RIVISTA INTERNAZIONALE DI FILOSOFIA E PSICOLOGIA
DOI: 10.4453/rifp.2015.0011
ISSN 2039-4667; E-ISSN 2239-2629
Vol. 6 (2015), n. 1, pp. 145-160

Ricerche

What do Neurosciences Talk About When They Talk About Free Will?

Federica Della Grotta

Ricevuto: 21 novembre 2014; accettato: 16 marzo 2015

Abstract In this paper, I will take into account and criticize two of the most celebrated neuroscientific experiments about free will, which seem to deny that agents freely deliberate about simple choices of their everyday life: the pioneering experiment of Benjamin Libet and the more recent one of John Dylan Hayes. My aim is to reject the relevance of their empirical results, which deny the existence of free will. However, such a rejection will not rely on criticisms about how the experiments are conducted. Instead, I would like to bring about a broad philosophical and methodological concern: namely, that the success or the failure of the experiments in arguing for the illusion of free will is strictly dependent on the *meaning* of the notion of free will which is put through an experimental investigation.
KEYWORDS: Free Will; Neuroscience; Experimental Investigation; Benjamin Libet; Dylan Hayes.

Riassunto *Di cosa parlano le neuroscienze quando parlano di libero arbitrio?* – In questo articolo mi occuperò criticamente di due tra i più noti esperimenti neuroscientifici sul libero arbitrio, i quali paiono negare che gli agenti possano liberamente prendere decisioni, anche molto semplici, nel corso della vita quotidiana: il pionieristico esperiemento di Benjamin Libet e quello più recente di John Dylan Hayes. Intendo mettere in discussione la rilevanza dei loro risultati empirici che negano l'esistenza del libero arbitrio. Questo rifiuto non sarà basato su una critica del modo in cui questi esperimenti sono stati condotti. Vorrei piuttosto portare l'attenzione su una questione più ampia da un punto di vista filosofico e metodologico: il successo o il fallimento degli esperimenti nell'affermare l'illusione del libero arbitrio è strettamente dipendente dal *significato* della nozione di libero arbitrio che messa sotto indagine da un punto di vista sperimentale.
PAROLE CHIAVE: Libero arbitrio; Neuroscienza; Indagine sperimentale; Benjamin Libet; Dylan Hayes.

IMAGINE YOU ARE ATTENDING A talk. When the debate starts you feel the desire to join the discussion; accordingly, you raise a finger, and wait for your turn. Intuitively, the act of raising a finger is a very simple one: you can easily perform it, and you certainly perceive it as a result of a conscious deliberation of yours to say something in that debate. However, if one looks at the brain's activities underlying the act of raising a finger, they seem to suggest something different: that our intuitive idea that such an action is "free"- namely, that one's raising a finger is the result of one's own free will and conscious activity – is mistaken.

The pioneering experiments performed by

F. Della Grotta - Faculty of Philosophy, University of Cambridge, 401 King's College, CB2 1ST, Cambridge (UK) E-mail: fd280@cam.acuk (✉)

the neuroscientist Benjamin Libet in the 80s[1] aimed to investigate the ability to exercise free will and they precisely took into account the simple act of raising a finger. Following Libet's experiments, different neuroscientific studies have tried to shed light on the faculty of free agency by means of an experimental approach.[2]

Neuroscientific studies ask an important question: whether it is true that rational agents exercise some form of conscious control over their decisions or actions. The picture offered in response to this question is, most of the time, a negative one: agents do not possess a conscious control over their decisions and actions, even the simple one of raising a finger, because a person's brain seems to initiate decisions before she becomes aware of having made them. Moreover, neuroscientists go further and deduce from the empirical findings that a plausible idea of free will – namely that we consciously cause and perform our own decisions and actions – is inconsistent with experiential data: it is, it is said, an *illusion.*[3]

In recent years, such neuroscientific results have raised a big clamour: a copious number of non-specialized magazines have indeed reported the data and given the alarming news that we are just machines without freedom.[4] Nonetheless, according to the vast majority of philosophers, the results the experiments bring about are strongly controversial, and surely not conclusive in proving the illusory nature of free will.[5] Such a negative assessment is mostly due to several empirical and conceptual inaccuracies detected in the experiments by philosophical analyses.

In this paper, I will take into account and criticize two of the most celebrated neuroscientific experiments about free will, which seem to deny that agents freely deliberate about simple choices of their everyday life: the pioneering experiment of Benjamin Libet and the more recent one of John Dylan Hayes.[6]

My aim is to reject the relevance of their empirical results. However, such a rejection will not rely on criticisms about how the experiments are conducted. Instead, I would like to bring about a broad philosophical and methodological concern: namely, that the success or the failure of the experiments in arguing for the illusion of free will is strictly dependent on the *meaning* of the notion of free will which is put through an experimental investigation.

In other words, "free will" can mean various things, and even neuroscientists, implicitly or explicitly, tend to rely on particular philosophical theses about free will *before* setting the experiments. The validity of the experimental findings, then, is deeply dependent on the particular conception of free will which is previously embraced. I aim to shed light on the meanings of free will which Libet and Haynes seem to take for granted, and to contextualize them in the philosophical debate about free will. As a result, I will show that such notions of free will are firstly controversial from a philosophical point of view. Accordingly, I will show that in both cases it is possible to reject the experiments' results by simply opposing to them two alternative understandings of free will, which are not challenged at all by the empirical findings.

Libet's experiment

Benjamin Libet's pioneering experiments in the 80s[7] intended to show that a serious trouble for the existence of free will is hidden in the notion of *conscious will*. Libet's experiments concern not only the question as to whether something (such as a pre-determined chain of events) causes our will, but mainly whether there is something like a conscious will which causes anything; that is, if there is something like a conscious will which is efficacious in causing our willed actions.

First, I will briefly describe the empirical setting of the experiment. Then, I will clarify the hidden theoretical framework that Libet seems to take for granted in his experiment. As a result, I will show that Libet's interpretation of the empirical data as evidence that free will is an illusion relies on a specific philosophical characterization of free will: *incompatibilism*. Accordingly, I will object this

position and I will show that it is possible to defend a notion of free will which is not threatened by Libet's empirical discoveries. My aim, then, is to argue that it is not certain that Libet's empirical findings are able to show that free will is an illusion.

