain activity that predicts the outcome of a decision even before the decision reaches awareness. This is a striking finding, and it is important to critically discuss several possible sources of artifacts and alternative interpretations. Moreover, it is necessary to make

sure that the report of the timing is correct, and that the information does not reflect a carryover from previous trials.

### Early decision—late action?

One question is whether the subjects are truly performing the task correctly. For example, a subject might decide early (e.g., at the beginning of the trial) which button to press, and then simply wait for a few seconds to execute his/her response. This could be the case if perhaps the entire group of subjects had grossly disregarded the instructions. A similar argument has already been made against the Libet experiment. It is conceivable that as the decision outcome gradually enters awareness, subjects adopt a very conservative criterion for their report, and wait for the awareness to reach its "peak" intensity.[11,14] Fortunately, there are reasons that make it implausible that subjects waited to report a decision that had already begun to reach awareness. In situations where subjects know which button they are going to press, the corresponding movement is already prepared in the primary motor cortex.[75] In contrast, in our study, the motor cortex contains information only at a very late stage of processing, following the conscious decision of which movement to make. This suggests that subjects did not decide early and then simply wait to respond.

### Carryover from previous trial?

It is also important to discuss whether the early prediction reflects a carryover of information from the previous trial, as would be expected because of the failure of human subjects to generate random sequences.[76,77] There are several reasons to doubt that the information reflects such a spillover between trials.

First, the distribution of response sequences clearly resembles an exponential distribution without sequential order, as would be expected if subjects decide randomly which button to press from trial to trial.[49] Presumably, this is because, in contrast to previous studies, we did not ask subjects to balance left and right button presses across trials, thus encouraging decisions that were independent of previous trials. However, it is important to note that a proper assessment of the randomness of a time series requires a large number of trials that is not always available in behavioral experiments with delays. In our experiments, subjects often took a long time until they made a decision, which might explain why

subjects behaved more randomly than in traditional random choice experiments, where subjects systematically violate randomness when explicitly asked to rapidly generate random sequences.[76,77]

Second, our chosen statistical analysis method, fitting a so-called finite impulse response function, is designed to separate the effects of the current trial from the previous and the following trial. This approach is highly efficient as long as both types of responses are equally frequent, with variable intertrial intervals.

Third, the early onset of predictive information in prefrontal and parietal regions cannot be explained by any trailing BOLD signals from the previous trials. The onset of information occurs approximately 12 seconds after the previous trial, which is far beyond the relaxation time of the hemodynamic response. Additionally, the predictive information increases with temporal distance from the previous trial, which is not compatible with the information being an overlap from the previous trial.

Fourth, time points that overlap with the next trial also revealed no carryover of information.

Taken together, the high predictive accuracy of the activation preceding the decision reflects prospective information encoded in prefrontal and PC that is related to the decision in the current trial.

### Implications for the free-will debate?

The revised Libet study shows that the brain can begin to unconsciously prepare decisions several seconds before they reach awareness. The potential implications of Libet's experiments for free will have been discussed at great length in the literature (for a recent review, see Ref. 78), which has helped sharpen what the contribution of such simple free choice paradigms might be. Obviously, however, they do not address real world decisions that have high motivational importance, they are not based on long-term reward expectations, and they do not involve complex reasoning. Libet's and our decisions have only little motivational salience for the individual and are experienced as random, rather than being based on in-depth trial to trial reasoning. However, Libet's and our findings do address one specific intuition regarding free will, that is, the naive folk-psychological intuition that at the time when we make a decision, the outcome of this decision is free in the sense of not being predetermined by prior brain activity. This intuition is scientifi-

cally implausible anyway, simply because it stands in contradiction to our belief in deterministic laws of physics. However, the direct demonstration that brain activity predicts the outcomes of decisions before they reach awareness has additional persuasive power in convincing people that they are more predictable than they believe to be. Similar dissociations between awareness and motor control have been demonstrated before.[79] What the findings highlight is that a cascade of unconscious brain processes beginning in prefrontal and PC unfolds across several seconds and prepares subjectively free, self-paced decisions.

