

Spontaneous Decisions and Free Will: Empirical Results and Philosophical Considerations

JOANA RIGATO,¹ MASAYOSHI MURAKAMI,² AND ZACHARY MAINEN²

¹*Center for Philosophy of Sciences of the University of Lisbon, Faculty of Sciences, 1749-016, Lisbon, Portugal*

²*Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown, 1400-038, Lisbon, Portugal*

Correspondence: zmainen@neuro.fchampalimaud.org; mjrigato@fc.ul.pt

Spontaneous actions are preceded by brain signals that may sometimes be detected hundreds of milliseconds in advance of a subject's conscious intention to act. These signals have been claimed to reflect prior unconscious decisions, raising doubts about the causal role of conscious will. Murakami et al. (2014. *Nat Neurosci* 17: 1574–1582) have recently argued for a different interpretation. During a task in which rats spontaneously decided when to abort waiting, the authors recorded neurons in the secondary motor cortex. The neural activity and relationship to action timing was parsimoniously explained using an integration-to-bound model, similar to those widely used to account for evidence-based decisions. In this model, the brain accumulates spontaneously occurring inputs voting for or against an action, but only commits to act once a certain threshold is crossed. The model explains how spontaneous decisions can be forecast (partially predicted) by neurons that reflect either the input or output of the integrator. It therefore presents an explicit hypothesis capable of rejecting the claim that such predictive signals imply unconscious decisions. We suggest that these results can inform the current debate on free will but must be considered with caution.

In classical experiments on self-initiated actions in humans, Kornhuber and Deecke (1965) discovered a “readiness potential” that could be recorded from the scalp electroencephalogram (EEG). In brain areas such as the supplementary motor area (SMA), these signals appear as slowly ramping activity starting as much as 1 sec in advance of a voluntary movement. Benjamin Libet et al. (1983), in a now famous follow-up of this work, showed that the agent's conscious awareness of an intention to act occurs much later—only 200 msec before the act itself (Fig. 1). Although these experiments have been criticized on empirical grounds, including the ability to accurately measure the timing of conscious awareness (van de Grind 2002), similar results by other researchers have revealed both single neuron and fMRI signals substantially preceding spontaneous movements (Haggard and Eimer 1999; Lau et al. 2004; Soon et al. 2008; Fried et al. 2011), in some cases by several seconds (Soon et al. 2008).

This line of experiments (which we will hereafter refer to as “Libet-type” experiments) has led to significant attention and debate. Libet originally interpreted the experiments as implying that decisions are made at an unconscious level and only later do they become available to the conscious self. Given the usual demand that a free decision be made consciously, this interpretation seems to pose a grave threat to free will. But it has been greeted with diverse opinions and arguments. Libet himself argued that free will existed but only in the possibility of vetoing the action in that short gap between awareness and movement (Libet 1985). In the field of social psychology it was used to defend the claim that conscious

will is an illusion (Wegner 2002). In philosophy, the tendency has been to diminish the relevance of Libet's work, even by philosophers who are skeptical about free will (Dennett 1991; Levy 2014). Among those who defend the existence of free will, Mele (2009, 2014a,b) has questioned the claims attributed to Libet's experiments, arguing that there is no proof that the readiness potentials are identical to an intention to act and that the type of decisions involved in these experiments is not representative of the kinds of important decisions about which one should care about free will.

Neuroscience may have something to contribute to this debate, as it has been studying the neural basis of decision-making for more than a decade (Gold and Shadlen 2007). The psychology and neurobiology of simple perceptual decisions such as discrimination tasks (Roitman and Shadlen 2002; Wang 2002; Mazurek et al. 2003), reaction time tasks (Hanes and Schall 1996), or simple value-based choices such as choosing one of two food items (Krajbich et al. 2010) have converged to a common if not consensual mechanistic hypothesis. According to this hypothesis, decisions are the result of the accumulation of evidence. Evidence may correspond to sensory signals or to more abstract variables such as subjective value. When a certain evidence threshold or bound is reached, the decision for the corresponding choice is made. Choices may correspond to observable actions or commitments to act in the future. Formal mathematical accounts of this sort are referred to as “integration-to-bound” models.