The experimental setting

Libet's experiments were designed to explore the role of conscious intentions in the production of voluntary actions. Indeed, he has experimentally investigated the temporal relation between the appearance in the experimental subjects of the conscious urge to perform an action, and the beginning in the brain of the neurophysiological processes that lay behind that urge.

In the most famous of his experiments,[8] he asked experimental subjects to periodically raise a finger, freely, while at the same time they had to look at a clock with a rotating spot. In order to time the appearance of the conscious will, they were supposed to remember where the dot on the clock-face was when they had the urge to move their finger. Moreover, Libet used electro-encephalogram machines (EEG) to record, during the task, electrical signals from the scalp of experimental subjects.

What Libet found was a 200 millisecond delay, on average, between the appearance of the conscious urge to move a finger (called W and registered with an electromyogram which shows relevant muscular motion to begin) and the movement itself. But the EEG recordings also revealed the presence of a scalp potential, called the readiness potential (RP), that appears in the brain even earlier – 550 milliseconds, on average – before the action.

Table 1. Summary of Libet's results

-550 ms	-200 ms	0 ms
RP	W	Action starts

Libet interpreted the data and concluded that the cerebral activity, identified with RP, represents the effective cause of the intentional process that leads to the execution of a free voluntary movement. In this picture, conscious will (W) seems to be only *epiphenomenal*: it has no role in causing actions, because it appears in the subject when the brain processes are already started: it is, then, constantly "after the facts".[9]

Free will or metaphysical freedom?

Libet's thesis is that his empirical findings clearly dismiss the possibility for human agents of possessing free will, that is, the ability of consciously cause their own actions. In fact, according to Libet's interpretation of the data, these show that while the brain is the effective originator of our actions, consciousness – which is straightforwardly equated to free will – has no role in causing our actions.

One common strategy in assessing Libet's results is to ask *whether* Libet's empirical data are a genuine problem for the existence of free will. My strategy in this paper, however, is slightly different: I aim to ask *why* Libet's results are perceived, by Libet himself to begin with, as a real threat to the existence of free will. The answer to this question is that the reason is in its essence a conceptual one: Libet's results are perceived as undermining the existence of free will because of one's taking for granted a particular understanding of such a concept.

It is possible to detect two main aspects in Libet's interpretation of the data which seem to make them worrisome: a *temporal aspect* and an *awareness aspect*.

(a) *Temporal aspect*. If the brain's activities start before the appearance of the conscious will, then conscious will, which is temporally delayed in respect to them, cannot be considered as the origin of our willed actions.

(b) *Awareness aspect*. Everything the agent is

> unaware of is a problem for free will: an action, whose causes are unconscious, cannot be considered a free action.

Libet's experiment is, then, immediately perceived as a threat to free will because it seems to undermine a powerful idea: that the agent has to be the *source* of his own decisions and actions. More explicitly, the two aspects joined together suggest that the agent has to be the conscious originator of the causal chain of events which lead to an action, and that this feature guarantees that he is the free source of his actions.

Furthermore, according to Libet, there is another aspect in the picture suggested by his data which strongly undermines the existence of free will: the hypothetical *deterministic character* of the brain's activity. In Libet's words:

> We have not answered the question of whether our consciously willed acts are fully determined by natural laws that govern the activity of the nerve cells in the brain, or whether acts and the conscious decisions to perform them can proceed to some degree independently of natural determinism. The first of these options would make free will illusory [...] We would not need to view ourselves as machines that act in a manner completely controlled by the known physical laws.[10]

In this passage, Libet admittedly perceives the hypothetical existence of deterministic trends in the brain's activities as an insurmountable problem for free will, though presenting his results as not conclusive in this direction.

The overall connotation of free will which emerges from these remarks is, accordingly, a very specific one: that the ability of exercising free will requires by definition a non-physical element, which can sidestep the risk of deterministic behaviour and which can be identified with the agent himself as a causal origin of his action. In this perspective, physical operations of the brain and, even more, their possibly deterministic trends are perceived by Libet as deeply incompatible with an acceptable conception of free will, while the conscious will seems to be the correct place where to locate free will.

The validity of such claims is, however, an old topic of discussion in the philosophical debate about free will. Specifically, Libet's position is easily attributable to a family of philosophical positions in the free will debate: *incompatibilism*. Indeed, according to such a view, free will is intrinsically incompatible with determinism and is characterized, in some particular versions,[11] by the following theses:

(1) *Source incompatibilism.* The agent, and not his physical substratum, has to be the origin of the causal chain which leads to the action.

(2) *Leeway incompatibilism.* When making a choice, the agent has to be the effective source of his own decisions, by possessing the ability to choose between different courses of action.

I would like to call the conception of free will that results from (1) and (2) Metaphysical Freedom (MF), and I suggest that both elements of this view are implicitly taken for granted by Libet. More precisely, I will argue that they are the reason why he perceives the temporal and the awareness aspects as a genuine threat to free will. However, I will suggest that it is possible to characterize free will in a way that does not presuppose the existence of MF, thereby avoiding the challenge of Libet's empirical results.

Metaphysical freedom: The temporal aspect

The *temporal aspect* of Libet's worry suggests that if there are brain's activities which start before the appearance of the conscious will, then conscious will, which is equated to

free will, cannot be considered the origin of our willed actions. Consequently, if the conscious free will is not the cause of our decisions and actions, *we* are not the cause of them.

The temporal aspect can be further elucidated in two related concerns: the *Epiphenomenalist concern* and the *Source concern*. I argue that Libet explicitly recognizes the former, and that he is brought from it to get close to share the latter.

The Epiphenomenalist concern asks whether, in the explanation of voluntary actions, consciousness plays a role in the initiation of bodily movement. It distinctively answers that the causal efficacy has to be entirely attributed to neural mechanisms, but not to consciousness. That is, while neural events cause bodily movement and consciousness, consciousness cannot cause neural events. In this picture, consciousness is only an *epiphenomenon*: we think that it is the cause of our voluntary activity while the truth is that it is not. Accordingly, the epiphenomenalist interpretation of Libet's results states that what we call free will is nothing more than a mistaken impression, an *illusion*, because our voluntary actions are completely caused by physical brain activities.[12]

It is a common view in philosophy that actions can be free only if there is a conscious, mental activity which directly causes them[13]. However, it is possible to recognize behind such a claim at least two flaws (both displayed even by epiphenomenalism). The first flaw lies in the certainty that only a causal-effective conscious dimension can guarantee that *the agent* is the cause of his actions. The second flaw lies in the straightforward equation of the mind's conscious activity to free will. It seems indeed that Libet understands the phenomenon of free will exactly in these flawed terms: if conscious mental events do not operate as a cause that moves or directs the body, then we do not possess free will.