### Causality?

An important point that needs to be discussed is to what degree the finding of choice-predictive information supports any causal relationship between brain activity and the conscious will. Such causal links have been demonstrated previously by direct cortical stimulation over parietal and frontal cortex.[37,80] However, it is unclear if the early predictive signals are also causally involved in the decision. As for the criterion of *temporal precedence*, there should be no doubt that our data finally demonstrate that brain activity can predict a decision long before it enters awareness. A different point is the criterion of *constant connection*. A constant connection would require that the decision could be predicted with 100% accuracy from prior brain activity. Libet's original experiments were based on averages, so no statistical assessment can be made about the accuracy with which decisions can be predicted. Our prediction of decisions from brain activity is statistically reliable, but far from perfect. The predictive accuracy of around 60% (which is significant, but only 10% above chance) can be improved if the decoding is tailored to each subject. However, even under optimal conditions, this is far from 100% for several reasons. One possibility is that the inaccuracy stems from imperfections in our ability to measure neural signals. Because of the limitations of fMRI in terms of spatial and temporal resolution, it is clear that the information we can measure can only reflect a strongly impoverished version of the information available from a direct measurement of the activity in populations of neurons in the predictive areas. A further source of imperfection is that an optimal decoding approach needs a large (ideally infinite) number of training samples to learn

exactly what the predictive patterns should be. In contrast, the slow sampling rate of fMRI imposes limitations on the training information available. So even if the populations of neurons in these areas would, in principle, allow a perfect prediction, our ability to extract this information would be severely limited. These limitations, however, cannot be used to argue that one day with better methods the prediction will be perfect; this would constitute a mere "promissory" prediction. Importantly, a different interpretation could be that the inaccuracy simply reflects the fact that the early neural processes might only be partially predictive of the outcome of the decision. In this view, even full knowledge of the state of activity of populations of neurons in FPC and in the precuneus might not permit the full prediction of a decision. In that case, the signals have the form of a biasing signal that influences the decision to a degree, but additional influences at later time points might still play a role in shaping the decision. The fact that decoding after the decision from motor cortex can be achieved with higher accuracy might point toward the fact that neural signals in BA10 and in PC are not fully predictive in principle. However, the exact topology of clustering of calls with similar tuning preferences in BA10/PC is, to date, unknown, and thus might turn out to be less suitable for fMRI decoding than in motor cortex.

Until a perfect predictive accuracy has been reached in an experiment, both interpretations—incomplete prediction and incomplete determination—remain possible. Importantly, even a complete, 100% prediction may not directly imply a causal link between the early predictive signals and the choice.

## Future perspectives

An important next step will be to establish whether early predictive signals are decision related at all. This might sound strange given that they predict the choices. However, this early information could hypothetically be the consequence of stochastic, fluctuating background activity in the decision network,[10] similar to the known fluctuations of signals in early visual cortex.[81,82] In this view, the processes relevant for the decision could occur late, perhaps in the last second before the decision. In the absence of any "reasons" for deciding for one option or the other, the decision network might need to break the symmetry, for example, by using stochas-

tic background fluctuations in the network. If the fluctuations in the network are in one subspace, the decision could be biased toward the "left," and if the fluctuations are in a different subspace, the decision could be biased toward the "right." But how could fluctuations at the time of the conscious decision already be reflected seven seconds before? One reason could be that the temporal autocorrelation of neural signals includes very slow fluctuations.[82] In contrast, the slow rise of the hemodynamic response might smear the ongoing fluctuations across time;[83] however, the fMRI signal itself is presumably not causally involved in decision making, as it is only an indirect way of measuring the neural processes leading up to the decision. Nonetheless, in future experiments it is important to further investigate how tightly the early information is linked to the decision. One prediction of the slow background fluctuation model is that the outcome of the decision would be predictable even in cases where a subject does not know that they are going to have to make a decision or where a subject does not know what a decision is going to be about. This would point toward a predictive signal that does not directly computationally contribute to decision making.