Spontaneous decisions, like starting to move out of one's own initiative, have been less studied than evi-

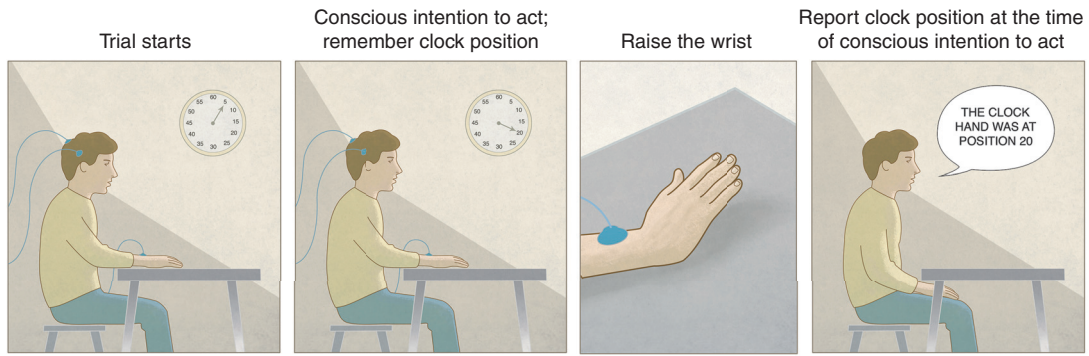


Figure 1. The spontaneous action generation task in Libet's experiment (1983). A human subject is instructed to spontaneously flex the wrist of his/her right hand at any time while looking at a clock-like visual display. After the trial, the subject reports the time he/she became consciously aware of his/her intention to act. (Adapted from Haggard 2008, with permission from Macmillan Publishers Ltd.)

dence-based decision-making and may seem initially to correspond to a different type of decision. Self-initiated acts do have characteristics that might lie outside the normal issues addressed by decision models. A first issue is how to account for the extreme unpredictability in the timing of spontaneous actions. A second, and related, issue is when there are not cues from the environment or well-defined values at stake, what plays the role of "evidence" driving the decision? Then most perplexingly, is there room for a role for the "self" or agent in this model?

The primary aim of this review is to consider what empirical results in neuroscience, by attempting to ad-

dress issues such as these, may have to say about the nature of free will. The central reference will be the recent work of Murakami et al. (2014), who presented new experimental evidence about the neural origins of spontaneous decisions in rats. We have taken as our mission the nontrivial task of providing an account accessible to both scientists and philosophers, without doing injustice to either by oversimplification. Surprisingly, perhaps, we will suggest that even studies in rats might be illuminating to the case of free will in humans. Nevertheless, great care must be taken in making the leap from empirical data to metaphysical conjectures.

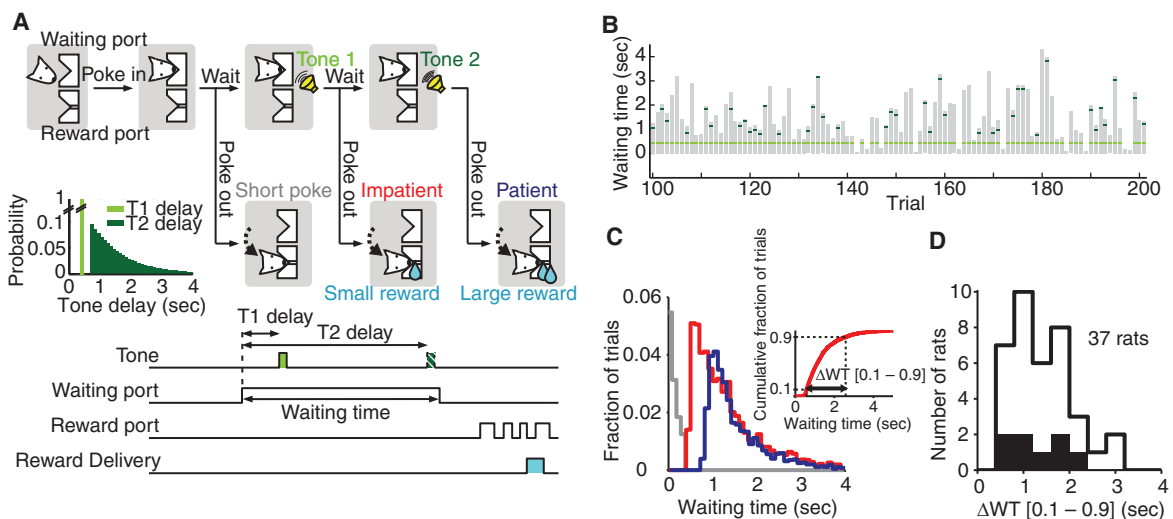


Figure 2. The waiting task and waiting behavior. (A) Schematic diagram of trial events in the rat waiting task (top). In each trial, the rat is required to wait for tone(s) and move to the reward port to obtain water reward. If the rat fails to wait for Tone 1 (T1), the reward is not available. If the rat waits for T1 but leaves the port before Tone 2 (T2), a small reward is available. If the rat waits until T2, a large reward is available. Probability distributions of the delays of T1 (light green) and T2 (dark green) are shown in the inset. Time line of the task events (bottom). The light green rectangle indicates presentation of T1; dark green rectangle, T2; light blue rectangle, reward. Tone 2 is represented by a hatched rectangle to indicate it was not played in the impatient trials. (B) Snapshot of the waiting behavior. The waiting period in each trial is indicated by a gray bar. Light green ticks represent the presentation of T1; dark green ticks, T2. (C) Waiting time histograms of short poke trials (gray), impatient trials (red), and patient trials (blue) for an example rat. The histograms show data pooled across sessions. Inset: Cumulative histogram of waiting times in impatient trials from this rat. The arrow indicates the range from 10th to 90th percentile waiting times ($\Delta WT [0.1-0.9]$), which represents the variability of waiting time for this rat. (D) Distribution of $\Delta WT [0.1-0.9]$, which shows the variability of waiting time across rats. Filled bars indicate rats used in electrophysiology experiments. (Adapted from Murakami et al. 2014.)

