

THE LIMITS OF THE NEUROSCIENCE OF MORAL RESPONSIBILITY¹

DANIEL DE VASCONCELOS COSTA²
(UFRJ/Brasil)

ABSTRACT

The findings of the neuroscientist Benjamin Libet are among the most discussed in moral philosophy. They present a clear challenge to the notion of intentional action as a consciously chosen action. According to them, the awareness of the decision to act by the subjects of his studies came only after the moment of preparedness of the action in our brains, called "readiness potential". Many, including Libet, saw these results as an evidence that we do not have free will nor moral responsibility. The aim of this article is to criticize the claim that moral responsibility would be in danger because of the Libet's findings. First, the concept of free will as intentional action will be explained in order to understand how the notion of being conscious in deciding when and how to act is relevant. Then, the findings from Libet's experiments and the argument of how they could be a challenge to the notions of free will and of moral responsibility are presented. At the end, it will be argued that the notion of moral responsibility involves more than psychological capacities, but, foremost, the attribution of social roles in a moral community.

Keywords: Neuroethics; Free will; Moral responsibility; Benjamin Libet.

Introduction

It would not be an overstatement to affirm that we are in the middle of a *neuroscientific turn* in moral philosophy (LEEFMANN & HILDT, 2018, 14-19). It is undeniable that discoveries in neuroscience have a great effect in how philosophers think about ethical questions. There are, at least, four different ways of how these discoveries could affect moral philosophy: (i) they could provide us an *empirical basis* for *moral analysis*; (ii) they could present us new *challenges* to important premises to the philosophical tradition; (iii) the neurosciences could also provide us *new foundations* for the comprehension of moral behavior; (iv) and, finally, there is the possibility that, in the end, neurosciences could, and perhaps should, *replace ethics* as the discipline *par excellence* about moral behavior.³

One of the most significant neuroscientific contributions for the ethical investigation was made by Benjamin Libet.⁴ In a series of experiments, he had found out that the *feeling of voluntary control* of the act came

milliseconds (msecs) after the decision to act had already been made (LIBET, GLEASON, *et al.*, 1993; LIBET, 1993b; LIBET, 2011). This discovery was taken by many as one of the most significant challenges to moral philosophy, in special, to the attribution of *moral responsibility*, since his findings were taken as evidence that the human action was just a consequence of neurophysiological mechanisms.⁵

The attribution of responsibility is one of the main elements in our moral judgments. When we ascribe someone responsibility about some event, we feel ourselves justified to criticize (or praise) them and even punish (or reward) them for it. Given its importance to our moral practices, it is not a coincidence that the concept of responsibility was one of the most investigated moral concepts through the history of philosophy. Indeed, if Libet is correct and our behavior is truly decided by brain mechanisms over which we have no direct control, we have in our hands a challenge to the practice of attribution of responsibility and to the philosophical concept of moral responsibility.

The aim of this article is to analyze this challenge to moral philosophy and to propose that it is not as much a challenge as a demonstration that our concepts of human agency, intention, and responsibility, cannot be well understood without the appreciation of their social aspects. First, it is very instructive to describe the compatibilist solution to the free will problem, and how the concept of intention plays a part in it, for only against this background we can appreciate Libet's argument in its fullest (I). Then, we are going to examine his findings and how they posed a challenge to our traditional concepts of free will and of moral responsibility (II). At the end of this paper, we intend to analyze the concept of moral responsibility in order to show how it is more complex than conceived by Libet and many philosophers (III).

I

In a sense, the *free will problem* could be described as the question whether our actions happen because we *wanted* it to. There would be an *intrinsic connection* between our *wanting* and our *actions*, so that when we want them to happen, they happen.⁶ Our will would be free *only if* this intrinsic connection holds. However, not only we conceive that this intrinsic connection between our actions and our wanting is real, but we *also* picture ourselves as living in a universe where all entities are *causally connected* and this causal connection is subjected to *constant laws of nature*.⁷ This causal connection among all entities can be understood through what Jaegwon Kim named as "the causal closure of the physical domain" (KIM,

1993, 280; BISHOP & ATMANSPACHER, 2011, 102-103; BISHOP, 2011, 604). This idea basically says that: for every *physical* event in the world there is a *preceding physical* event that is *causally sufficient* for the occurrence of the former.

In itself, the causal closure does not say anything about the laws that rule this causal relationship. Nevertheless, we could assume that Kim and others that espouse physicalism would adopt the laws of nature discovered by the natural sciences. Hence, the assumption that the causal closure principle (CCP) would include the laws of nature is strong. We could reformulate the CCP as expounded above in the following way: for every *physical* event in the world there is a *preceding physical* event that is *causally sufficient* for the occurrence of the former in the way *stated* by the *laws of nature*. With this description, we have now a causal connection among all entities governed by physical laws. Therefore, the history of the causal connection *could not have been different*, since the constancy of these laws allows us to make predictions about the future (BISHOP, 2011, 605). The consequence of the conjunction between the causal closure and the laws of nature creates the thesis of *physical determinism*.

Also, according to Kim, we could infer another principle from the CCP, namely, the *principle of causal exclusion* (1993, 281; KIM, 2005, 17). His reasoning is that the CCP would not be compatible either with *partial* or *independent sufficient causation*. This incompatibility has as consequence the *exclusion of mental causation* from a complete, sufficient causal explanation. Therefore, from the acceptance of the CCP, we see ourselves also committed with the principle of causal exclusion (PCE). For Kim, this argument shows definitively that a nonreductive physicalism, where mental categories could play some role in the causal explanation of behavior, is not possible. Although, *prima facie*, mental categories need not be eliminated from our explanatory vocabulary, there would not be also any good reason for maintaining them.

This is the traditional way of understanding the free will problem, and it is not a new one. Almost all modern philosophers cared about this question or, at least, a version close enough to it (O'CONNOR & FRANKLIN, 2020). Most likely, this persistent willingness to deal with this problem is because it is not just an interesting theoretical problem for the philosophers, but it has ramifications over many of our social practices, such as morality and the law. Arguably, if we are not truly able to control our actions or if these are caused by elements outside our power, it would be unfair to be punished for something out of our control. After all, if we cannot control our actions, how can we be held responsible for them?

There are two ways of dealing with this problem if we want to defend the possibility of free will. The first is to deny determinism and to defend that we are, as agents, not subjected to causality in the way as formulated by the physicalism.⁸ The other is to deny that there is a dilemma; not only determinism and free will are compatible but also, in some understandings, the notion of free will makes sense only when coupled with some notion of determinism (HUME, 2007, 2.3.1-2; AYER, 1972, 275-276). In an almost obvious way, this position is called *compatibilism* and those that defended it were known as *compatibilists*. What makes the compatibilism interesting is that it does not deny physical determinism, but it tries to *disarm* it, showing how both the ideas of determinism and of free will could coexist. Furthermore, its focus on the human agency puts it nearer to the question of moral responsibility, while the search to deny the thesis of determinism says nothing about it and is purely metaphysical.

Aristotle pointed out that we connect our moral practices to our capacity of *voluntarily* acting, and if our actions are out of our control, the practices of blaming or praising us for them or their consequences would be pointless or, even, unfair (ARISTOTLE, 2011, 1109b30-1111b3). He divides our action between voluntary and involuntary. In a first approach, he conceives the voluntary action *negatively*, that is, he begins with what could make an action *involuntary*. An action would be involuntary if it was *forced* or the result of *ignorance*.

The criteria he uses to define when an action was forced or the result of ignorance help us to comprehend what voluntary action *is*. A forced action has its origins in an *external* source, that is, the springs of the action cannot be found in the agent, but in something external to her. Now, an action was the result of ignorance when the agent acted *not knowing* all the *relevant aspects* for her action. According to Aristotle, the ignorance causes feelings of regret or pain because, when the agent realizes that her action failed on account of her ignorance, she feels bad. This probably happens because had she taken more care about how to act, about the necessary information for the action to be successful, she would have achieved what she had aimed.