It should be emphasized that the question whether prediction reflects a carryover between trials is independent from the question whether it reflects slow fluctuations. Subjects took a long time to make decisions that could have counteracted any autocorrelation in the signals affecting successive trials. Importantly, any slow signals would necessarily have to be content selective in order to be able to be predictive of specific decisions. Thus, a hypothesis could be that the decisions are based on a stochastic process that is predictable across short time scales but not at longer timescales (as say different trials).

A further interesting point for future research is the comparison of self-paced with rapid decisions that occur in response to sudden and unpredictable external events. At first sight it seems implausible that rapid, responsive decisions could be predicted ahead of time. How would we be able to drive a car on a busy road if it always took us a minimum of seven seconds to make a decision? However, even unpredictable decisions are likely to be determined by "cognitive sets" or "policies" that are likely to have a much longer half-life in the brain than mere seven seconds.
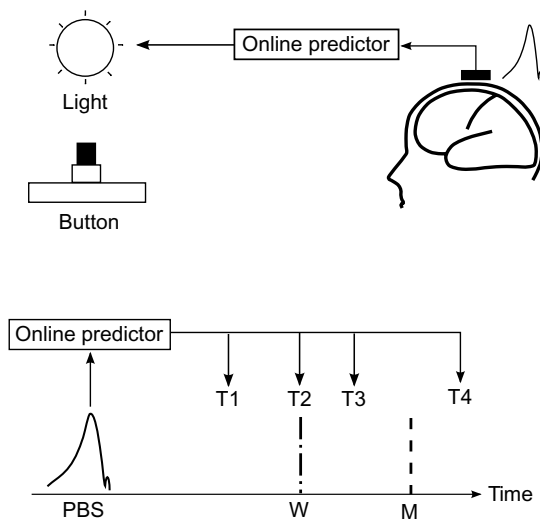
**Figure 4.** Hypothetical "decision prediction machine"[84] that performs online prediction of free decisions based on early predictive brain signals (PBS). (A) The subject is asked to spontaneously press a button unless a light is turned on. An online predictor is then trained to turn on the light as soon as an upcoming decision is predicted from early PBSs, but the subject can't press the button because the light is turned on. (B) The timing at which the prediction is fed back to the subject should have an important effect on the experiences of the subject. If the light is turned on at T1, before the subject has decided to move, the subject might have the impression the light is turning on randomly. If the light is turned on at T4, after the movement is made, the subject will perceive it to be too late to have any relevant effect. If the light is turned on at T2, the subject should have the impression the light turns on exactly at the time their choice was made. If the light is turned on at T3, after the choice but before the movement can be executed, the subject might have the impression that the light turned on just after their choice, but they were unable to cancel their movement. This experiment requires rapid and highly accurate online prediction of upcoming decisions.

Finally, it would be interesting to investigate whether decisions can be predicted in real-time before a person knows how they are going to decide. Such a real-time "decision prediction machine" (DP-machine) would us allow to turn certain thought experiments[12,84] into reality, for example, by testing whether people can guess above chance which future choices are predicted by their current brain signals even though a person might not have yet made up their mind (Fig. 4). Such forced-choice judgments would be helpful in revealing whether there is evidence for subtle decision-related information that might enter a person's awareness at an

earlier stage than would be apparent in the conventional Libet tasks.[12,21]

One experiment could be to ask a person to press a button at a time point of their own choice, with the one catch that they are not allowed to press it when a lamp lights up.[84] Using real-time decoding techniques it might then be possible to predict the impending decision to press the button and to control the lamp to prevent the action. The phenomenal experience of performing such an experiment would be interesting. For example, if the prediction is early enough, the subject is not even aware that they are about to make up their mind and should have the impression that the light is flickering on and off randomly. If the prediction occurs after the decision but before the onset of the movement, the subjects might have the impression that the light turned on just after their choice, but too late for them to stop the movement, similar to stop-signal experiments.[21,85]