SPONTANEOUS ACTIONS FROM RODENTS TO HUMANS

Murakami et al.'s experiment went as follows (see Fig. 2). A thirsty rat initiated waiting by inserting its snout into the waiting port. Shortly thereafter, a first tone was played, after which the rat could garner a small amount of water reward at the reward port. Alternatively, if the rat successfully waited for a second tone, played at a random delay, it would get a larger reward. Of particular interest were those trials in which the rat responded between the first and second tones. In those "impatient" trials rats sometimes responded quickly, but often waited for a second or more before leaving the port. These leaving decisions were said to be "self-initiated" or "spontaneous" because of the high variability in their timing despite otherwise identical conditions.

The authors then recorded neuronal activity in the rat secondary motor cortex. Of the 385 neurons recorded in eight rats, 27 were identified as having "ramp-to-threshold" activity, which means that their firing rate ramped up during the waiting period, reaching a constant threshold just before movement initiation (Fig. 3). Moreover, the timing of threshold crossing was correlated with waiting time in such a way that when the rat waited longer the neurons' firing rate ramped up more slowly. The characteristic activity of these ramp-to-threshold neurons suggested that they might serve as an internal trigger for the behavioral response in the rat. An additional 64 neurons were identified as "transient neurons," because their activity also correlated with the waiting time, but they fired in a brief burst rather than a ramp (Fig. 4A). Remarkably, these transient neurons fired in correlation with waiting time even during a period extending to well over 1 sec before the actual movement. Some of the transient neurons were positively correlated with waiting time, whereas others were negatively correlated (Fig. 4B,C).

To attempt to explain the relationships between the observed electrophysiological recordings and the behavior, a computational model was created (Fig. 5). In this "integration-to-bound" model, a set of transiently active "units" constitute the input to the integrator, and the ramping neurons are taken to reflect the output of the integrator. When the integrator crosses a certain positive threshold,

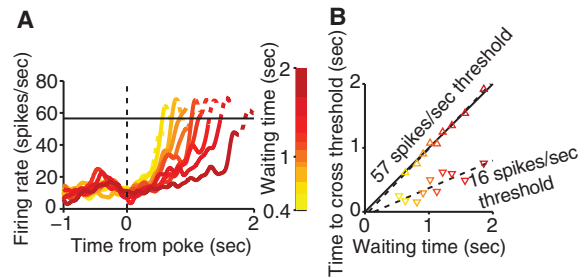


Figure 3. Waiting time predictive activity of ramp-to-threshold neurons. (A) Spike density functions (SDFs) of an M2 neuron in different waiting time trials, aligned to poke-in and smoothed with a Gaussian filter ($SD = 50$ msec). Impatient trials are grouped according to the waiting time, indicated by the color scale. Dashed lines in SDFs indicate times when the rat already left the port in some of the trials in that group. A threshold of 57 spikes/sec is indicated by the horizontal solid line. (B) Time to cross a given threshold firing level as a function of mean waiting time. The analyses with the highest and lowest thresholds with significant correlation (57 spikes/sec, triangle, and 16 spikes/sec, inverted triangle, respectively) are shown. Dashed lines indicate the linear regression line for the time to cross a threshold as a function of waiting time. The color represents the waiting time group as indicated by the color scale in A. (Adapted from Murakami et al. 2014.)

the decision to abort waiting is taken. Each input unit is activated at a certain time point at a level that varies randomly from trial to trial, matching the transient neuron data. Each unit is assumed to connect to the integrator with a different positive or negative weight and therefore promotes shorter or longer waiting accordingly.

This integration-to-bound model could parsimoniously account for the data collected at both the neuronal and the behavioral level. Most importantly, the properties of the ramp-to-threshold neurons strongly resembled the output of the integrator in the model, and the model could explain the correlation between both ramp-to-threshold and transient neurons' activity and waiting time. Moreover, the authors noticed that if transient neurons were assumed to fire independently from each other, the correlation of each one's activity with waiting time systematically decreased to zero as the number of transient neurons increased. Instead, by introducing a common noise source