Taking these two criteria about involuntary action in consideration, we are able to formulate a *positive* conception of action in Aristotle. An action is voluntary if it had originated from an *internal* source *and* the agent was *well-informed* about how to act in order to achieve what she wants. Of course, what would count as internal source or well-informed can be tricky, however, compatibilists also said something, at least, similar to him as they tried to answer these questions.

Arguments from A. J. Ayer (1972), Harry Frankfurt (1998a; 1998b; 1998c), John Martin Fischer and Mark Ravizza (1998), among others, bear

resemblance with what Aristotle proposed; they all point to the same aspect of our attribution of moral responsibility in normal contexts, namely, the *voluntarily or intentional action*.

They conceive voluntary action as one guided by the *will*. The action is the way it is because the agent *wanted* it to. Thus, an agent would be morally responsible for her action only if there was not any influence on it other than *her own will* (AYER, 1972, 277-280; FRANKFURT, 1998b, 23-25). As in Aristotle, this precludes different kinds of external forces, such as coercion or compulsion, which move our bodies but that we cannot truly see them as motivations of our own. External forces *impose* a behavior on us. This can be true even in cases where an agent sees a desire as hers. In these cases, although she sees it as hers, she does not see it as *her own*. Her desire is perceived as something *external* to herself, to her identity. Compulsions act exactly like this; even though it is her desire to do *x*, the agent does not truly *want* to do it, but she is *made* to act like it by this desire. Her actions are *internal* when they are *her own*, when they *are*, in a certain sense, *herself* (FRANKFURT, 1998c, 164-172).

It is exactly this identification of an action with the agent made by herself that explains why it is necessary for the agent to have some *conscious* relationship with her own actions (FRANKFURT, 1998c, 162-164). When she acts for a *reason* of her own, she behaves in a way that she *recognizes herself* in that action. It was her *decision* to act in this way, and this decision is also a statement of who she is. When she *knows* that her action *represents* what she wanted and she knows *why* she acted, she is the *source* of her action. About this point, Frankfurt says:

The decision determines what the person really wants by making the desire on which he decides fully his own. To this extent the person, in making by which he identifies with a desire, *constitutes himself*. The pertinent desire is no longer in any at external to him [...] It comes to be a desire that is incorporated into him by virtue of the fact that he has it *by his own will* (FRANKFURT, 1998c, 170).

However, when she does not *see* herself in the action and has *doubts* about this action being what she *truly* wanted it to, that means that her action does not make any sense to her, even though it was her body that produced it. In this sense, she acts and she knows that she acts, but, at the same time, she does not know *why* she acted in that way. She does not *feel* as if this action was her own. When her own intentions about the ends, and even about the means, cannot be seen in these actions by her, the agent does not have any sense of *ownership* over them. That is why they are felt

as if *external* to her. To be *conscious* about *why* she acts is, therefore, a *necessary condition* for an agent to own her actions.

Without this self-consciousness about her actions and the reasons of why she acts, a person would not be an agent in the sense that she could act through her own will. Talking about this kind of reflexivity that marks any voluntary action, Frankfurt says the following:

There is also another sort of reflexivity of self-consciousness [...] It is a salient characteristic of human beings, one which affects our lives in deep and innumerable ways, that we care about what we are [...] We are particularly concerned with our own motives. It matters greatly to us whether the desires by which we are moved to act as we do motivate us because we want them to be effective in moving us or whether they move us regardless of ourselves or even despite ourselves (FRANKFURT, 1998c, 163).

II

Based on what we have seen above, we can say that the model of action presupposed by the traditional accounts of moral responsibility has as one of its tenets the notion that any agent is *aware* or *conscious* of her actions, her reasons to act, and acted *consciously* on these reasons. In a certain sense, we could say that an agent is always someone able to give a proper response to the question: "*Why* did you do this?".

However, this paradigm of voluntary action as conscious action, and the notion of moral responsibility related to it, came under attack. Neuroscientific studies were heralded as the beginning of an empirical guided ethics and the abandonment, or a scientific led reconceptualization, of folk concepts of our ordinary moral practices, such as volition, intention, punishment, and, specially, free will and moral responsibility (RACINE & DUBLJEVIĆ, 2017, 405-406; LILIENFELD, ASLINGER, *et al.*, 2018).⁹ Some of the most known and commented studies in neuroscience that directly resonate on moral philosophy were those from Benjamin Libet (1993a). Through his studies he had shown that *actions were not consciously chosen*, for the conscious phenomena of the decision to act appears *only after* the *preparation* to act has already begun. The conclusion made by him and others was that we are not truly free to decide *when* and *how* we will act, at least, in a positive sense.

In an article clearly aimed to a philosophical audience, where he avoided a heavily scientific loaded language, Libet summarizes these findings and philosophically analyzes them (2011). Without doubt, one of

the most important neuroscientific discoveries related to the voluntary act would be the “readiness potential” (RP) (JAHANSHAHI & HALLETT, 2003, 1-2, 5-7; SHIBASAKI & HALLETT, 2006).¹⁰ The RP is an electrophysiological pattern that signals the *preparedness* of the agent to *voluntarily* act. The RP is just one of different electrophysiological patterns that indicate specific processes that happen in the brain (DONCHIN, 1979, 24; KAPPENMAN & LUCK, 2012, 3; FABIANI, GRATTON & FEDERMEIER, 2007, 87-88). Seeing, hearing, thinking, speaking, etc., could be distinguished from each other through their patterns. Also, movements have their own electrophysiological patterns, and they can be recorded through *electroencephalography* (EEG), a modern medical method capable of measuring and recording the electrical activity in the brain.

In a very simplified description, we could say that the brain is a bundle of neurons that communicate with each other through electrochemical discharges, known as *action potential* or *nerve impulse* (CRAVER, 2007, 114-122; BAARS & GAGE, 2013, 61-64; STERNBERG e STERNBERG, 2012, 61-62). Any brain activity occurs through action potentials, but, since there are many possibilities for them to occur, action potentials are differentiated according to its activation, function, and other criteria. The electrical activity associated to the response of the agent to stimulus of internal or external environment are known as *event-related potential* (ERP) (BLACKWOOD & MUIR, 1990, 96). Often, the ERP reflects an ongoing neural process that has a specific mental event correlated to it.¹¹ The neuroscientists Steven Luck and Emily Kappenman say that: “In a general sense, we can define the term *ERP component as a scalp-recorded voltage change that reflects a specific neural or psychological process*” (KAPPENMAN & LUCK, 2012, 4). This definition can also be found in other articles: “Conceptually, ERPs are regarded as neural manifestations of specific psychological functions” (FABIANI, GRATTON & FEDERMEIER, 2007, 85). If we follow them, we can say that when we act, feel pain, or think about something, there are ERPs that can be recorded by EEG. The measurement appears in waveform that represents the electrical fluctuation of the ERP, which indicate the activity of a neural process in a time frame.

It seems to be an agreement among neuroscientists that the ERP can be classified into two kinds: *exogenous* or *evoked* ERP and *endogenous* or *cognitive* ERP (DONCHIN, 1979, 26; FABIANI, GRATTON & FEDERMEIER, 2007, 86-87). Exogenous ERP are those, whose response is *necessary* when the adequate *external* stimuli happen. For instance, if we have a functional visual system and we are with open eyes, it is not possible not to see what is in front of us in case of sufficient and adequate stimuli. We do not have the power to control their occurrence and the exogenous ERPs happen with

environmental stimuli and a functional neural basis of the affected bodily system. We could say that this follows the stimulus–response model.