It would also be possible to use the DP-machine to inform the subject of their impending decision and get them to "veto" their action and not press a button. This will allow testing to what degree a decision that is predicted by a specific brain signal can still be averted. If a decision can't be averted, this would lend plausibility to the idea of a causal link between the preparatory signal and the conscious decision. Currently such "veto" experiments rely on trusting a person to make up their mind to press a button and then to rapidly choose to terminate their movement.[86] A DP-machine would finally allow one to perform true "veto" experiments. If it were possible to predict not only *when* a person is going to decide, but also *which specific option* they are going to take, one could ask them to change their mind and take the *opposite* option. It seems plausible that a person should be able to change their mind across a period as long as seven seconds. However, there is a catch: how can one change one's mind if one doesn't even know what one has chosen in the first place? As mentioned previously, it is unclear yet whether fMRI or any other neural signal will provide sufficient decoding accuracy to predict decisions before they are made. However, if we were one day able to create a machine that could accurately predict our free decisions,[87] it would allow us to better understand the relationship between our conscious thoughts and our brain activity.

## Box 1

Predicting intentions—five key questions for future research:

(1) Accuracy: What is the highest accuracy with which early signals can predict future decisions? Clarifying this question could reveal how tight the connection is between early signals and the subsequent decision. This might require measurement of intracranial signals during self-paced decision experiments.

(2) Real-time feedback: To what degree can feedback of the prediction alter the subject's behavior? Answering this question could help reveal how flexible people are in overcoming their own predictability when confronted with a technical device that feeds back their choices. Can subjects behave in an unpredictable fashion?

(3) Commitment: This question is closely linked to the feedback experiment. To what degree can the effect of an early signal that biases a decision be avoided or revoked? Up to what point in time preceding a decision can an established plan still be changed?

(4) Slow functions: What role do slow cortical fluctuations play in predicting signals early? Is the time scale of predictive neural signals systematically related to the time scale of their intrinsic autocorrelation? Can such background fluctuations play a role in breaking the symmetry between different choices? Similarly, what is the relationship between the so-called default network or resting state activity and the timepoints for choices in Libet-style experiments with similar parieto-frontal topographies.[88]

(5) Phenomenology: The timepoint of a decision needs to be measured with high accuracy. It is also important to precisely measure what subjects are thinking during the period leading up to the decision to avoid the influence of conscious deliberation. For example, if a subject were oscillating between different choices, but biased toward one option, it might appear that choice-predictive signals precede the conscious decision (although the subject was already leaning toward a specific outcome). For motor signals, this can be ruled out by monitoring any buildup of motor commands in primary motor cortex.[75]

But for complex decisions, such signatures are more difficult to find.

## Acknowledgments

## Conflicts of interest

The author declares no conflicts of interest.

## References

1. Libet, B., C.A. Gleason, E.W. Wright & D.K. Pearl. 1983. Time of conscious intention to act in relation to onset of cerebral activities (readiness-potential): the unconscious initiation of a freely voluntary act. *Brain* **106:** 623–642.

2. Libet, B. 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav. Brain Sci.* **8:** 529–566.

3. Kornhuber, H.H. & L. Deecke. 1965. Hirnpotentialänderungen bei Willkürbewegungen und passiben Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Arch. Ges. Phys.* **284:** 1–17.

4. Colebatch, J.G. 2007. Bereitschaftspotential and movement-related potentials: origin, significance, and application in disorders of human movements. *Mov. Disord.* **22:** 601–610.

5. Groll-Knapp, E., J.A. Ganglberge & M. Haider. 1977. Voluntary movement-related slow potentials in cortex and thalamus in man. *Progr. Clin. Neurophysiol.* **1:** 164–173.

6. Sochurkova, D., I. Rektor, P. Jurak & A. Stancak. 2006. Intracerebral recording of cortical activity related to self-paced voluntary movements: a bereitschaftspotential and even-related desynchronization/synchronization SEEG study. *Exp. Brain Res.* **173:** 637–649.

7. Rektor, I. 2002. Scalp-recorded Bereitschaftspotential is the result of the activity of cortical and subcortical generators: a hypothesis. *Clin. Neurophys.* **13:** 1998–2005.

8. Eccles, J.C. 1982. The initiation of voluntary movements by the supplementary motor area. *Arch. Psychiatr. Nervenkr.* **231:** 423–441.

9. Breitmeyer, B.G. 1985. Problems with the psychophysics of intention. *Behav. Brain Sci.* **8:** 539–540.

10. Eccles, J.C. 1985. Mental summation: the timing of voluntary intentions by cortical activity. *Behav. Brain Sci.* **8:** 542–543.