For the purposes of rejecting Libet's position, I will show why the first claim is flawed.[14] Following Daniel Dennett, Libet's position can be called Cartesian materialism.[15] According to Cartesian materialism, somewhere in the brain there is a place where a hypothetical observer could always "find" the content of conscious experience. Dennett's arguments show that Cartesian materialism is an out-dated position in the debate about consciousness.

Indeed, such a position is still centred in Descartes' idea of the existence in the brain of some centralized and separated "storage", where the contents of consciousness are combined and assembled, a place Dennett calls the "Cartesian theatre". In these respects, both epiphenomenalism and Libet seem to suggest an obsolete picture of conscious free will, in which consciousness is still a sort of *homunculus*, which contemplates as a spectator what happens in the theatre of the mind.

Such an interpretation of his data seems to get Libet close to the Source concern (1) of the conception of free will that I have called Metaphysical Freedom (MF). In short, MF says that the agent, and not his physical substratum, has to be the origin of the causal chain which leads to the action. Such a position holds that an action is free if it is caused and controlled by the agent in a distinctive *self-determining* way, a way that is incompatible with deterministic causation. According to this view, even if we are apparently able to conduct our everyday actions, to be truly free agents we must be the "ultimate sources" of our decisions. In other words, we can make free choices only if we cause our choices and nothing causes us to cause them.

The soundness of such a position in the free will debate is however strongly objected. More generally, it is at least a matter of discussion if a position which perceives as deeply problematic the existence of physical, mechanistic causation in our bodies is at all acceptable from the point of view of our current scientific theories. The affinity between Libet's position and *Source Incompatibilism* shows that what is missing in Libet's theory is an acceptance of an even *basic* physicalist position: that is, that mental events are correlated in some way with physical events in the brain.[16]

But if this is true, it is possible to reject as a false problem the fact that our actions are caused by a chain of (physical) events. In such a picture, it is not very surprising that the brain is doing something while the agent is making a decision or an action. From a basic physicalist point of view, the temporal factor in Libet's experiments is no longer a deep source of worry: why should we consider the timing of consciousness as a singular instant? It is more realistic, also at a phenomenological level, to understand consciousness as a *process* which is linked with some physical activities of the brain: in this light, there *should* be some extended brain event that underlies the process of conscious awareness.[17]

My suggestion, then, is that Libet's implicit endorsing of an incompatibilist position about free will is the actual reason of his deep worry about the temporal aspect of his findings. Only under such theoretical assumptions he is brought to perceive as "mysterious" the small temporal gap between brain processes and the appearance of the conscious awareness. If one rejects Libet's incompatibilist position, the "mysterious thing" would become instead the absence of some process in the brain that corresponds, and even precedes, our decisions.

Metaphysical freedom: The awareness aspect

The *awareness aspect* of Libet's worry suggests that if the causes of an action are not immediately transparent to the conscience of the subject, namely if they are "unconscious", the ability to exercise free will is in serious troubles. Indeed, if the causes of an action are unconscious (as the physical ones seem to be), the agent is not really *participating* in choosing to perform that action, that is instead brought about by the activity of the brain.[18]

Such interpretation of his findings seems to get Libet close to thesis (2) of MF. In this sense, Libet seems to accept the controversial philosophical claim that, when making a choice, the agent has to be the effective source of his own decisions by possessing the ability to causally choose between different courses of action. Indeed, if the causal determinants of an action are unconscious, as Libet's data seem to suggest, then one intuitive and unpleasant consequence would be the impossibility for the agent to have any conscious role in the choice of starting one course of action instead of another.[19]

In philosophical terms, this particular ability for the agent is usually called 'metaphysical ability to do otherwise' and it is a central feature of the MF. It states that the existence of real, metaphysical alternative possibilities (or the agent's power to do otherwise) is a necessary condition for acting freely, and that determinism is not compatible with alternative possibilities, because it precludes the power to do otherwise for the agent.[20]

However, the claim that the agent has to necessarily possess metaphysical alternative possibilities in order to be a free agent is a much contested statement in the free will debate. According to the proponents of compatibilist theories, it is indeed possible to exercise a robust ability to act freely without possessing the ability to do otherwise and, thus, even if determinism is true. A significant argument in this direction is represented by the series of counterexamples developed by Harry Frankfurt,[21] which were intended to argue against the common thesis that moral responsibility necessarily requires the ability to do otherwise for the agents.

Frankfurt's examples involve agents who are intuitively morally responsible for their behaviour even though they lack metaphysical alternative possibilities. Frankfurt's argument was developed exactly to argue for moral responsibility; however, it is intuitively plausible to affirm that in Frankfurt-scenarios agents also act in a *free* manner. Here is a typical Frankfurt-scenario:

> In the next election, Donald is likely to vote for the Democrats; in fact, he will not vote in such a way only in one particu-

lar circumstance: that is, if he thinks about the possibility of American defeat in Iraq just prior to voting. Ms White, a member of the Democratic Party, wants to make sure that Donald votes Democratic, so she secretly put a device in Donald's head that, if activated, will force him to vote Democratic. However, as to not reveal her presence, Ms White will activate the device only if Donald will think about the Iraq War prior to voting, thus risking to not voting Democrat. As things happen, Donald does not think about Iraq prior to voting, and Ms White has no reason to activate the device: Donald votes Democratic of his own free will.

Such an example clearly seems to suggest that there are cases in which an agent does act freely even if he does not *practically* possess alternative possibilities in the actual scenario of events: Donald, in the previous example, could only vote Democratic, but he has nevertheless voted in this way in full consciousness and freedom.

Frankfurt-scenarios help in underlying that the ability to do otherwise is not a necessary requisite for exercising free will. By making use of a terminology introduced by John Martin Fisher,[22] it is possible to distinguish between two kinds of control which an agent can have when performing an action. If the agent possesses *regulative control*, he possesses the power to perform an action freely and, at the same time, the power to do otherwise; if an agent possesses *guidance control*, he only possesses the power to perform an action freely but not the power to perform another action instead.