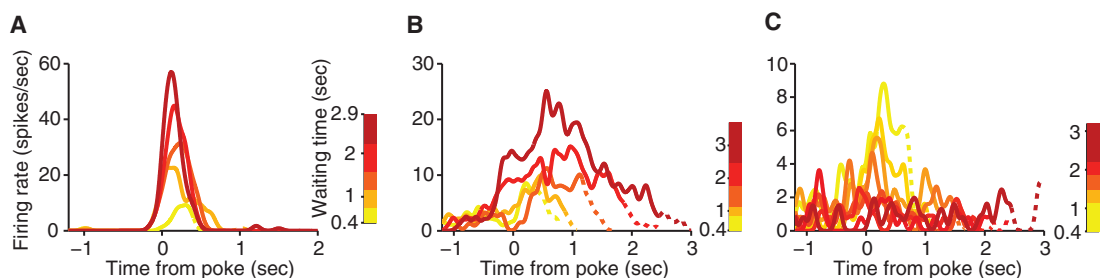


Figure 4. Waiting time predictive activity of M2 transient neurons. (A–C) Spike density functions (SDFs) of three M2 neurons. (A) An M2 neuron that shows phasic activation at the beginning of waiting and whose firing rate is positively correlated with waiting time. (B) An M2 neuron that shows sustained activation during waiting and whose firing rate is positively correlated with waiting time. (C) An M2 neuron that shows negative correlation between the firing rate and the waiting time. The format is the same as is in Figure 3A. (Adapted from Murakami et al. 2014.)

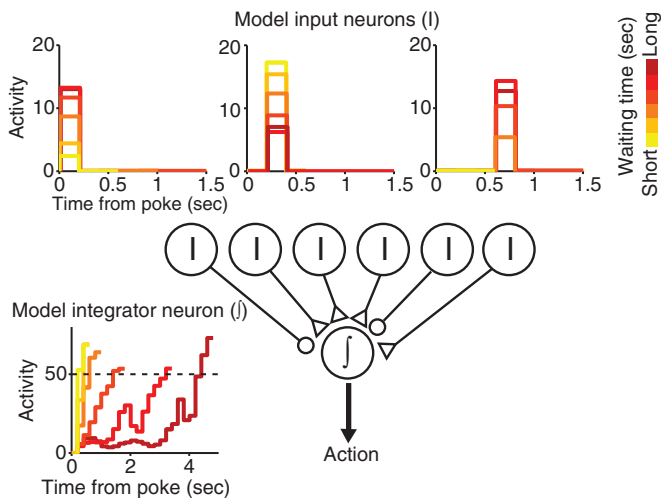


Figure 5. A schematic diagram of an integration-to-bound model for spontaneous actions. In the model, multiple input neurons (I) connect to one temporal integrator neuron (J) with variable synaptic weights (circles indicate inhibitory weights and triangles indicate excitatory weights). *Inset* panels show activity of example model neurons (*top* three panels are example input neurons and the *bottom left* panel is an integrator neuron, the same format as in Fig. 3A). (Adapted from Murakami et al. 2014.)

by which transient units became correlated with each other, the contribution of each single unit became much more significant, even at large population sizes. Interneuronal correlations with the predicted magnitude were in fact observed in M2, an independent measurement that supports the merits of the model.

The integration-to-bound model suggests a critical perspective on the causal relationship between the two kinds of neural data and the behavioral response. The causal connection is mediated by the integrator circuit, which can be thought of as analogous to a ballot system during an election. Transient neurons are considered as input to the integrator. They play the role played by sensory evidence in perceptual decisions, “voting” for or against giving up waiting. Each one has some causal bearing in the timing of the action, but none can be said to determine it. The ramping neurons’ activity, considered as output of the integrator, reflect a cumulative tally of votes. But if read out at a level less than the decision threshold, the tally is only suggestive and not determinant of an outcome. Even a tally reaching very close to threshold might happen to meander back away from it. The appearance of ramping is a result of averaging of trajectories that hit the threshold.

The dynamics of individual trials are tortuous paths reflecting the random bombardment of votes both for and against. The origin of trial-by-trial variability in behavior is due to the variability in transient neurons’ firing together with their strong interneuronal correlations.

With these data, Murakami et al. show the integration-to-bound model to be a useful explanatory tool in relating the timing of spontaneous decisions to the neuronal level, providing a hypothesis about what is taking place in the rats’ brains. We suggest that they can potentially inform us about choices in Libet-type experiments as well. There are different levels of comparison in which we can see the parallels between the Murakami and Libet experiments. First of all, the similarity of behavior: Spontaneous wrist movements in humans and spontaneous waiting aborts in rats are *prima facie* similar (although not identical) kinds of “self-initiated” behaviors, and both occur with highly unpredictable timing. The second parallel is the similarity of shape and time course between ramping neurons’ activity recorded in rat secondary motor cortex and readiness potentials recorded over the homologous human supplemental motor area (Fig. 6). Fried’s (2011) study, showing something like ramping neurons in humans per-