11. Latto, R. 1985. Consciousness as an experimental variable: problems of definition, practice and interpretation. *Behav. Brain Sci.* **8:** 545–546.

12. Marks, L.E. 1985. Toward a psychophysics of intention. *Behav. Brain Sci.* **8:** 547–548.

13. Van de Grind, W. 2002. Physical, neural, and mental timing. *Conscious. Cogn.* **11:** 241–64.

14. Ringo, J.L. 1985. Timing volition: questions of what and when about W. *Behav. Brain Sci.* **8:** 550–551.

15. Hume, D. 1777. *An Enquiry Concerning Human Understanding*. Reprinted as Harvard Classics, Collier & Son. New York, NY.

16. Moutoussis, K. & S. Zeki. 1997. Functional segregation and temporal hierarchy of the visual perceptive systems. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **264:** 1407–1414.

17. Rollman, G. B. 1985. Sensory events with variable central latencies provide inaccurate clocks. *Behav. Brain Sci.* **8:** 551–552.

18. Wundt, W. 1904. *Grundzüge der physiologischen Psychologie.* Engelmann. Leipzig. 2.

19. Lau, H.C., R.D. Rogers & R.E. Passingham. 2006. On measuring the perceived onsets of spontaneous actions. *J. Neurosci.* **26:** 7265–7271.

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

21. Matsuhashi, M. & M. Hallett. 2008. The timing of the conscious intention to move. *Eur. J. Neurosci.* **28:** 2344–2351.

22. Trevena, J.A. & J. Miller. 2002. Cortical movement preparation before and after a decision to move. *Consc. Cogn.* **11:** 162–190.

23. Blankertz, B., G. Dornhege, C. Schäfer, R. Krepki, *et al*. 2003. Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. *IEEE Trans. Neural Syst. Rehabil. Eng.* **11:** 127–131.

24. Deiber, M.P., R.E. Passingham, J.G. Colebatch, *et al*. 1991. Cortical areas and the selection of movement: a study with positron emission tomography. *Exp. Brain. Res.* **84:** 393–402.

25. Merikle, P.M. & J. Cheesman. 1985. Conscious and unconscious processes: same or different? *Behav. Brain Sci.* **8:** 547–548.

26. Lau, H.C., R.D. Rogers, P. Haggard & R.E. Passingham. 2004. Attention to intention. *Science* **303:** 1208–1210.

27. Pesaran B., M.J. Nelson, R.A. Andersen. 2008. Free choice activates a decision circuit between frontal and parietal cortex. *Nature* **453:** 406–409

28. Jahanshahi, M., I.H. Jenkins, R.G. Brown, *et al*. 1995. Self-initiated versus externally triggered movements. I. An investigation using measurement of regional cerebral blood flow with PET and movement-related potentials in normal and Parkinson's disease subjects. *Brain* **118:** 913–933.

29. Haggard, P. & M. Eimer. 1999. On the relation between brain potentials and the awareness of voluntary movements. *Exp. Brain Res.* **126:** 128–133.

30. Georgopoulos, A.P., A.B. Schwartz & R.E. Kettner. 1986. Neuronal population coding of movement direction. *Science* **233:** 1416–1419.

31. Tanji, J. 2001. Sequential organization of multiple movements: involvement of cortical motor areas. *Annu. Rev. Neurosci.* **24:** 631–651.

32. Nakajima, T., R. Hosaka, H. Mushiake, & J. Tanji. 2009. Covert representation of second-next movement in the pre-supplementary motor area of monkeys. *J. Neurophysiol.* **101:** 1883–1889.

33. Quintana, J. & J.M. Fuster. 1999. From perception to action: temporal integrative functions of prefrontal and parietal neurons. *Cereb. Cortex* **9:** 213–221.

34. Snyder, L.H., A.P. Batista & R.A. Andersen. 1997. Coding of intention in the posterior parietal cortex. *Nature* **386:** 167–170.