However, in endogenous ERPs, the system is not passive and it has an effective participation in the occurrence of the ERP, for it *actively interacts* with the stimuli, instead of only responding to them. Endogenous ERPs are not evoked, but rather *invoked*, since they are not merely response to environmental stimuli, but an *interaction* to them. The system as a whole organism *steers* its own behavior in front of its necessities and contingencies presented by the environment:

[...] evidence accumulated that a variety of processes that are *invoked* by the psychological demands of the situation rather than *evoked* by the presentation of stimuli manifest themselves on the scalp in the form of distinct components of the ERP [...] Of particular importance has been the weakening of the stimulus-response paradigm. No longer is psychology bound by the concept of the organism as a passive system, inertly lying in wait for “stimuli,” upon whose arrival it rises to emit responses [...] Contemporary models tend to view the organism as continuously processing information: hypotheses are generated and tested, expectations are established and revised, decisions are made and acted upon (DONCHIN, RITTER & MCCALLUM, 1978, 350).

The readiness potential is one of the many endogenous ERPs that can occur in our brains. More specifically, it is also *just* one of the possible ERPs that can happen with the operation of the *motor system* (FABIANI, GRATTON & FEDERMEIER, 2007, 94-98). In the case of potentials related to motor events, they can be, at least, of three different *modes*, according to the moment of their occurrence in the event to which they are related, in this case, any bodily movement. There is (i) the moment *prior* to the onset of the movement, a *preparation* moment of the movement; (ii) the *actual* onset of the movement, with the *prior* potential generated to the working of the muscles; and (iii) an *evaluative* moment that happens *after* the movement, in search of necessary motor adaptations to the next movements. In the case of *voluntary movements*, some of the most known ERPs involved in each of these moments are, respectively, the *readiness* potential (RP), the *motor* potential (MP), and the *reafferent* potential (RAF) (KORNHUBER & DEECKE, 1965; DEECKE, SCHEID & KORNHUBER, 1969; FABIANI, GRATTON & FEDERMEIER, 2007, 94-96; SMULDERS & MILLER, 2012, 209-212).

Leaving aside innumerable particularities about each one of these potentials, we can say that the occurrence of a voluntary action begins with the RP, followed by the MP, and, lastly, the RAF. The RP begins before the event that it is related to, the voluntary movement. It prepares a response from the motor system that comes to be with the MP, which generates the movement. After the movement, there is feedback stimuli generated by the relation movement-environment to correct the following movements according to new contingencies. This informational analysis of this relation is reflected by the RAF. In order to illustrate this, we can see in the following figure (FABIANI, GRATTON & FEDERMEIER, 2007, 95) the different moments in which each one of these ERPs appear relative to the onset of the movement signaled by the msec 0:

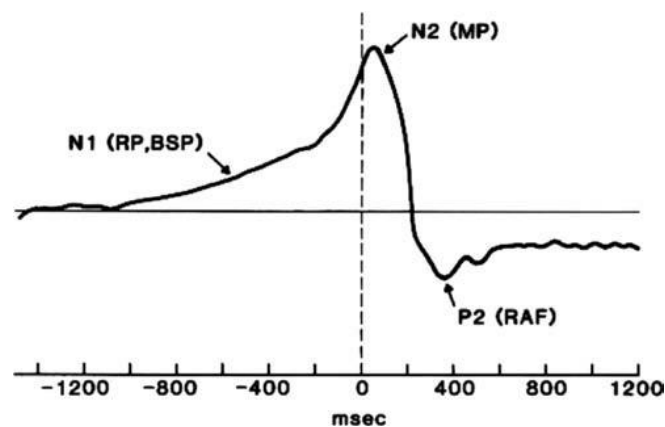


Figure:

By itself, the RP represents no problem to the conception of voluntary action as a *consciously chosen* action. The natural supposition would be that the action is consciously chosen *before* the RP. This supposition goes as the following: since the RP is the preparedness of the brain to act, the action itself was already chosen, for the preparation to act presupposes the choice of *how* the agent will act. But, if the action was already chosen and the moment of *voluntary* choice is *conceptually* taken as a *conscious decision* by the agent, then the moment of the conscious choice could *only* have happened *before* the preparation to act. This natural assumption is accepted by Libet (2011, 2): "In the traditional view of conscious will and free will, one would expect conscious will to appear before, or at the onset of, the RP, and thus command the brain to perform the intended act".

The focus of Libet's studies was the *moment* when an agent becomes *aware* of her *intention to act* (W) and compared W with the RP. If the action is truly voluntary, the agent should be aware of the moment of her intention to act, that is, the moment of the decision to act that *triggers* the action, before the onset of the RP. He recognized that the results of his experiments

could have a big impact on how we conceive our notions of free will and of moral responsibility:

The brain was evidently beginning the volitional process in this voluntary act well before the activation of the muscle that produced the movement. My question then became: *when* does the *conscious* wish or intention (to perform the act) appear? [...] It was clearly important to establish the time of the conscious will relative to the onset of the brain process (RP); if conscious will were to *follow* the onset of RP, that would have a fundamental impact on how we could view free will (2011, 2).

For his experiments, he *operationally* defines voluntary action as an action that fulfils three conditions (LIBET, 1993b, 269-270; 2011, 1): (i) it is an *endogenous* stimulus; (ii) without *external restrictions*; lastly, (iii) the agent has an *introspective feeling of intentionally initiating* the action. The first condition remembers what Aristotle had already said, and that was adopted by the tradition, that a voluntary action is one whose source is internal. Libet presupposes that the voluntary action cannot be initiated *spontaneously* by any *external* stimulus, as in the case of exogenous ERPs. Although it can be a reaction to the contingencies of the environment, the action is not a merely response of the agent. Therefore, this condition could be comprehended as putting aside any action that is a consequence of a *compulsion*.

About the second condition, once again we remember Aristotle and the tradition after him. In certain cases, we could say that the action was endogenously initiated, but we could not say that it was *truly* voluntary, because external factors influenced the agent in such a way that she could not choose any other action but that one. This reminds us of how Aristotle and others said that the use of *coercion* eliminates the voluntary choice of the action, and the second condition could be understood in this way.

The third and final condition says nothing about the structure of the voluntary action, but its *phenomenology*. There is a *feel* of deciding to act. Agents conceive themselves as the source of their action. But this self-conception is not only cognitive; it also involves something *qualitative*, a *subjective* side of the experience of being *free*, of choosing one's own action, not possible to be reduced to reasoning. The agent *knows that* she is free, but she is not able to explain it with words alone, for it also has a *feeling* of being free. Clearly based on Thomas Nagel (1979), Libet says: "Conscious subjective experience, in this case an awareness of the endogenous urge or intention to move, is a primary phenomenon; it cannot be defined in an a

priori way by recourse to any externally observable physical event [...]” (1993b, 272).

With respect to the RP, Libet’s findings were not different from those who had investigated it before him (LIBET, 1993b, 270-274; LIBET, 2011, 2-4; LIBET, GLEASON, *et al.*, 1993, 258-261). The average RP found in his studies was about –550 msec before the act. However, the moment of the awareness of the intention to act (W) from his subjects was, in average, –200 msec before the onset of the MP. He had discovered that the preparation to act, which was the onset of the RP, was, in average, –350 msec *before* the conscious decision to act. The feel of choosing to act, the “I will do *x* now”, came *after* the preparation to act. There was no *effective conscious* decision to act, for the action was already decided before their awareness of it.

As if acknowledging how the impact on our moral practices could be without the notion of free will, Libet tries to rescue this notion, even if it had to be modified (LIBET, 1993b, 276-278; LIBET, 2011, 4-6; LIBET, GLEASON, *et al.*, 1993, 266-267). Although he does not deny the findings of his experiments and stays behind the conclusion that our actions are not voluntarily initiated, since this would require a conscious decision to act, he still sees a way of how human beings could exert their conscious will on their actions. Instead of a *trigger* control or what we could call, adopting some of Fischer’s and Ravizza’s terminology, a *positive guidance* control over our actions, we had, nonetheless, a *veto* or a *negative guidance* control over them.¹²

Libet remarks that we become conscious of our decision to act before the act itself. As seen above, in average, W comes –200 msec before the MP. Theoretically, this would still allow the agent to exert some power over the way she acts. The conscious decision would have to be made between this –200 msec and the MP. The human agency would be conceived not anymore as a kind of power to consciously initiate an act, but more as a switch through which an agent could consciously *allow* or *stop* this volitional process, as if it could block the RP, preventing the onset of the MP. He even says that this veto guidance control is in accord with some religious and secular views of moral responsibility and free will.