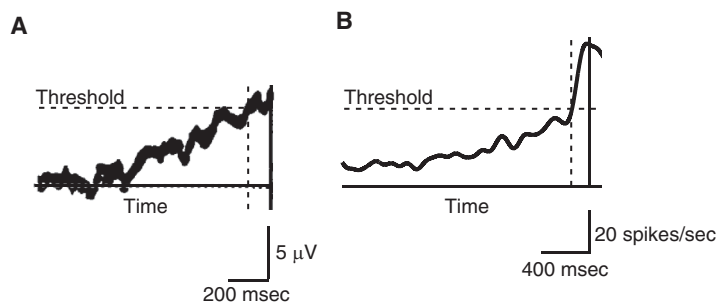


Figure 6. Ramping activity in human and rat motor cortex preceding spontaneous action. (A) An example EEG trace recorded from the vertex cortex corresponding to the supplementary motor area. The EEG trace is an average of 40 trials, aligned to the movement onset (solid vertical line). In this session, the subject felt a conscious intention to act, on average, 118 msec before the movement. A dashed horizontal line indicates a hypothetical threshold line for the conscious intention to act. (Adapted from Libet et al. 1983.) (B) Example single neuron activity recorded in rat M2 during spontaneous giving up trials. The trace is an average of 10 trials with the longest waiting times (1.7–2.0 sec), aligned to the give-up time (solid vertical line). A hypothetical threshold line (dashed horizontal line) and the time the activity crosses the threshold (dashed vertical line) are indicated. (Adapted from Murakami et al. 2014.)

forming the Libet task, is also strongly suggestive of this parallel. The third parallel is the applicability of the integration-to-bound model to describing the data. The case for the rodent data has been made above. In the case of Libet-type experiments, Schurger et al. (2012) have recently made this case strongly at the behavioral level.

A final very intriguing parallel between the two sets of experiments is the timing of threshold crossing in the rat experiments and the moment of subjective intention to act reported by Libet's subjects (Fig. 6). Both occur at ~ 200 msec before the movement. Accepting a similar set of mechanisms between the two experiments, it could be hypothesized that threshold-crossing corresponds to the moment of decision both at the neural and the mental level. Clearly, however, concrete evidence that threshold-crossing of neurons with ramping activity corresponds to the moment of awareness of decision, in one species or the other, would be needed to justify this idea.

PHILOSOPHY AND NEUROSCIENCE OF FREE WILL

The relevance of these results for the understanding of free will, defined as the agent's ability to control her decisions and actions, may not be entirely straightforward but we will argue for their importance nonetheless. Similarly to the now common distinction between the easy and the hard problems of consciousness (Chalmers 1995), we believe it can be useful to distinguish "easy" and "hard" problems of free will (Shariff et al. 2008). The "hard" questions of free will are the ones that have typically concerned philosophers for 25 centuries and seem to remain squarely out of reach by scientific experiments. Can a person act freely in a deterministic world? (A world is deterministic if, given a full description of all its elements and laws at t_1 , only one possible state can follow at t_2 . This means that, given a certain cause, only one effect can follow. A world is said to be indeterministic if this is sometimes not the case.) Can anyone make sense of an indeterministically caused action that is not random? In other words, is it coherent to think that a person might make a decision based on reasons and still have the possibility to act otherwise given the exact same circumstances and laws of nature?

In contrast, the "easy" problems are in principle scientific in nature, and today's neuroscience can contribute at least to illuminating the empirical aspects that are related to these issues. Is the variability that we perceive in the behavior of biological systems originated within the organisms themselves or is it just the deterministic output of undetectable changes in the environment? Do we have evidence that brain processes at the neuronal level are sensitive to genuine indeterminacy originated at a more fundamental level (the quantum domain, say)? What do we know about how the brain weighs alternatives and decides what to do when facing different possibilities of action? What does Libet's readiness potential represent in the psychological process of decision-making? What is the neural basis of the self who endorses the decision?

The relationship between the mental level, where deliberation is described in terms of beliefs and desires, and the neuronal level, where scientists measure spikes per second, has been and remains an object of great debate in philosophy. However, no one denies that what can be provisionally proven or disproven about the brain must be somehow taken into account when we speak about the mind. It is generally accepted that the mind "super-venes" on the brain: That is, that there cannot be changes at the mental level without there being corresponding changes at the neural level. This is not considered a strong thesis, as it regards only the interdependence of mind and brain, not their nature nor how one produces the other, but it is enough to justify the importance of understanding how the physiological substrate works in order to know the limits of what philosophers are entitled to conjecture. It would be pointless to argue for the philosophical plausibility of a certain account of free will if its empirical commitments were totally implausible from a scientific point of view.

However, one must be aware of several aspects of the philosophical concept of free will before purporting to infer any conclusions from what is possible to observe in experiments such as this one.