35. Toni, I., N.J. Shah, G.R. Fink, *et al*. 2002. Multiple movement representations in the human brain: an event-related fMRI study. *J. Cogn. Neurosci.* **14:** 769–784.

36. Sirigu, A., E. Daprati, S. Ciancia, *et al*. 2004. Altered awareness of voluntary action after damage to the parietal cortex. *Nat. Neurosci.* **7:** 80–84.

37. Desmurget, M., K.T. Reilly, N. Richard, *et al*. 2009. Movement intention after parietal cortex stimulation in humans. *Science* **324:** 811–813.

38. Desmurget, M. & A. Sirigu. 2009. A parietal-premotor network for movement intention and motor awareness. *Trends Cogn. Sci.* **13:** 411–419.

39. Haggard, P. 2005. Conscious intention and motor cognition. *Trends Cogn. Sci.* **9:** 290–295.

40. Averbeck, B.B., J.W. Sohn & D. Lee. 2006. Activity in prefrontal cortex during dynamic selection of action sequences. *Nat. Neurosci.* **9:** 276–282.

41. Andersen, R.A. & H. Cui. 2009. Intention, action planning, and decision making in parietal-frontal circuits. *Neuron* **63:** 568–583.

42. Wessberg, J., C.R. Stambaugh, J.D. Kralik, *et al*. 2000. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* **408:** 361–365.

43. Miller, K.J., G. Schalk, E.E. Fetz, *et al*. 2010. Cortical activity during motor execution, motor imagery, and imagery-based online feedback. *Proc. Natl. Acad. Sci. USA* **107:** 4430–4435.

44. Lebedev, M.A. & M.A. Nicolelis.2006. Brain-machine interfaces: past, present and future. *Trends Neurosci.* **29:** 536–546.

45. Schwartz, A.B. 2004. Cortical neural prosthetics. *Annu. Rev. Neurosci.* **27:** 487–507.

46. Birbaumer, N. 2006. Breaking the silence: brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology* **43:** 517–532.

47. Haynes, J.D., & G. Rees. 2006. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7: 523–534.

48. Haynes, J.D., K. Sakai, G. Rees, *et al*. 2007. Reading hidden intentions in the human brain. *Curr. Biol.* **17:** 323–328.

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

50. Haynes, J.D. & G. Rees. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat. Neurosci.* **8:** 686–691.

51. Kamitani, Y. & F. Tong. 2005. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* **8:** 679–685.

52. Kriegeskorte, N., R. Goebel & P. Bandettini. 2006. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. USA* **103:** 3863–3868.

53. Norman, K.A., S.M. Polyn, G.J. Detre & J.V. Haxby. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn. Sci.* **10:** 424–430.

54. Friston, K.J., A.P. Holmes, J.B. Poline, *et al.* 1995. Analysis of fMRI time-series revisited. *Neuroimage* **2:** 45–53.

55. Brass, M. & D.Y. von Cramon. 2004. Decomposing components of task preparation with functional magnetic resonance imaging. *J. Cogn. Neurosci.* **16:** 609–620.

56. Dove, A., S. Pollmann, T. Schubert, *et al.* 2000. Prefrontal cortex activation in task switching: an event-related fMRI study. *Brain Res. Cogn. Brain Res.* **9:** 103–109.

57. Bode, S. & J.D. Haynes. 2009. Decoding sequential stages of task preparation in the human brain. *Neuroimage* **45:** 606–613.

58. Hampton, A.N. & J.P. O'Doherty. 2007. Decoding the neural substrates of rewardrelated decision making with functional MRI. *Proc. Natl. Acad. Sci. USA* **104:** 1377–1382.

59. Swisher, J.D., J.C. Gatenby, J.C. Gore, *et al.* 2010. Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *J. Neurosci.* **30:** 325–330.

60. Kamitani, Y. & Y. Sawahata. 2010. Spatial smoothing hurts localization but not information: pitfalls for brain mappers. *Neuroimage* **49:** 1949–1952.

61. Op de Beeck, HP. 2010. Probing the mysterious underpinnings of multi-voxel fMRI analyses. *Neuroimage* **50:** 567–571.