However, even if Libet still believes that a kind of *negative* free will was still possible, that is not the same verdict of many philosophers or neuroscientists. Not only proponents of a “free will free” world see Libet’s findings as damaging to the folk notions of free will and of moral responsibility, but also proponents of the human capacity of acting in accordance with one’s own will saw them as potentially damaging to our

moral practices. They could be even more problematic for the notion of free will than the idea of determinism.

As shown above, Ayer's and Frankfurt's arguments are attempts to answer to the dangers of physical determinism. Nonetheless, the compatibility of determinism and free will that these authors defended is based on the human capacity of having a free will or, at least, a will of one's *own*. A will that is free is a will capable of choosing what it wants. Similarly, to be morally responsible means to act through a will of one's own. But Libet's findings put a question mark whether we can truly have a will of our own in a positive way. If correct, they could take out the ground from the compatibilists' arguments for free will and moral responsibility, for, now, even our intentions would not be truly ours, but only the consequence of a neurophysiological mechanism that human beings have. If compatibilism is an attempt to answer to the dangers of physical determinism, what could rescue the notion of moral responsibility found in our moral practices?

There are, at least, two ways of trying to save our folk understanding of moral responsibility and, perhaps, free will from Libet's conclusions. In a completely expected way, the first is to rebut his findings. This could happen either through new studies that prove them wrong or through epistemological criticisms of his experiments.¹³ This would cast an overall doubt on his findings as such or as the result of faulty experiments. This first way would save our folk moral concepts by eliminating what put them at stake. The second way is to show that, even if his findings were correct, they would not be a danger to our moral practices, for Libet left out of his experiments many aspects of the notions of free will, of moral responsibility, and even of the human agency necessary for the correct comprehension of them, in special, their *social* aspects. In the following section we are focusing only on the folk concept of moral responsibility.

III

One of the most insightful examinations about the *social* nature of moral responsibility can be found in the article "Freedom and resentment" from P. F. Strawson (2008). It is also one of the earliest attempts to understand whether the concept of freedom could be compatible with the notion of determinism. His aim was to comprehend why freedom is important to our moral practices and to show why determinism is not a threat as many think it is. Similar to Ayer and Frankfurt, he recognizes that freedom is important to us because it seems to underlie our practices of moral condemnation and praise, and of punishment. How we understand it in our moral practices, freedom is connected to our folk notion of moral

responsibility. However, instead of searching for the necessary or sufficient conditions of the morally responsible agent, as Ayer and Frankfurt did, he analyses *how* our moral practices *actually* happen and *what* our attribution of responsibility involves in order to see *how* freedom contributes to it.

Strawson speaks about the gamut of attitudes we normally have when interacting with others (2008, 5-7). Many of these attitudes, if not most of them, are *non-detached attitudes*, that is, attitudes infused with emotions and feelings. We can love, hate, desire, suspect, resent, respect, be grateful of, forgive, etc., others. Some of these feelings do not need to presuppose some interpersonal relationship. For instance, love or hate do not need an actual relationship between two people for someone to be affected by them. However, there is a kind of feelings that arouse *only* when we find ourselves in an *interpersonal relationship*. They appear as a *reaction* to the actions or attitudes of others toward us. We can feel thankful for someone's help; we can be angry with another for breaking something of us, etc. That is why Strawson called these attitudes as *reactive attitudes*.

If someone injury us in some way, at a first moment, we could become angry at this. But, if we find out the one who injured us was a young child or a dog, most of us would judge that they had no intention to it, and it was just an accident. Our anger would, most likely, subdue quickly in this scenario. However, in the case that the one who injured us was older enough for us to consider her having some kind of discernment, the angry we feel could not fade away simply with time. Actually, they could evolve to *another* kind of sentiment depending on the response we receive from her. If she says that she wanted to injure us, or shows us no *following consideration*, not excusing herself or denying some compensation to us in the case of material damage, our feelings can, and normally they do, evolve to another. Now we feel ourselves *wronged*, not merely injured by the other person. Our anger becomes *resentment*. Resentment carries this feeling of wrongness toward us and is, somewhat, moral, since we *feel* that she *disregarded* us as someone *deserving special consideration*.

Nevertheless, our moral practices allow more than point the finger to others. We can also *forgive* them. As resentment, forgiveness is not a reactive attitude that appears at a first moment. We forgive someone only after feeling other reactive attitudes and after some further consideration of the person who injured us and her following attitudes. According to Strawson, the act of forgiveness can be of two kinds (2008, 7-10). Normally, they can come after a special consideration made by the other toward us, or with the modification of the circumstances of our first reaction.

The first of these asks us to see the other as someone who did not have the *intention* or the appropriated *knowledge* to act the way she did.

In these cases, to say an excuse is to give reasons of why she did not do that intentionally. She could say that she has not seen us, or that she tripped over her own feet and hit us, or many others possible ways of expressing excuse or regret. Once again, Aristotle comes to mind with this kind of response. Strawson says basically the same as him, though with other words. Up to a certain point, the lack of intention or knowledge could be used as excuses to an error. This kind of excuse does not make us to see the other as someone that we could not hold responsible in other circumstances; she just made a mistake *this time*.

The second way of forgiving the other is not truly forgiveness, but rather a complete change of how we see her. Typically, we forgive the other because she asked us for forgiveness. Without some recognition that she committed a mistake and without some plausible excuse for why this happened, normally, we would not forgive her. However, when we see that the injury was made by some non-human animal, a young kid, or even an adult with diminished agency, our feelings change. We do not demand some excuse or an apology from them, for they are *not capable* of making them. Our attitudes tend to change, for we do not see them as *agents*. They cannot be held responsible for their actions because they lack the *minimal* criteria to be a full-fledged agent. Either they do not have the *cognitive* capacities to judge what they can do, what has value, etc., or they do not have a developed *control* of their *emotions*, so that they simply act on it. That is why we do not really forgive them. In a sense, they cannot even be forgiven.

These differences of excusing point to two distinct ways of treating others (STRAWSON, 2008, 10-12, 23). The relationship between our reactive attitudes and our interaction with others marks the existence of a *moral community*. When we are a member of this moral community, we are treated as someone deserving special consideration from the other members. We interact among ourselves with a *participant attitude*, through which we can be morally evaluated by others, who can, then, criticize or praise us for our actions and attitudes. Reactive attitudes toward others and ourselves is the *sign* that we find ourselves in this moral community and that, therefore, we can be held responsible for our actions.¹⁴ We forgive the other because she made a mistake and acknowledges it. Nonetheless, we still maintain our participant attitude toward her and treat her as a member of our moral community and as a full-fledged agent.

On the contrary, we excuse young children and non-human animals exactly because we perceive them as non-agents and non-participants of our moral community. Although they can be objects of moral consideration by the moral community and be treated as *persons*, they are not *members*

of it. They have no duty toward others for they have no capacity to exercise them. We almost see them as natural events such as a storm or tornado. They can behave, we can predict their behavior, but they do not act, for they are not agents. Seeing this, we tend to assume a *detached attitude* toward them. That is why our feelings of anger or disrespect normally subdue when we learn that our injury was caused by a young child or a non-human animal. Strawson names this kind of attitude as *objective attitude*.¹⁵

He stresses the fact that to take a participant attitude is something that we *actually do*; it is in our actual moral practice to conceive others morally responsible when they do something we consider morally wrong. He says that: “[...] participant reactive attitudes are essentially *natural human reactions* to the good or ill will or indifference of others towards us [...]” (2008, 10-11, emphasis mine). Attribution of moral responsibility is a normal moral practice; it is something we do. He does not try to explain how and why we do this; he just takes it as a fact “as we know them”.