First of all, the philosophical literature about free will has intricately associated it with moral responsibility (Strawson 1962; Kane 1996; Pereboom 2001, 2014). To consider that to be free is to be fit for responsibility attributions is the standard way of framing the problem of free will, and one that only occasionally has been called into question (Steward 2012). In contrast, experiments such as Libet's or Murakami's have obviously nothing to do with morality. Yet, we believe morally neutral decisions are just as relevant for the debate as any other, for what matters for free will is the agent's control over her choice or her lack thereof.

Second of all, the idea that free will is an exclusively human ability is often taken for granted from the start (Clarke 2003), and that of course would prevent us from considering that animal models can provide us with anything other than evidence for the evolutionary antecedents of this privilege of ours. We do not favor this view, however, for there is no obvious reason why this should be taken for granted from a scientific standpoint. Even if there are relevant differences, we are likely to learn something about our situation from the evolutionarily common aspects we share with other animals.

Moreover, when trying to use these studies as arguments in the philosophical debate about free will, one must be very careful with terminology. What philosophers call "reasons for action" neuroscientists call "evidence"; what philosophers call "intentional," neuroscientists call "goal-directed"; and, most importantly, while philosophers (in the context of this debate) take "decision" to mean a conscious and active formation of an intention to act, usually following a deliberative process, neuroscientists apply the term both to the commitment made after some sort of reasoning in the case of evidence-based decisions *and* to more general mechanisms of settling an indeterminate matter, such as where to turn the head. It is also important to note that, in simple "decisions," the

indeterminacy that precedes choice may be only epistemic (i.e., it may depend only on the contingent limitations of an external observer's knowledge of the stimulation). In fact, the problem of determinism, from an empirical point of view, is mainly an epistemic issue, whereas the sort of indeterminism that philosophers consider important is also ontological (i.e., it regards the underlying nature of reality). Also for that reason, philosophers and neuroscientists tend to be talking about different things when discussing these matters.

PHILOSOPHICAL IMPLICATIONS OF MURAKAMI'S RESULTS

Reinterpreting Libet's Readiness Potential

The arguments based on Libet-type cases typically use as a premise the fact that an unconscious ramping activity at the neuronal level is correlated with the decision made consciously by the agent in such a way that we are able to predict that decision to some degree. Since Libet et al. interpret the readiness potential as part of the decision, they suppose this observation to *mean* that the decision was unconsciously made by the agent's brain, which becomes aware of it only some time later. The gap between the start of the brain's predictive activity and the subjective awareness of the urge to move is said to be as long as 10 sec in some of the experiments (Soon et al. 2008). This is considered to undermine free will because it is hard to see how the agent might control what she is not aware of.

In the above experiment, Murakami et al. defend a model according to which the ramping activity of neurons reflects the accumulation of input in favor or against a certain spontaneous action, to which the agent is decisively committed only when a certain threshold is crossed. Until that moment, the action might be predicted with a certain degree of accuracy, like election polls, but the outcome is open until the very last moment when the decision is made—that is, when threshold is crossed, and overt action takes place immediately after. The authors hypothesized that the ramping activity of the integrator, which in the case of the waiting task in rats is reflected in the ramping neurons in the secondary motor cortex, is, in Libet's experiment, reflected in the slowly building readiness potential.

This model and its use in Murakami's experiment fleshes out what Alfred Mele (2014b) has argued when discussing putative neuroscientific evidence that free will is just an illusion: "when the rise starts about half a second before the muscle burst in the main experiment, the beginning of the EEG reading—or the first half of it—is correlated with something that precedes an intention rather than with an intention itself" (Mele 2014b, p.19).

Free Will in Rodents?

The relevance of experiments with rodents for the debate around human free will might be called into question by philosophers who would maintain that free will is a specifically human feature. But so far evidence has confirmed the success of integration-to-bound models in ex-

plaining simple evidence-based decisions not only in nonhuman primates and in humans, but also in rats and even in invertebrates (e.g., DasGupta et al. 2014), consistent with a wealth of biological data arguing for evolutionary continuity in biological systems. With the present data, Murakami et al. added a further argument for extending the case from perceptual or value-based judgments to self-initiated actions. So, from rats to monkeys to humans, the same type of process seems to take place when simple decisions are made, both when they are based on evidence provided by the environment and when they are spontaneously made.

From Waiting Tasks to Rational Decisions

Another question has to do with the relevance of this type of task, given that, in the philosophical literature, free will is taken to be the agent's ability to control a choice that is made for reasons. What philosophers mean by "reasons" in this context are beliefs and desires, and it may not be entirely clear how such reasons translate into factors in spontaneous decision tasks. In humans, these tasks typically involve apparently arbitrary urges. Agents are usually asked to flex their wrists or press a button whenever they feel like it, and they are explicitly instructed not to plan their action in advance. There is no extrinsic reason why it should be preferable to flex the wrist *now* rather than *later*, apart from maybe boredom or the intention to keep a random pattern. It may thus be objected that Libet-type experiments do not probe the right kinds of decisions because they are without reasons.