In a Frankfurt-scenario, an agent acts freely but at the same time he is not able to do otherwise, due to an external intervention, which would block an alternative course of action by manipulating the agent's brain. In such a case, the actual course of action is a free one, and the agent is responsible for his choice in performing that action, even if he is not really able to do otherwise. In Fisher's terminology, in such a case the agent possesses guidance control while he clearly lacks regulative control. Therefore, in order to be a free agent, an agent only needs to possess guidance control over his actions.

I would like to suggest, then, that Libet's implicit endorsing of an incompatibilist position about free will is the actual reason of his deep worry about the awareness aspect of his findings. Libet's incompatibilist claim, according to which the presence of physical (probably deterministic) and unconscious causes lying behind our actions reveals the illusory of free will, can thus be rejected from a philosophical point of view, by rejecting incompatibilism and making use of Frankfurt's cogent counterexamples.

Against Libet: The epistemic freedom

In this section, I would like to suggest an alternative view to Metaphysical Freedom and argue that in order to exercise free will it is only necessary to possess a more modest ability, namely the ability to exercise *Epistemic Freedom* during the act of deliberation. My aim is then to point out that it is possible to maintain a different understanding of free will which is not influenced by Libet's experimental discoveries, because they are by no means relevant with respect to the existence of such ability for human agents.

The idea of epistemic freedom can be made clear by focusing on the process of deliberation. It seems that a typical human being is able to engage, most of the time, in deliberation and practical reasoning. In the process of practical deliberation, agents usually reflect upon and weight reasons for acting (or not acting) in a certain way. In the scenario brought about by Libet, it is alleged that an agent lacks the ability to do otherwise, because the causes of his actions are not consciously available to him.

However, it can be argued that even in such a scenario there is still a sense and a purpose for practical deliberation. Consider again Donald's situation. He has to choose

between voting Democratic or Republican, but this time, while he deliberates about this decision, he is also aware that he lives in a deterministic world and that his choice is already determined by his biological makeup and brain's functions. It seems that, even if this is the case, Donald cannot avoid choosing which party he is going to vote. This is because, even if he knows that determinism is true, and that his antecedent psychological and physical states are causally sufficient to determine his decision, it does not follow that he *knows* what decision he will make and so which party he will vote. He still has to make a decision.

More precisely, according to David Velleman what happens in a scenario like this is that there always are *different epistemic descriptions* of the future available to the agent when he deliberates.[23] In other words, Donald is entailed to say whenever he wants, because he will vote the party he wants. Of course in a deterministic scenario there is only one possible choice for Donald, so he will actually vote only one party: the one that is predetermined. However, the evident lack of a unique true answer for him to the question of what he will vote makes him feel that his future is really open, that is metaphysically open.

In such a scenario, even if Donald does not possess metaphysical alternative possibilities, he still possesses *epistemic alternatives*. Therefore, what happens in a scenario like this is that, in feeling that Donald's choice is open, one is mistaking epistemic for causal freedom.[24] All that is open is not what party Donald is going to vote, but rather, as David Velleman suggests,

> what you would be correct in saying you are going to have. You mistake your license to say any one of various things about what you'll have for the possibility that you'll have any one of various things.[25]

If this is the case, then it seems possible to maintain a different understanding of the notion of free will which does not presuppose the existence of a Metaphysical Freedom.

Moreover, I suggest that Libet's empirical data are not relevant at all with respect to the existence of an Epistemic Freedom for the agents during the process of deliberation. In Libet's experiment, the experimental subjects deliberate about raising a finger. Such a process of deliberation is in fact *free* with respect to their epistemic freedom: they possess the epistemic freedom, during the deliberation, of thinking and evaluating alternatives and to say that they will not raise their finger, even if they do not possess the ability to perform such different action instead. And such a possibility would remain true even if the empirical findings of Libet's experiment are correct, namely if the effective beginning of the action is situated in an unconscious brain activity which takes place before the conscious deliberation of the subject begins.

Haynes' experiment

In the previous section, I offered an alternative view about free will and argued that the existence of an epistemic freedom characterizes an understanding of the ability to act freely which does involve neither the existence of real alternative possibilities nor the necessity for the agent of being a metaphysical kind of cause of his actions. Moreover, I have shown that Libet's experimental results are not relevant in respect of such an alternative characterization of free will.

A recent experimental study conducted in 2008 by John Dylan Haynes and colleagues[26] seem to represent a potential worry for the existence of epistemic freedom for agents. The experiment's results are relevant for our purposes because they focus on the notion of *predictability.* Instead of showing, as in Libet's case, that our actions are in general terms determined by unconscious cerebral causes, the experiment aims to investigate if it is possible to predict future actions by looking at cerebral processes at work during the act of deliberation.

First, I will briefly describe Haynes' exper-

imental setting. Then, I will address the challenge of predictability for a satisfactory conception of free will. Accordingly, I will ask whether it is possible to describe a theoretically plausible notion of freedom even in the presence of revealed predictions. Although I hope to show that Haynes's experiment cannot be considered in the end as a real worry for epistemic freedom, I will however, for the sake of the argument, follow Haynes' directions in setting up an imaginary neuroscientific and Laplacian scenario, and ask what would remain of free will in such a picture.

The experimental setting

In Haynes' experiment, experimental subjects had to decide whether to push a right or left button. They were supposed to perform this action periodically and freely. Haynes used Functional Magnetic Resonance Imagine (fMRI), an advanced technique of analysis in neuroimaging, in order to look at patterns of activations in small regions of the brain, and see whether those patterns could be interpreted to predict future actions.

Experimental data suggest that there are two main regions of the brain which are involved in the decision process: front-polar cortex and parietal cortex. The decoding of these areas, by means of a dedicated software, shows that, on average, 7 to 10 seconds before the subjects press the button, neuroscientists are able to predict whether they are going to push the right or left button.

Empirical results are then interpreted sharply by Haynes. According to him, the controversy as to whether subjectively "free" decisions are determined by brain activity ahead of time is at least partially solved by his study, because the empirical results show that the outcome of a decision can be programmed in brain activity of prefrontal and parietal cortex up to 10 s before the appearance of awareness. In this sense, even if subjectively "free", our decisions are actually up to the early activity of the brain, thus reducing the activity of free will to a mere *illusion*.