He goes, then, to criticize those who think that the veracity of determinism has as consequence the abolition or, at least, a reconfiguration of our moral practices, in special, the attribution of moral responsibility. If to adopt a point of view of a determinist is the same as to adopt a point of view that there is no moral responsibility, for no one could truly act freely, then, this would call for a radical adoption of the objective attitude. Now, not only those whom we see as less than a full-fledged agent are seen through this perspective, but *all* of us. There is no need for the participant attitude because none of us can be an agent anymore. Strawson thinks that, at this point, the question becomes: is this overall adoption of the objective attitude feasible? He says it is not.

Strawson has two basic reasonings for it (2008, 12-14, 17-20). One is straightforward. The veracity of the thesis of determinism and of the thesis of incompatibilism are *theoretical*. He claims that, even if we are completely certain that both of these theses are true, this is a theoretical conviction, but the participant attitude is a *practical* conviction. Perhaps the best way of clarify this point is through a personal anecdote told by the philosopher Shaun Nichols:

From the first time I encountered the problem of free will in college, it struck me that a clear-eyed view of free will and moral responsibility demanded some form of nihilism. Libertarianism seemed delusional, and compatibilism seemed in bad faith. Hence I threw my lot in with philosophers like Paul d’Holbach, Galen Strawson, and Derk Pereboom who conclude that no one is truly moral responsible. But after two

decades of self-identifying as a nihilist, it occurred to me that I had continued to treat my friends, colleagues, and acquaintances as morally responsible. Hardly ever did I call on my philosophical views to excuse people's actions. *I'm increasingly inclined to think that my practice was appropriate and that it was my philosophical view that was defective* (2007, 405, emphasis mine).

As Nichols told us, even if he was *theoretically* convinced of the thesis of determinism and that this belief entailed some nihilism about our moral practices, he could not abandon them; *practically*, he continued to act as if he believed in those practices. Strawson invite us to try to put aside our moral practices and to adopt a sustained objective attitude, and to observe whether our reactive attitudes will, in fact, change. He is clearly convinced, as Nichols became later, that they will not.

This conviction takes us to his second reasoning, that the participant attitude and reactive attitudes are a *part of who we are as human beings*. He does not defend any kind of naturalism, at least, not in an explicit way. Nonetheless, we can see it there when he says, for instance, that:

The human commitment to participation in ordinary inter-personal relationships is, I think, too thoroughgoing and *deeply rooted* for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them; and being involved in inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question (2008, 12, emphasis mine).

To enter into interpersonal relationships and to have reactive attitudes are "deeply rooted" in us. We have a psychological makeup that make us unable to run away from our moral practices in the same way Nichols tried, but could not. They are ingrained in us and we define ourselves through them. Trying to abandon them is trying to abandon *who* we are. In the end, he says that determinism could actually be true, but our moral practices are inescapable.

Perhaps, the most important lesson of Strawson is that *the way we hold others responsible is a matter of our moral practices*. We see others as morally responsible because this is *what we do*. But, to see them as morally responsible means that we consider them as agents and as members of our moral community. Those who are perceived as incapable of being a full-

fledged agent are not included as members of our moral community. We adopt an objective attitude toward them for we do not see them as capable of *assume responsibilities* that the members of it normally do. Therefore, to ascribe moral responsibility is more than attributing some cognitive and affective capacity to others, rather, it is a matter of seeing them as *fulfilling* some *role* in our moral community. In the most general case, the role they fulfill is the role of being an agent.

If we observe our moral practices of holding others responsible, we will see that the most fundamental part of judging others responsible is a matter of considering them fulfilling a particular role that has some *role-specific duties*. For instance, parents have duties toward their children. In some societies, their duties are more than just to assure the welfare of their children, but also to provide them the better upbringing that they are able to. Parents that use their resources to their own satisfaction rather than to their children are perceived as *bad* parents. They do not correctly fulfill their roles as parents. Even in cases where they try to do whatever they can to fulfill them, they could be judged as not being good enough. In cases of childhood obesity, for instance, if the causes of it are unhealthy eating or lack of physical exercise, we could hold the parents responsible for it, for they should have done more for the health of their children.

Moreover, even if they tried to change the unhealthy eating or to establish an exercise routine to their children, simply by not achieving this, they could be seen as morally *failing* them. To say that they are morally responsible for their children is the same as saying that whatever happens to them, it is their responsibility. Such cases are extremely common; we judge others responsible when they are not able to fulfill the duties given by their roles. There are, sometimes, acceptable excuses for failing, but, in other times, there are not, even if they tried to fulfill them in the best way possible. In a similar way, to see someone as an agent is to attribute her the capacity of being morally responsible *in general*, that is, she has some duties to all members of her moral community due to her role as an agent. Many of our attributions of moral responsibility would not make sense if we do not consider this idea of responsibility as assuming a role. As seen above, the notion of negligence cannot be completely understood only by the notion of responsibility as intentional action. Many moral mistakes are not a matter of wanting to do something, but of *not caring about doing it*. Probably many of us would agree that not to do something to help someone who suffered an accident that could be fatal without any help would be morally wrong, even in cases where the agent is just a bystander who had nothing to do with the accident. But the fact alone that she did nothing to help the victim is still morally reproachable. She could even be held

responsible for the death of victim since she could have avoided it with a simple call.

In an interesting article, H. L. A. Hart points to the fact that our concept of responsibility has different senses (2008, 211-230). He mentions four different ways of how it could be interpreted. They can be classified as: (i) *role-responsibility*; (ii) *causal responsibility*; (iii) *liability-responsibility*; and (iv) *capacity-responsibility*.

Three of these senses are not as much categories of different kinds of responsibility as *criteria* of how we could attribute someone responsibility for something. Different of those, the category of liability-responsibility is better conceived as a generic term to say that someone is responsible for something (HART, 2008, 215-227). When we say to someone, “you are responsible for this”, according to Hart, in this moment, we are employing the notion of liability-responsibility. To say that someone *A* is liable or morally responsible for something *S*, is the same as saying that “according to our criteria *C*, *A* is morally blameworthy and could be punished for doing *S*.”

Hart claims that there are, normally, four criteria for saying that someone is liable-responsible for *S* in most legal or moral systems (2008, 218-221). Three of them correspond exactly to the other three senses of responsibility and the fourth condition is that *S* is morally condemnable or legally prohibited in our *moral* or *legal code*. About the necessity of this fourth condition to the attribution of liability, he says: “When legal rules require men to act or abstain from action, one who breaks the law is usually liable, according to other legal rules, to punishment for his misdeeds [...] He is thus liable to be 'made to pay' for what he has done [...]” (HART, 2008, 215). But just the description of a moral or legal wrong is not sufficient for holding someone responsible or liable, for we have to establish some *connection* between the agent and the wrongdoing. The other three senses of responsibility fulfill this role.

The first is the *causal* responsibility (HART, 2008, 214-215). This way of understanding responsibility is well known by many competent English, or Portuguese, speakers. When we say, for instance, “the climate change is responsible for the melting of the polar ice caps”, we just saying that there is a *causal correlation* between the climate change and the melting of the polar ice caps. Nonetheless, causal responsibility is not a necessary nor a sufficient condition to establish moral responsibility. In cases of negligence or omission, for instance, there is no causal relationship between the events of negligence or omission and of the moral or legal wrong. Also, for instance, although we can establish a causal connection between some broken object and a young child or a non-human animal, we would not impute them moral

responsibility. They are not considered capable to bear responsibilities. It is true, nonetheless, that some kind of relationship between the occurrence and its perpetrator is needed, though this *does not need* to be a causal one. The *capacity-responsibility* is, perhaps, a more known and acceptable notion of responsibility that we use (HART, 2008, 227-230). This is nearer of what Aristotle, Ayer, and Frankfurt comprehended about responsibility, that it involves the possession of psychological and behavioral capacities, such as those already mentioned above, as not being coerced, not having a compulsion, and being capable of forming the own intention in an appropriate way and being cognitively able to assess the context. When someone is in possession of these capacities, we can say that she is an agent; and whenever she acts, if she was not coerced or does not have any compulsion, she is responsible for her actions and their consequences. Since this kind of responsibility is attributed only *after* some act she did, some also called this as *retrospective* responsibility, since it looks for the *effects* of the action that has already happened (DUFF, 2007, 30; COWLEY, 2014, 2-5). This conception of responsibility is, indeed, relevant for our moral practices. We morally assess others through the actions they did and also their intentions in order to blame or praise them.