With respect to this issue, Murakami's task and interpretation would seem to help the case for relevance. Murakami's rats had conflicting motivations related to the content of the decisions themselves: the desire to drink plus the certainty of the small reward versus the possibility of a larger reward, despite the inconvenience of having the snout poked in while waiting for the delayed tone. One could thus argue that their experience is more closely related to the human decisions that philosophers discuss as free than to the tasks that have been tested in Libet-type experiments. And the fact that the rats hesitated before leaving the waiting port in impatient trials, while moving promptly in patient trials, was evidence for how much they had learned the task and understood what could be expected. In other words, one can argue that subjects in Murakami's waiting task did have beliefs—that water would be delivered in the reward port, that a larger amount would be given if they waited for the second tone—as well as desires—above all, the desire to drink. Therefore, their decisions were not random but made for reasons. On the other hand, although these reasons are clearly relevant to the rats' behavior in general, they do not necessarily dictate the *precise timing* of giving up in a particular trial, which varies substantially and randomly even given apparently similar level of thirst and so on. Thus, the precise moment of giving up in a particular time would seem to be a less-reasoned decision, much as is argued for the Libet experiments.

Dual Theories of Decision-Making

The main trend in neurobiology is to assume that “the path from simple decisions to complex ones may be more straightforward than it appears” (Gold and Shadlen 2007, p. 562). However, further evidence that the integration-to-bound decision model applies to truly “deliberative” decision-making may be needed in order to more reliably move on to the sort of cases philosophers enjoy discussing. Are rapid perceptual decisions truly representative of the process by which, for instance, a human agent decides to change jobs and move with her family to a different city? Is the rational, abstract, and linguistically mediated consideration of reasons in this case just a more complex version of what animals such as Murakami’s rats experience? It has indeed been argued that complex human choices, like deciding to marry someone, are at a different qualitative level with respect to more simple ones. In particular, Kahneman’s defense (2011) of a two-systems approach to judgment and choice in which a fast, automatic, and often unconscious System 1 is distinguished from the slow, effortful, and controlled System 2, has a long and influential history in psychology (Tversky and Kahneman 1974; Kahneman et al. 1982; Gigerenzer et al. 1999). In the context of moral psychology, a dual-process theory has also been proposed (Greene et al. 2004), suggesting that competing subsystems (emotional versus cognitive) in the brain are responsible for moral judgments under different situations.

It is tempting to think that decisions based on deliberation are more likely to be free than rapid choices based on “gut,” in the sense that the longer process of weighing different reasons, such as very abstract ones like moral questions, seems to allow for more control by the agent’s conscious self than an automatic mechanism. However, processes underlying System 1, despite being, by definition, not conscious or effortful, are associated with experiences of agency and choice, just like the ones underlying System 2. Also, there is not yet any evidence for such a dichotomy at the neurophysiological level (Sugrue et al. 2005), which would be fundamental given our assumption of the supervenience of the mind on the brain. Moreover, we do not have reasons to believe a putative deliberative system would be less deterministic than its counterpart based on fast heuristics.

All or Nothing at All

The questions raised by the differences between Murakami’s rats and human agents undergoing complex deliberative processes remind us how the use of neuroscientific results in philosophical discussions about free will needs to be conducted with caution. However, one thing that becomes clear when we review the results from this experiment and others is that the same integration-to-bound model that has successfully been applied to simple decisions in humans and monkeys allows for a parsimonious explanation in the case of rats and, most probably, in the case of Libet-type subjects. So the similarity between the neural patterns in all these cases makes it likely

that either there is a continuity between all of them and the more complex cases considered by philosophers when discussing free will, or none of them is relevant for the discussion—Libet cases included.

CONCLUSIONS

The integration-to-bound model and the evidence provided by Murakami et al.’s results of its general applicability to situations of self-initiated decisions in biological systems is a good challenge to the common idea that neuroscience is inimical to free will. In cases of human actions, it provides us a hypothetical explanation for the timing of the phenomenological experience of conscious decision, identified as the moment of threshold crossing within the integrator mechanism. At the same time, and importantly, it accounts for the existence of causal antecedents that make one decision or another more likely. The outcome is neither inevitable beforehand nor random.

Some free will defenders argue that Libet-type experiments are not representative of the sort of situations that are typically under discussion in the philosophical context. We believe this argument is open for discussion, but if Murakami and his co-workers are right, it is actually irrelevant. Libet’s threat to free will has been overrated simply because it wrongly assumed that because a choice is preceded by a signal it follows that a signal is always followed by a choice. Instead, even though the choice can be forecast, it remains open until decision is made. Together with growing consensus in favor of an inner source of variability in biological systems (Brembs 2011), these results leave the door open even for the most demanding indeterministic accounts of free will.