62. Andersen, R.A., E.J. Hwang & G.H. Mulliken. 2010. Cognitive neural prosthetics. *Annu. Rev. Psychol.* **61:** 169–190.

63. Forstmann, B.U., M. Brass, I. Koch & D.Y. von Cramon. 2006. Voluntary selection of task sets revealed by functional magnetic resonance imaging. *J. Cogn. Neurosci.* **18:** 388–398.

64. Burgess, P.W., A. Quayle & C.D. Frith. 2001. Brain regions involved in prospective memory as determined by positron emission tomography. *Neuropsychologia* **39:** 545–555.

65. Ramnani, N. & A.M. Owen. 2004. Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat. Rev. Neurosci.* **5:** 184–195.

66. Rao, R.P. & D.H. Ballard. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2:** 79–87.

67. Schacter, D.L., D.R. Addis & R.L. Buckner. 2008. Episodic simulation of future events: concepts, data, and applications. *Ann. N.Y. Acad. Sci.* **1124:** 39–60.

68. Lisman, J. & A.D. Redish. 2009. Prediction, sequences and the hippocampus. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364:** 1193–1201.

69. Eichele, T., S. Debener, V.D. Calhoun, *et al.* 2008. Prediction of human errors by maladaptive changes in event-related brain networks. *Proc. Natl. Acad. Sci. USA* **105:** 6173–6178.

70. Sirotin, Y.B. & A. Das. 2009. Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. *Nature* **457:** 475–480.

71. Handwerker, D.A. & P.A. Bandettini. 2010. Hemodynamic signals not predicted? Not so: a comment on Sirotin & Das (2009). *Neuroimage* doi: 10.1016/j.neuroimage. 2010.04.037.

72. Logothetis, N.K., J. Pauls, M. Augath, *et al.* 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412:** 150–157.

73. Gold, J.I. & M.N. Shadlen. 2007. The neural basis of decision making. *Annu. Rev. Neurosci.* **30:** 535–574.

74. Brass, M. & P. Haggard. 2008. The what, when, whether model of intentional action. *Neuroscientist* **14:** 319–325.

75. Mars, R.B., M.G. Coles, W. Hulstijn & I. Toni. 2008. Delay-related cerebral activity and motor preparation. *Cortex* **44:** 507–520.

76. Nickerson, R.S. 2002. The production and perception of randomness. *Psych. Rev.* **109:** 330–357.

77. Wagenaar, W.A. 1972. Generation of random sequences by human subjects: a critical survey of literature. *Psych. Bull.* **77:** 65–72.

78. Roskies, A.L. 2010. How does neuroscience affect our conception of volition? *Annu. Rev. Neurosci.* **33:** 109–130.

79. Fourneret, P. & M. Jeannerod. 1998. Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia* **36:** 1133–1140.

80. Fried, I., A. Katz, G. McCarthy, *et al.* 1991. Functional organization of human supplementary motor cortex studied by electrical stimulation. *J. Neurosci.* **11:** 3656–3666.

81. Arieli, A., A. Sterkin, A. Grinvald & A. Aertsen. 1996. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* **273:** 1868–1871.

82. Leopold, D.A., Y. Murayama & N.K. Logothetis. 2003. Very slow activity fluctuations in monkey visual cortex: implications for functional brain imaging. *Cereb. Cortex* **13:** 422–433.

83. Cordes, D., V.M. Haughton, K. Arfanakis, *et al.* 2001. Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data. *Am. J. Neuroradiol.* **22:** 1326–1333.

84. Chiang, T. 2005. What's expected of us. *Nature* **436:** 150.

85. Logan, G.D., W.B. Cowan & K.A. Davis. 1984. On the ability to inhibit simple and choice reaction time responses: a model and a method. *J. Exp. Psychol. Hum. Percept. Perform.* **10:** 276–291.

86. Brass, M. & P. Haggard. 2009. To do or not to do: the neural signature of self-control. *J. Neurosci.* **27:** 9141–9145.

87. Feigl, H. 1958. *The "Mental" and The "Physical"*. University of Minnesota Press. Minneapolis, MN.

88. Gusnard, D.A., M.E. Raichle & M.E. Raichle. 2001. Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci.* **2:** 685–694.