However, the capacity-responsibility is only capable of saying that the doer had the sufficient capacities to be a full-fledged agent. It cannot explain *why* these capacities are relevant nor it has a connection with the *criteria* of *when* a wrong is committed. In our moral practices we say that one committed a moral wrong when she went against what is accepted as morally correct in our moral community. But, any moral or legal code is extremely complex, and actions that are prohibited in some contexts, are not in others. For instance, as owners of a house, we can enter it whenever we want it, but others do not have this same privilege-right. They can enter in our house *only after* we give them this permission. In a similar way, no one can say which school someone else's children should attend, but their parents. These privilege-rights come accompanied by duties as well. In the same way someone cannot dictate anything to someone else's children, she has no special duty towards them, only their parents.

Our moral or legal code establishes different roles that are applicable in some contexts depending on the fulfillment of certain *role* conditions. As citizens of a country we have, for instance, the duty to pay taxes for the services provided by the state. If we do not pay them, we can suffer some penalty or punishment. But this duty is not applicable to tourists who travel through our country. They are not citizens of it, so they do not need to fulfill the civic duties of a country that is not theirs. This is the *role-responsibility* (HART, 2008, 212-214). The duties we have are given to us depending on

the roles we fulfill. They are defined *beforehand* according to our roles, that is, to the social *relationships* we bear in our moral community. In this sense, responsibilities are *relational* and *prospective* (DUFF, 2007, 23-36; COWLEY, 2014, 135-137). We can be blamed in case we fail to fulfill them. That is why we have *role-responsibility*, also called by some as *prospective responsibility*.

A very interesting analysis of the centrality of the role-responsibility for holding someone responsible was done by R. A. Duff (2007). He lays down his whole idea about responsibility as relational in the following paragraph:

It is a commonplace that responsibility involves a dyadic relationship: an agent is responsible for something. The *relational conception of responsibility* that concerns us here is not merely dyadic, but *triadic*: I am responsible for X, to S—to a person or body who has the standing to call me to answer for X. I am also responsible for X to S as Φ —in virtue of satisfying some normatively laden description that makes me responsible (prospectively and retrospectively) for X to S. *To be responsible is to be answerable*; answerability is answerability to a person or body who has the right or standing to call me to account; and *I am thus answerable in virtue of some normatively laden description, typically a description of a role, that I satisfy* (DUFF, 2007, 23, emphasis mine).

Along with Strawson, Duff claims that ascribing responsibility is a matter of, first, *identifying* which role the agent fulfills. We can say *which duties* someone has and their *conditions of fulfillment* only *after* we identify which roles she has. We cannot say *a priori* which kind of duty someone has toward another if we do not look at which role she has toward her. The role an agent can occupy differs depending on the kind of relationship she has with others (COWLEY, 2014, 137-148). Toward her children, she has the duties of a mother; toward her students, she has the duties of a professor; toward her parents, she has the duties of a daughter; toward every person, she has the duties as an agent; and so on. Each relationship demarcates a role she satisfies, and each of these roles has specific duties, and also rights, attached to it. Also, each role identifies *whom* she must *answer* to (DUFF, 2007, 25, 32).

However, these roles and relationships are not simply given, they are socially and culturally constructed. The role of mother, daughter, professor, and even of agent can vary from society to society. We cannot just say that the role *R* has the duties *d*, for this involves the values *v* of the context *C*

in which we find ourselves. That is why Duff insists on the notion that responsibilities are “normatively significant/laden descriptions” (DUFF, 2007, 24-25, 33-34). Even the notion of capacity-responsibility does not escape from this normative context. The differences about the minimum age of criminal responsibility in different nations show clearly that the criteria of when someone becomes a full-fledged agent can differ. Also, which capacities make someone beyond of controlling her own actions can diverge through cultures and societies. Without a role within society and the normative context that shapes this role, we cannot really know which responsibilities we have and, therefore, we cannot be held responsible for any action. That is the same as saying that capacity-responsibilities make sense only after role-responsibilities are well defined. About this point, Duff says:

Retrospective responsibilities of all these kinds depend, however, both on there being someone to whom I must answer, and who has the standing to call me to answer, and on the prospective responsibilities that I have to those who can thus call me to answer (DUFF, 2007, 30, emphasis mine).

Conclusion

Just as there are different meanings of “responsibility”, there are different ways of understanding what “determinism” means. Even though the theories espoused by compatibilists, such as Frankfurt and Ayer, could push aside the troubles created by *physical* determinism, they cannot escape the threat presented by Libet’s findings. Central to their conception of free will and moral responsibility was the notion of *voluntary* action, that is, of a *consciously chosen* action, even if this was implicit. When I consciously choose an action that means that I act through *my own will*; there is not external or internal motivation beyond those of my agency, and I *feel* the action as mine. Libet’s findings challenged this conception of consciously chosen action.

Based on earlier research about readiness potential (RP), a state of preparation for the voluntary action, whose beginning was, in average, – 550 msec before the action itself, Libet discovered that the awareness state of choosing the action comes, in average, 350 msec *after* the RP. When we say, “I will do *x*”, and feel that we are choosing how to act, the action had already been chosen. His rationale is that we can assume that the preparedness to act could only activate *after* the decision to act. But, since the state of consciously choosing the action comes, actually, only after the preparedness to act, then the decision to act is not a conscious one, but

unconscious. Libet does not try to explain the mechanism behind this unconscious decision; he only remarks that what we have been counting as voluntary action was not voluntary at all, for an action could only be called voluntary if it was consciously chosen. Even if he defends a *veto* conception of the voluntary action, this is not capable of rescuing the conception of moral responsibility as an *intentional* relationship between an event and our actions from his findings.

Nevertheless, this conception of moral responsibility leaves out many relevant aspects of our attribution of moral responsibility from our moral practices. More than connecting an event to a voluntary action of an agent, moral responsibility is a matter of attributing social *roles* to members in a community. There are many specific roles with specific duties that only people in some relationships have, but there is also a general role that every person in possession of some psychological capacities have; this is the role of *agent*. Being an agent means to be able to respond for one's own actions and, thus, to participate in the moral community. This aspect of moral responsibility was noted by Strawson. This is how we *do* attribute moral responsibility in our normal moral practices. Legal philosophers such as Hart and Duff also emphasized this aspect of moral responsibility and called it, not without a motive, *role-responsibility*. To be morally responsible is a matter of being in a social relationship and of being recognized as fulfilling a role in this relationship. Even if Libet's findings could say something about the neurophysiological inner workings of the voluntary action, they are limited and do not touch other important aspects of moral responsibility, as the notion of role-responsibility. Neurophysiological determinism could be a threat to our moral practices only if it had, from the beginning, set them aside. However, this begs the question.

Notes

¹ I would like to thank Marcelo de Araújo for his comments.

² Postdoctoral degree at the Programa de Pós-Graduação em Bioética, Ética Aplicada e Saúde Coletiva at the Universidade Federal do Rio de Janeiro (PPGBIOS/UFRJ) and PhD in Philosophy at the Johann Wolfgang Goethe Universität - Frankfurt am Main. ORCID-iD: <https://orcid.org/0000-0001-9724-8229>; e-mail: danieldevcosta@gmail.com

³ For a similar categorization see (ROSKIES, 2002).