The challenge of predictability

The result of Haynes' experiment is interesting exactly because the neuroscientists are apparently able to *predict* the outcome of an agent's decision, and they are able to do this a few seconds before the decision enters the subject's awareness.

What seems to happen in this scenario is a shift from determinism to predictability or, in other terms, from the possibility of an *external predictability* to the possibility of an *embedded predictability*.[27] On the one hand, determinism in itself only entails external predictability. This is the possibility in principle for an external observer, not part of the universe, to predict all future states of that universe. However, Haynes' experiment seems to go further and to suggest the existence of an embedded predictability. This is the possibility for an embedded subsystem in the universe to make a prediction about a future state of the universe itself.

At a closer look, however, Haynes' results can be shown to be at least partially harmless.

First, overall predictions success is only an average over a large number of trials. In other words, in order to get any interpretable result, experimenters need to consider the average over many trials. Accordingly, with respect to any individual trial, it is impossible to predict which button will be pushed by an experimental subject. Secondly, the percentage of success in predicting the outcomes of experimental trials is decisively lower than 100%. At a closer glance, it turns out that Haynes' predictions are successful about 55-57% of the trials.

Although the result is certainly significant, however, as Adina Roskies suggests, a slight step over the chance is not enough to assert the predictability of human actions and to infer from this that free will is an illusion.[28]

However, for the purposes of understanding how much the challenge of predictability can be harmful for free will, I would like to suggest the following thought experiment, designed to push the limits of Haynes' results

in predicting human actions.

Let's then imagine a Laplacian scenario in which every outcome of any decisions of every human being is actually predicted, and in which the resulting previsions are revealed to the agent before he acts. Whereas the Laplacian demon was imagined by Laplace[29] as a kind of God-like observer that could know the positions, velocities, and forces of all the particles in the universe at one time, and thus know every future state of the universe[30], the predictor in our fictional scenario is rather a predictions device, a sophisticated development of our current neuroscientific technologies. The difference between the classical Laplacian Demon and the prediction's machine is significant, because whereas the former only accounts for the possibility of an external predictability, the latter realizes the hypothetical possibility of an embedded one, by extending the range of predictability to the totality of human actions.

The question now is: how is it possible to preserve free will against a collection of revealed predictions?[31]

First, I would like to suggest that there are powerful reasons to believe that a scenario which realizes an embedded predictability is logically impossible, as the latter is the source of well-known paradoxes. Nonetheless, my suggestion is that such kinds of paradoxes are not to be considered as evidence in favor of the existence of a metaphysical free will. On the contrary, I would like to propose that it is possible to argue for a meaning of free will which is not challenged even by such an extreme scenario, in which the possibility of realized embedded predictability is taken for granted.

It has been argued by many authors that there are serious *epistemic* limitations on the ability of an embedded subsystem in a deterministic universe to make predictions of future events.[32] Such epistemic limitations arise, for example, because it is unlikely that a finite subsystem of the universe will ever be able to make exact measurements of the initial states of the universe.

More radically, it can be argued that until the moment of the choice it is impossible to make a prediction about that choice because, in order to provide such a prediction, one would need all the information until the moment of the choice *included*. In short, the objection rules out the very possibility of early predictions as impossible.

Moreover, authors have insisted that there are also fundamental and stronger *non-epistemic* limitations on the ability of any subsystem in the universe to make predictions on future behaviours of other subsystems embedded in the same universe.[33] Consequently, embedded predictability does not obtain even in the deterministic universe. The most powerful way of proving such conclusion is by constructing a paradoxical situation in which a subsystem of the considered universe makes a prediction which comes up in the end as necessarily *self-defeating*.

Such a paradoxical situation is explained by Rummens and Cuyper[34] by putting forward the hypothesis that these limitations arise because the predictions themselves are *physical events* which are part of the law-like causal chain of events in the deterministic universe. According to these authors,

> a general and complete causal uncoupling of our past and future activities from the predicted subsystem, would require, among other things, that we were capable of obtaining all the information needed for our predictions without actually disturbing the predicted system.[35]

However, due to the Uncertainty Principle, it is unlikely that a requirement of this sort can ever be met. Considerations of this sort, which seem to deny *in principle* the very possibility of embedded predictability, are certainly useful to put into perspective the (already partial) results of Haynes' experiment as a real threat to the existence of free will.

However, it seems correct to say at the same time that the inference from the lack of embedded predictability to the possibility of

human freedom is hard to be drawn. This is mostly because there are several elements which constitute the ability to act freely, and none of them is explained or captured by the mere demonstration of the impossibility of embedded predictability.[36] For this reason, in the next section, I would like to pursue the proposed thought experiment and suggest a characterization for the notion of free will which is not (or at least not entirely) challenged even if we allow the possibility of revealed predictions.

My aim is then to point out that it is possible to maintain a different understanding of free will which is not influenced by Haynes' experimental discoveries, because they are not relevant with respect to such characterization of free will.

Against Haynes: The sense of free will

Imagine that the following situation holds. I live in such a Laplacian world, in which every outcome of any decisions of mine is actually predicted by a prediction's device. Today, I am in front of a much serious choice: to carry on working on this paper or, instead, to go out for a relaxing walk under the sun of this early afternoon. Furthermore, I have in front of me a piece of paper with the relevant prediction made by the prediction's machine, and I read it carefully: it says that I will decide to go out for a walk. Indeed, through a process of deliberation, I effectively decide to take a pause and go out for a walk. Now, is it possible to consider such an action as freely performed?

To explain how it is conceivable to answer the affirmative, I would like to carry on a modified version of a line of argument introduced by Peter Strawson in his well-known paper *Freedom and Resentment*.[37]

Strawson's purpose, in this famous paper, is to overcome the problem of the compatibility between determinism and moral responsibility. His strategy consists in leaving apart common conceptual issues about the analysis of "freedom" and "responsibility", by considering what actually happens when we hold a person responsible. His argument is that our 'reactive attitudes' towards others and ourselves, attitudes such as gratitude, anger or resentment, on which moral responsibility is based, are *natural* and unchangeable.

That is, they cannot be considered from an objective stance. In this sense, why should we think that accepting the truth of determinism will change our stance towards reactive attitudes? First, we are not practically able to give them up, because they are too deeply rooted in our human nature. Secondly, according to Strawson, to give them up because of the truth of determinism is not even rational because, in practical terms, we would not receive any kind of benefit from such a rejection.