REFERENCES

- Brembs B. 2011. Towards a scientific concept of free will as a biological trait: Spontaneous actions and decision-making in invertebrates. *Proc R Soc B* **278**: 930–939.
- Chalmers D. 1995. Facing up to the problem of consciousness. *J Conscious Studies* **2**: 200–219.
- Clarke R. 2003. *Libertarian accounts of free will*. Oxford University Press, New York.
- DasGupta S, Ferreira CH, Miesenböck G. 2014. FoxP influences the speed and accuracy of a perceptual decision in *Drosophila*. *Science* **344**: 901–904.
- Dennet D. 1991. *Consciousness explained*. Little, Brown, Boston.
- Fried I, Mukamel R, Kreiman G. 2011. Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron* **69**: 548–562.
- Gigerenzer G, Todd PM, the ABC Research Group. 1999. *Simple heuristics that make us smart*. Oxford University Press, New York.
- Gold JI, Shadlen MN. 2007. The neural basis of decision making. *Annu Rev Neurosci* **30**: 535–574.
- Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* **44**: 389–400.
- Haggard P. 2008. Human volition: Towards a neuroscience of will. *Nat Rev Neurosci* **9**: 934–946.

- Haggard P, Eimer M. 1999. On the relation between brain potentials and the awareness of voluntary movements. *Exp Brain Res* **126**: 128–133.
- Hanes DP, Schall JD. 1996. Neural control of voluntary movement initiation. *Science* **274**: 427–430.
- Kahneman D. 2011. *Thinking fast and slow*. Farrar, Straus & Giroux, New York.
- Kahneman D, Slovic P, Tversky A, ed. 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press, New York.
- Kane R. 1996. *The significance of free will*. Oxford University Press, New York.
- Kornhuber HH, Deecke L. 1965. Hirnpotentialänderungen bei willkürbewegungen und passiven bewegungen des menschen: Bereitschaftspotential und reafferente potenziale. *Pflugers Arch Gesamte Physiol Menschen Tiere* **284**: 1–17.
- Krajbich I, Armel C, Rangel A. 2010. Visual fixations and the computation and comparison of value in simple choice. *Nat Neuroscience* **13**: 1292–1298.
- Lau HC, Rogers RD, Haggard P, Passingham RE. 2004. Attention to intention. *Science* **303**: 1208–1210.
- Levy N. 2014. *Consciousness and moral responsibility*. Oxford University Press, Oxford.
- Libet B. 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav Brain Sci* **8**: 529–566.
- Libet B, Gleason CA, Wright EW, Pearl DK. 1983. Time of unconscious intention to act in relation to onset of cerebral activity (Readiness-Potential). *Brain* **106**: 623–642.
- Mazurek ME, Roitman JD, Ditterich J, Shadlen MN. 2003. A role for neural integrators in perceptual decision making. *Cereb Cortex* **13**: 1257–1269.
- Mele AR. 2009. *Effective intentions: The power of conscious will*. Oxford University Press, New York.
- Mele AR. 2014a. *A dialogue on free will and science*. Oxford University Press, New York.
- Mele AR. 2014b. *Free. Why science hasn't disproved free will*. Oxford University Press, New York.
- Murakami M, Vicente MI, Costa GM, Mainen ZF. 2014. Neural antecedents of self-initiated actions in secondary motor cortex. *Nat Neurosci* **17**: 1574–1582.
- Pereboom D. 2001. *Living without free will*. Cambridge University Press, New York.
- Pereboom D. 2014. *Free will, agency and the meaning of life*. Oxford University Press, Oxford.
- Roitman JD, Shadlen MN. 2002. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* **22**: 9475–9489.
- Schurger A, Sitt JD, Dehaene S. 2012. An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proc Natl Acad Sci* **109**: e2904–e2913.
- Shariff AF, Schooler J, Vohs KD. 2008. The hazards of claiming to have solved the hard problem of free will. In *Are we free? Psychology and free will* (ed. Baer J, Kaufman JC, Baumeister RF), pp. 181–204. Oxford University Press, New York.
- Soon CS, Brass M, Heinze HJ, Haynes JD. 2008. Unconscious determinants of free decision in the human brain. *Nat Neurosci* **11**: 543–545.
- Steward H. 2012. *A metaphysics for freedom*. Oxford University Press, Oxford.
- Strawson PF. 1962. Freedom and resentment. *Proc Br Acad* **48**: 187–211.
- Sugrue LP, Corrado GS, Newsome WT. 2005. Choosing the greater of two goods: Neural currencies for valuation and decision making. *Nat Rev Neurosci* **6**: 363–375.
- Tversky A, Kahneman D. 1974. Judgment under uncertainty: Heuristics and biases. *Science* **185**: 1124–1131.
- van de Grind W. 2002. Physical, neural, and mental timing. *Conscious Cogn* **11**: 241–264.
- Wang XJ. 2002. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* **36**: 955–968.
- Wegner D. 2002. *The illusion of conscious will*. MIT Press, Cambridge, MA.