⁴ Libet has a series of articles where he presents his findings, many collected in a book (LIBET, 1993a). However, for sake of brevity and our interests, we are just focusing on a few of them, in special those which offer a better philosophical analysis of the consequences of these findings to moral responsibility. It must be added that, although most of these articles and experiments had the contribution of different researchers, also for the sake of brevity, only Libet will be mentioned.

⁵ There are too many books and articles dedicated to his findings or that take them seriously. Some of them are (ROESSLER & EILAN, 2003; WEGNER, 2002; WALLER, 2011; LEVY, 2014; SINNOTT-ARMSTRONG & NADEL, 2011; GLANNON, 2015; HAGGARD & EITAM, 2015; BAUMEISTER, MELE & VOHS, 2010).

⁶ The vocabulary used here is meant to be as neutral as possible with respect to the question whether our actions should be described as being *caused* by us or as *intentionally connected* to us.

⁷ We set aside possible new questions created by the modern physics about indeterminism and probabilities, especially those from quantum mechanics.

⁸ This is the path taken by the *libertarianism*. Libertarianism is one of the possible stances from *incompatibilism*. The incompatibilism takes the possibility of free will and the thesis of determinism as contradictories; only one of them can be true. Libertarians defend that the determinism is false and the free will true, whereas the *hard determinists* assert that the determinism is true and free will impossible in such comprehension of reality.

⁹ Two of the most famous names that advocate for a scientific based philosophy are Patricia (1986; 2002; 2008; 2011) and Paul Churchland (1981; 1989; 2007). Among neuroscientists and neuropsychologists, the belief that how the mind works should be reduced to brain activity is the default thought. Aside from Libet, some of them known for their work about the psychophysiological basis of free will and voluntary action are Patrick Haggard (2005; 2011; 2019), John-Dylan Haynes (HAYNES, 2011; HAYNES, 2014; HAYNES & REES, 2006; SOON, BRASS, *et al.*, 2008), Daniel M. Wegner (2002), John A. Bargh (2005; 2008; BARGH, LEE-CHAI, *et al.*, 2001).

¹⁰ By potential it is meant electrical voltage.

¹¹ We should be aware, however, that, even though *every mental event* has an ERP, since this is a signal of a functional brain, *not necessarily every ERP* has a correlated mental event, since not every neural activity is correlated to a mental state or event (DONCHIN, 1979, 30-2).

¹² Fischer and Ravizza conceive guidance control as the capacity of acting in accordance with one's own will or because of reasons (FISCHER & RAVIZZA, 1998, 31-41). We can understand as *positive* guidance control as the capacity of willing the way of how the action will be, that is, the capacity of initiate an action in the way one wanted, whereas the *negative* guidance control is the capacity of the agent of *not* acting in a way she *want* it *not* to be, that is, she can *stop* the initiation of the action willingly.

¹³ One of such attempts can be found in (ROSKIES, 2011).

¹⁴ The importance of the reactive attitudes for Strawson, however, is more than just showing that we find ourselves in a moral community. They *found* which actions we consider *morally* correct or wrong on our *feelings* of gratitude, resentment, and indignation. We believe *x* is *morally* wrong for we have feelings of resentment or of indignation when we see *x* being committed against us or others. They are not, though, the foundation for the attribution of moral responsibility, since we attribute others moral responsibility because we consider them as *participants* of our moral community.

¹⁵ There are also, of course, cases where an agent can become less than it. Some cases of compulsion or mental disorder can put those agents outside of our moral community, at least, temporarily, for they are not able of being an agent anymore. When we see them in this light, it is normal for us to assume an objective attitude toward them. They cannot give reasons of why they act that way. They cannot respond anymore for their behavior and, therefore, they cannot be a participant of our moral community anymore, though they are still persons and can be objects of moral consideration.

References

- ARISTOTLE. *Nicomachean Ethics*. Chicago: The University of Chicago Press, 2011.
- AYER, A. J. Freedom and necessity. In: AYER, A. J. *Philosophical Essays*. London: Palgrave Macmillan, 1972. p. 271-284.
- BAARS, B.; GAGE, N. *Fundamentals of Cognitive Neuroscience: a beginner's guide*. Waltham: Academic Press, 2013.
- BARGH, J. A. Bypassing the will: toward demystifying the nonconscious control of social behavior. In: HASSIN, R. R.; ULEMAN, J. S.; BARGH., J. A. *The New Unconscious*. New York: Oxford University Press, 2005. p. 37-58.
- BARGH, J. A. Free will is un-natural. In: BAER, J.; KAUFMAN, J. C.; BAUMEISTER, R. F. *Are We Free?: psychology and free will*. New York: Oxford University Press, 2008. p. 128-154.
- BARGH, J. A. et al. The automated will: nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology*, v. 81, n. 6, p. 1014-1027, 2001.
- BAUMEISTER, R. F.; MELE, A. R.; VOHS, K. D. (Eds.). *Free Will and Consciousness: how might they work*. New York: Oxford University Press, 2010.
- BISHOP, R. C. Free will and the causal closure of physics. In: CHIAO, R. Y., et al. *Visions of Discovery: new light on physics, cosmology, and consciousness*. New York: Cambridge University Press, 2011. p. 601-611.
- BISHOP, R. C.; ATMANSPACHER, H. The causal closure of physics and free will. In: KANE, R. *The Oxford Handbook of Free Will*. 2^a. ed. New York: Oxford University Press, 2011. p. 101-111.
- BLACKWOOD, D. H. R.; MUIR, W. J. Cognitive brain potentials and their application. *The British Journal of Psychiatry*, v. 157, n. S9, p. 96-101, 1990.
- BUNGE, M. *Causality: the place of the causal principle in modern science*. Cambridge: Harvard University Press, 1959.
- CHURCHLAND, P. M. Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, v. 78, n. 2, p. 67-90, 1981.

CHURCHLAND, P. M. Folk psychology and the explanation of human behavior. *Philosophical Perspectives*, v. 3, n. Philosophy of Mind and Action Theory, p. 225-241, 1989.

CHURCHLAND, P. M. *Neurophilosophy at Work*. Cambridge: Cambridge University Press, 2007.

CHURCHLAND, P. S. *Neurophilosophy: toward a unified science of the mind-brain*. Cambridge: The MIT Press, 1986.

CHURCHLAND, P. S. *Brain-Wise: studies in neurophilosophy*. Cambridge: The MIT Press, 2002.

CHURCHLAND, P. S. The impact of neuroscience on philosophy. *Neuron*, v. 60, n. 3, p. 409-411, 2008.

CHURCHLAND, P. S. *Braintrust: what neuroscience tells us about morality*. Princeton: Princeton University Press, 2011.

COWLEY, C. *Moral Responsibility*. New York: Routledge, 2014.

CRAVER, C. F. *Explaining the Brain: mechanisms and the mosaic unity of neuroscience*. New York: Oxford University Press, 2007.

DAVIDSON, D. Actions, reasons, and causes. In: DAVIDSON, D. *Essays on Actions and Events*. New York: Oxford University Press, v. 2^a, 2001. p. 3-19.

DEECKE, L.; SCHEID, P.; KORNHUBER, H. H. Distribution of readiness potential, pre-motion positivity, and motor potential of the human cerebral cortex preceding voluntary finger movements. *Experimental Brain Research*, v. 7, n. 2, p. 158-168, 1969.

DONCHIN, E. Event-related brain potentials: a tool in the study of human information processing. In: BEGLEITER, H. *Evoked Brain Potentials and Behavior*. New York: Plenum Press, 1979. p. 13-88.

DONCHIN, E.; RITTER, W.; MCCALLUM, W. C. Cognitive psychophysiology: the endogenous components of the ERP. In: CALLAWAY, E.; TUETING, P.; KOSLOW, S. H. *Enoch Callaway - Event-Related Brain Potentials in Man*. New York: Academic Press, 1978. p. 356-411.

DUFF, R. A. *Answering for Crime: responsibility and liability in the criminal law*. Oxford: Hart Publishing, 2007.