As for the case of moral responsibility, I would like to suggest that what is really impossible (and not even rational) to get rid of for human agents is the *sense of freedom* which we usually perceive during the process of deliberation. Such a sense of freedom, I suggest, is the minimum requirement for a satisfactory notion of free will.

To see this clearly, consider again the imagined Laplacian scenario and my choice of going out for a walk, even when the outcome of such a choice is available to me as a revealed prediction. In this case, if I do not see any plausible reason to not going out, and I effectively go out, there is a deep sense for which my choice would be perceived by me as actually *mine*. The prediction, from my point of view, just happens to coincide with the choice I weighted in a process of deliberation. As Strawson suggests, then, what we do when we consider a person (or ourselves) a *free agent* is to take a particular stance on them, a stance from which I cannot avoid to consider others and myself as a free agent. That is because our sense of free agency is so profoundly incorporated in our agential practices to render practically impossible its abandon, even in front of revealed predictions.

However, I suggest that there is another deeper reason to maintain that it is impossible

for human agents to abandon a sense of freedom during the act of deliberation. In the Laplacian scenario I am taking into account, the reason why I perceive the choice of going out for a walk as *mine*, even if a prediction of such decision is available to me, is that such a choice of mine is perfectly *in line* with my personality and my reasons for acting: I can recognize myself as present in such a decision.

I propose that a possible explanation for this can be found in the claim that agency possesses not only a physical dimension but also a normative one:[38] when we deliberate, from a first-person point of view we are forced to build a personal identity which counts as a normative regulation for our actions.

First of all, it should be noticed that the process of deliberation, which is essential in order to effectively act, is made possible by the reflective structure of our mind.[39] To say that the human mind is self-reflective is to underline a distinguishing feature of our cognitive architecture, namely the fact that human beings acquire, during their natural development, the ability to step back from impulses and to undertake a process of reflection upon perceptions. In this reflection, we are able to form *reasons* from perception, which are essential in order to engage in practical activities.

Then, when we undergo a process of deliberation to answer the question "what to do", the reflective structure of the mind is the source of a sort of "self-consciousness", because it compels us to have a conception of ourselves: there is a "unified self", a coherent personality, which is able to find reasons in favour or against most of the decisions of our everyday life.

I would like to suggest that, even in the presence of revealed predictions, from a subjective stance we would be compelled to feel that there actually is a "unified self", which chooses actions in accordance with our own conception of ourselves. It is under such a "practical identity", which is continuously shaped by the same process of deliberation, that a desire can be recognized as chosen *by us* as a reason for acting, thus counting as a normative regulation for our choices.

The sense of freedom during deliberation is then justifyed by the fact that, from a subjective and normative stance, we are forced to understand our decisions as *in line* with our practical identities. That we occupy such an intentional, deliberative stance most of the time, and that we consider ourselves as free and active agents through this, is not in any way a pragmatic choice. On the contrary, it is exactly the necessary outcome of how our agency is constituted: by learning how to weigh reasons in the process of deliberation, we also learn the ability of being, from a first-person and normative perspective, free agents.[40]

It seem then possible to argue that there is a strong sense in which we are structurally forced to perceive ourselves as active and free agents even in the imagined scenario brought about by Haynes' experiment. In this different understanding, free will is not regarded as a metaphysical and theoretical property of the self (as both Libet and Haynes seem to presuppose), but rather as a consequence of the reflective structure of our mind, as a natural ability of the agent. As a result, I aimed to suggest that there is a possible characterization of free will which can survive even the challenge of predictability. And if this is correct, Haynes' empirical results are by no means significant with respect to such an alternative understanding of free will.

Conclusions

The aim of this paper was to suggest a different reading of some well-known neuroscientific experiments about free will. While philosophical analyses of these experiments are usually focused on asking *whether* experimental findings represent a threat to free will, I rather asked *why* such empirical data are perceived by many as worrisome.

In answering this question, I showed that the reason is at its essence a philosophical one: experimental results are perceived as undermining the existence of free will because of one's taking for granted a particular under-

standing of such a notion. In the light of this, I aimed to suggest that the same success or failure of the experiments is strictly dependent on the notion of free will which is put through experimental investigation. I sketched a number of arguments aiming to show that it is possible to reject the experiment's results by simply suggesting alternative but equally plausible understandings of free will, which are not challenged by neuroscientific empirical findings.

In the case of Libet's experiment, I argued that Libet's implicit endorsing of an *incompatibilist* position about free will is the actual reason of his deep concern about its existence. Then, I went on to suggest an alternative view of free will which is not threatened by Libet's experiment at all.

In the case of Hayne's experiment, I argued that his conclusions are too rushed in evaluating as irremediably dangerous the challenge of *predictability* for free will. Indeed, after expounding why predictability is intuitively seen as a worry for free will, I showed how an alternative understanding of free will would remain untouched even in a Laplacian scenario in which an embedded predictability systematically obtains.

In short, this paper's main message is that it is not entirely clear what neurosciences are talking about when they talk about free will. This is because, crucially, 'free will' is said in many ways. To be sure, one can legitimately ask if the varieties of free will which I pushed against the empirical data are in the end satisfactory or plausible. The answer to this question, I suppose, is inevitably uncertain, especially from a philosophical point of view. However, the methodological point I intended to stress remains valid: without preliminary conceptual analysis, neuroscientific experiments on free will run the risk of being blind or to miss theoretically interesting points.

In this respect, conceptual inquiries have indeed the merit of investigating and clarifying what we are talking about when using a concept like that of free will. In this light, their bigger merit is probably the one of asking and discussing what the relevant features are of a notion of free will which is in the end worth defending[41] and even worth to put through an experimental investigation.

Notes

[1] See B. LIBET, C.A. GLEASON, E.W. WRIGHT, D.K. PEARL, *Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential) in the Unconscious Initiation of a Freely Voluntary Act*, in: «Brain», vol. CVI, n. 3, 1983, pp. 623-642; B. LIBET, *Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action*, in: «Behavioral and Brain Sciences», vol. VIII, n. 4, 1985, pp. 529-566.

[2] See, e.g., P. HAGGARD , M. EIMER, *On the Relation Between Brain Potentials and the Awareness of Voluntary Movements*, in: «Experimental Brain Research», vol. CXXVI, n. 1, 1999, pp. 128-133; M. MATSUHASHI, M. HALLETT, *The Timing of the Conscious Intention to Move*, in: «European Journal of Neuroscience», vol. XXVIII, n. 11, 2008, pp. 2344-2351.

[3] See, e.g B. LIBET, *Do We Have Free Will?*, in: «Journal of Consciousness Studies», vol. VI, n. 8-9, 1999, pp. 47-57; D. WEGNER, *The Illusion of Conscious Will*, MIT Press, Cambridge (MA) 2002; S. HARRIS, *Free Will*, Free Press, New York 2012.

[4] See, e.g., T. WOLFE, *Sorry, but your Soul Just Died*, in: «Forbes Magazine», vol. CLVIII, n. 13, 1996.

[5] For philosophical criticisms on Libet's experiments about free will see contributions in W. SINNOT-ARMSTRONG, L. NADEL, *Conscious Will and Responsibility*, Oxford University Press, Oxford 2011.

[6] See C.S. SOON, M. BRASS, H.J. HEINZE, J.-D. HEYNES, *Unconscious Determinants of Free Decisions in the Human Brain*, in: «Nature Neuroscience», vol. XI, n. 2, 2008, pp. 543-545.

[7] Libet's works on conscious intentions are a copious number. In this paper, I will particularly focus on B. LIBET, C.A. GLEASON, E.W. WRIGHT, D.K. PEARL, *Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-potential) in the Unconscious Initiation of a Freely Voluntary Act*, cit.; B. LIBET, *Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action*, cit.; B. LIBET, *Mind Time: The Temporal Factor in Consciousness*, Harvard University Press, Cambridge (MA) 2004; B. LIBET, *Do we Have Free Will?*, cit.

[8] See B. LIBET, C.A. GLEASON, E.W. WRIGHT, D.K.

PEARL, *Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential)*, cit.; B. LIBET, *Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action*, cit.

[9] In Libet's words: «The brain [...] "decides" to initiate or, at least, prepares to initiate [certain actions] at a time before there is any reportable subjective awareness that such a decision has taken place» (B. LIBET, C.A. GLEASON, E.W. WRIGHT, D.K. PEARL, *Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential)*, cit., p. 640.); «If the "act now" process is initiated unconsciously, then conscious free will is doing nothing» (B. LIBET, *Mind Time: The Temporal Factor in Consciousness*, cit., p. 62); «The conscious feeling of exerting one's will would then be regarded as an epiphenomenon, simply a byproduct of the brain's activities but with no causal power of its own» (B. LIBET, *Do we Have Free Will?*, cit., p. 55).

[10] *Ivi*, pp. 55-57.

[11] There are two major kinds of incompatibilists: those who accept the incompatibility between determinism and free will and deny human freedom (usually called *hard determinists*), and those who, in accepting the incompatibility of free will and determinism, also assert the existence of free will (usually called *metaphysical libertarians*). Libet's position, according to the analysis I am suggesting, stands in the first category: as a result of his empirical findings, he thinks that it is unlikely that free will exists. However, and complementary, he seems to think that in order for free will to obtain, it should be characterized as a *Metaphysical Freedom*, by getting himself close in terms of intuitions about free will to metaphysical libertarians.

[12] See, e.g., D. WEGNER, *The Illusion of Conscious Will*, cit.

[13] This position can be traced back at least to Descartes. The Cartesian concept of the mind states that the agent controls his own thoughts and actions because of the existence of a separated 'mental space'. In this respect, for Descartes only mental actions can be free, while mere actions of the body cannot be free, because they are governed by physical laws. In Descartes' words: «Now the action of the soul consists entirely in this, that simply by willing it makes the small [pineal] gland to which it is closely united move in the way requisite for producing the effect aimed at in the volition [...] when we will to walk or to move the body in any manner, this volition causes the gland to impel the spirits toward the muscles which bring about this effect» (R. DESCARTES, *A Discourse of a Method for the Well Guiding of Reason, and the Discovery of Truth in the Sciences* (1637), Thomas Newcombe, London 1649, §§ xli, xliii).

[14] For a possible reply to the latter flaw see S. GALLAGHER, *Where's the Action? Epiphenomenalism and the Problem of Free Will*, in: W. BANKS, S. POCKETT, S. GALLAGHER (eds.), *Does Consciousness Cause Behavior? An Investigation of the Nature of Volition*, Cambridge MA, MIT Press, 2006, pp. 109-124.

[15] See D. DENNETT, *Freedom Evolves*, Viking, New York 2003.

[16] The following passage from Libet is significant in this respect. «Finally, there is the most mysterious of these questions: how can the physical activities of nerve cells in the brain give rise to the nonphysical phenomena of conscious subjective experiences, which include sensory awareness of external world, thoughts, feelings of beauty, inspiration, spirituality, soulfulness and so on? How can the gap between the 'physical' (the brain) and the 'mental' (our conscious subjective experiences) be bridged?» (B. LIBET, *Mind Time: The Temporal Factor in Consciousness*, cit., p. 140, emphasis in the original).

[17] See A. ROSKIES, *Why Libet's Studies don't Pose a Threat to Free Will*, in: W. SINNOT-ARMSTRONG, L. NADEL, *Conscious Will and Responsibility*, cit., for a detailed version of such a criticism.

[18] This claim can also be criticized from a phenomenological perspective. Indeed, what Libet seems to presuppose is a doubtful phenomenological thesis, i.e. that we always consciously perceive the urge to act *before* the performance of a voluntary action. At a closer glance, however, it seems plausible that our ordinary volitional actions are not preceded by intentions which are always consciously available. Indeed, we can imagine different actions of which we are not immediately aware, but that we still are able to consider conscious and voluntary, as for example the act of drinking water, at a first glance unconsciously, while speaking animatedly. Examples of this kind are very common, and show that to consciously perceive the urge to act is neither a sufficient nor a necessary condition in order to consider an action free and voluntary. However, Libet's belief that the presence of a conscious urge is a necessary condition for a free action is not avoidable, since it derives from the operational definition of

free will he has chosen: to perform a free action, the subject should consciously *feel* that she wanted to do it on her own initiative (see for such a line of argument A. LAVAZZA, M. DE CARO, *Not so Fast. On Some Bold Neuroscientific Claims Concerning Human Agency*, in: «Neuroethics», vol. III, n. 1, 2010, pp. 23-41).

[19] According to Libet, a satisfactory characterization of free will should contemplate that an agent possess the ability to choose different courses of actions. In fact, he proposes to posit the existence of a *conscious veto*. In order to 'save the phenomenon' of free will, he develops the conclusion that if the role of conscious free will cannot be the one to initiate a voluntary act by choosing between different possibilities, we can at least control whether the act takes place; free will can operate in this sense in the 200 ms which divide the appearance of the conscious will from the beginning of a bodily action. What we would have in this picture is, therefore, a sort of *free won't*: we can consciously stop the performance of already initiated actions. «The mere appearance of an intention to act could not be controlled consciously; only its final consummation in a motor act could be consciously controlled» (see B. LIBET, *Do we Have Free Will?*, cit., p. 55).

[20] R. KANE, *Oxford Handbook of Free Will*, cit., p. 11.

[21] See H. J. FRANKFURT, *Alternate Possibilities and Moral Responsibility*, in: «Journal of Philosophy», vol. LXVI, n. 3, 1969, pp. 829-839.

[22] See J.M. FISCHER, M. RAVIZZA, *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, Cambridge 1998.

[23] According to Velleman, intentions are a special sort of beliefs that the agent has the power to make true: an intention is the *self-fulfilling belief* that an agent has that she will perform action G as a result of having that very belief. Intentions as beliefs possess truth conditions: an intention of the form "I will G" is true iff I will indeed G. In this sense intentions, according to Velleman are self-fulfilling beliefs: one can affirm correctly different inconsistent propositions without being wrong in affirming them. Such a power is the power of epistemic freedom.

[24] As I will clarify in the next section of the paper, it is possible to easily confuse, even in philosophical arguments, between different issues: *determinism*, *mechanism* and *predictability*. While their consequences for free will are apart, they are often mixed up and treated as a whole challenge for free will.

[25] See J. D.VELLEMAN, *The Possibility of Practical Reason*, Oxford, Oxford University Press, 2000, ch. 2.

[26] See C.S. SOON, M. BRASS, H.J. HEINZE, J.-D. HEYNES, *Unconscious Determinants of Free Decisions in the Human Brain*, cit.

[27] See T. HONDERICH, *A Theory of Determinism. The Mind, Neuroscience, and Life-hopes*, Clarendon Press, Oxford 1988; K. POPPER, *Indeterminism in Quantum Physics and in Classical Physics*, in: «The British Journal for the Philosophy of Science», vol. I, n. 2-3, 1950; S. RUMMENS, S. E. CUYPERS, *Determinism and the Paradox of Predictability*, in: «Erkenntnis», vol. LXXII, n. 2, 2010, pp. 233-249.

[28] See A. ROSKIES, *Why Libet's Studies don't Pose a Threat to Free Will*, cit.

[29] Before Laplace, Gotfried Leibniz imagined a scientist who could see the events of all times, as they are present to the mind of God. «Everything proceeds mathematically [...]if someone could have a sufficient insight into the inner parts of things, and in addition had remembrance and intelligence enough to consider all the circumstances and take them into account, he would be a prophet and see the future in the present as in a mirror» (see G. LEIBNIZ, *The Leibniz-Clarke Correspondence*, edited by H.G. ALEXANDER, Barnes & Noble, New York 1956).

[30] See P. LAPLACE, *Essai Philosophique sur les Probabilités*, V. Courcier, Paris 1814, p. 4 : «We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes».

[31] Such hypothetical scenario is also called, in the analytic literature, the problem of the "Book of Life": one can imagine having beside her a book which contains all the revealed predictions relevant to her entire life. If this book could be constantly accessible to her, what would the consequences be on her ability to act freely?

[32] See, e.g., P. SUPPES, *Explaining the Unpredictable*, in: «Erkenntnis», vol. XXII, n. 2-3, 1985, pp. 187-195; R.C. BISHOP, *On Separating Predictability and Determinism*, in: «Erkenntnis», vol. LVIII, n. 2, 2003, pp. 169-188.

[33] See, e.g., K. POPPER, *Indeterminism in Quantum Physics and in Classical Physics*, cit.; M. SCRIVEN, *An Essential Unpredictability in Human Behaviour*, in: B.B. WOLMAN, E. NAGEL (eds.), *Scientific Psychology: Principles and Approaches*, Basic Books, New York 1965; D.A EVANS, P.T. LANDSBERG, *Free Will in a Mechanistic Universe? An Extension*, in: «The British Journal for the Philosophy of Science», vol. XXIII, n. 4, 1972, pp. 336-343.

[34] It is not in the purposes of the present work to go further in the analysis of the paradox of predictability. For further and interesting clarifications see S. RUMMENS, S.E. CUYPERS, *Determinism and the Paradox of Predictability*, cit., to which I am indebted for the present analysis of the problem of predictability.

[35] *Ivi*, p. 245.

[36] This line of reasoning follows the well-known claim that, from the truth of indeterminism, there is no direct inference to the existence of free will, until one is able to show *how* such an inference obtains.

[37] P.F. STRAWSON, *Freedom and Resentment*, in: «Proceedings of the British Academy», vol. XLVIII, 1962, pp. 1-25.

[38] On agency as a self-constituted, normative phenomenon (in a neo-Kantian fashion) see C.M. KORSGAARD, *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, Oxford University Press, Oxford 2008; C.M. KORSGAARD, *Self-Constitution: Agency, Identity, and Integrity*, Oxford University Press, Oxford 2009.

[39] The feature of self-reflection was importantly characterized by Harry Frankfurt as the most distinguishing feature of human persons in H. FRANKFURT, *Freedom of the Will and the Concept of a Person*, in: «Journal of Philosophy», vol. LXVIII, n. 1, 1971, pp. 5-20.

[40] The line of argument I am pushing in this last paragraph is based on the main argument that I am currently developing for my PhD thesis. There, I hope, I will be able to clarify the plausibility of a compatibilist account of free will based on a normative conception of the phenomenon of agency.

[41] See D. DENNETT, *Elbow Room: The Varieties of Free Will Worth Wanting*, MIT Press, Cambridge (MA) 1984, for an analysis of some varieties of free will which are really worth wanting.