FABIANI, M.; GRATTON, G.; FEDERMEIER, K. D. Event-related brain potentials: methods, theory, and applications. In: CACIOPPO, J. T.;

TASSINARY, L. G.; BERNTSON, G. *Handbook of Psychophysiology*. New York: Cambridge University Press, v. 3, 2007. p. 85-119.

FISCHER, J. M.; RAVIZZA, M. *Responsibility and Control: a theory of moral responsibility*. New York: Cambridge University Press, 1998.

FRANKFURT, H. G. Alternate possibilities and moral responsibility. In: FRANKFURT, H. G. *The Importance of What We Care About: philosophical essays*. New York: Cambridge University Press, 1998a. p. 1-10.

FRANKFURT, H. G. Freedom of the will and the concept of a person. In: FRANKFURT, H. G. *The Importance of What We Care About: philosophical essays*. New York: Cambridge University Press, 1998b. p. 11-25.

FRANKFURT, H. G. Identification and wholeheartedness. In: FRANKFURT, H. G. *The Importance of What We Care About: philosophical essays*. New York: Cambridge University Press, 1998c. p. 159-176.

GLANNON, W. (Ed.). *Free Will and the Brain: neuroscientific, philosophical, and legal perspectives*. Cambridge: Cambridge University Press, 2015.

HAGGARD, P. Conscious intention and motor cognition. *TRENDS in Cognitive Sciences*, v. 9, n. 6, p. 290-295, 2005.

HAGGARD, P. Does brain science change our view of free will? In: SWINBURNE, R. *Free Will and Modern Science*. Oxford: Oxford University Press, 2011. p. 7-24.

HAGGARD, P. The neurocognitive bases of human volition. *Annual Review of Psychology*, v. 70, p. 9-28, 2019.

HAGGARD, P.; EITAM, B. (Eds.). *The Sense of Agency*. New York: Oxford University Press, 2015.

HART, H. L. A. Postscript: responsibility and retribution. In: HART, H. L. A. *Punishment and Responsibility: essays in the philosophy of law*. 2^a. ed. Oxford: Oxford University Press, 2008. p. 210-237.

HAYNES, J.-D. Beyond Libet: long-term prediction of free choices from neuroimaging signals. In: SINNOTT-ARMSTRONG, W.; NADEL, L. *Conscious Will and Responsibility: a tribute to Benjamin Libet*. New York: Oxford University Press, 2011. p. 85-96.

HAYNES, J.-D. The neural code for intentions in the human brain: implications for neurotechnology and free will. In: SINNOTT-ARMSTRONG, W. *Moral Psychology: free will and moral responsibility*. Cambridge: The MIT Press, v. 4, 2014. p. 157-175.

HAYNES, J.-D.; REES, G. Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, v. 7, p. 523-534, 2006.

HUME, D. *A Treatise of Human Nature: a critical edition*. New York: Oxford University Press, v. 1, 2007.

IRWIN, T. H. Reason and responsibility in Aristotle. In: RORTY, A. O. *Essays on Aristotle's Ethics*. Berkeley: University of California Press, 1980. p. 117-155.

JAHANSHAH, M.; HALLETT, M. The Bereitschaftspotential: what does it measure and where does it come from? In: JAHANSHAH, M.; HALLETT, M. *The Bereitschaftspotential: movement-related cortical potentials*. New York: Springer Science+Business Media, 2003. p. 1-17.

KAPPENMAN, E. S.; LUCK, S. J. ERP components: the ups and downs of brainwave recordings. In: LUCK, S. J.; KAPPENMAN, E. S. *The Oxford Handbook of Event-Related Potential Components*. New York: Oxford University Press, 2012. p. 3-30.

KIM, J. The myth of nonreductive materialism. In: KIM, J. *Mind in a Physical World: an essay on the mind-body problem and mental causation*. Cambridge: Cambridge University Press, 1993. p. 265-284.

KIM, J. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press, 2005.

KORNHUBER, H. H.; DEECKE, L. Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Archiv - European Journal of Physiology*, v. 284, n. 1, p. 1-17, 1965.

LEEFMANN, J.; HILDT, E. Neuroethics and the neuroscientific turn. In: JOHNSON, L. S. M.; ROMMELFANGER, K. S. *The Routledge Handbook of Neuroethics*. New York: Routledge, 2018. p. 14-32.

LEVY, N. *Consciousness and Moral Responsibility*. New York: Oxford University Press, 2014.

LIBET, B. *Neurophysiology of Consciousness: selected papers and new essays by Benjamin Libet*. New York: Springer Science+ Business Media, 1993a.

LIBET, B. Unconscious cerebral initiative and the role of conscious will in voluntary action. In: LIBET, B. *Neurophysiology of Consciousness: selected papers and new essays by Benjamin Libet*. New York: Springer Science+Business Media, 1993b. p. 269-306.

LIBET, B. Do we have free will? In: SINNOTT-ARMSTRONG, W.; NADEL, L. *Conscious Will and Responsibility: a tribute to Benjamin Libet*. New York: Oxford University Press, 2011. p. 1-10.

LIBET, B. et al. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): the unconscious initiation of a freely voluntary act. In: LIBET, B. *Neurophysiology of Consciousness: selected papers and new essays of Benjamin Libet*. New York: Springer Science+Business Media, 1993.

LILIENFELD, S. O. et al. Neurohype: a field guide to exaggerated brain-based claims. In: JOHNSON, L. S. M.; ROMMELFANGER, K. S. *The Routledge Handbook of Neuroethics*. New York: Routledge, 2018. p. 241-264.

NAGEL, T. What is it like to be a bat? In: NAGEL, T. *Mortal Questions*. New York: Cambridge University Press, 1979. p. 165-180.

NICHOLS, S. After incompatibilism: a naturalistic defense of the reactive attitudes. *Philosophical Perspectives*, v. 21, n. Philosophy of mind, p. 405-428, 2007.

O'CONNOR, T.; FRANKLIN, C. Free Will. *The Stanford Encyclopedia of Philosophy*, 2020. Disponível em: <<https://plato.stanford.edu/archives/fall2020/entries/freewill>>. Acesso em: 06 dez. 2020.

RACINE, E.; DUBLJEVIĆ, V. Behavioral and brain-based research on free moral agency: threatening or empowering? In: ILLES, J. *Neuroethics: anticipating the future*. New York: Oxford University Press, 2017. p. 388-410.

ROESSLER, J.; EILAN, N. (Eds.). *Agency and Self-Awareness: issues in philosophy and psychology*. New York: Oxford University Press, 2003.

ROSKIES, A. Neuroethics for the new millenium. *Neuron*, v. 35, n. 1, p. 21-23, 2002.

ROSKIES, A. L. Why Libet's studies don't pose a threat. In: SINNOTT-ARMSTRONG, W.; NADEL, L. *Conscious Will and Responsibility: a tribute to Benjamin Libet*. New York: Oxford University Press, 2011. p. 11-22.

SHIBASAKI, H.; HALLETT, M. What is the Bereitschaftspotential? *Clinical Neurophysiology*, v. 117, n. 11, p. 2341-2356, 2006.

SINNOTT-ARMSTRONG, W.; NADEL, L. (Eds.). *Conscious Will and Responsibility: a tribute to Benjamin Libet*. New York: Oxford University Press, 2011.

SMULDERS, F. T. Y.; MILLER, J. O. The Lateralized Readiness Potential. In: KAPPENMAN, E. S.; LUCK, S. J. *The Oxford Handbook of Event-Related Potential Components*. New York: Oxford University Press, 2012. p. 209-229.

SOON, C. S. et al. Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, v. 11, p. 543-545, 2008.

STERNBERG, R. J.; STERNBERG, K. *Cognitive Psychology*. 6^a. ed. Belmont: Wadsworth, Cengage Learning, 2012.

STRAWSON, P. F. Freedom and Resentment. In: STRAWSON, P. F. *Freedom and Resentment and Other Essays*. London: Routledge, 2008. p. 1-28.

WALLER, B. N. *Against Moral Responsibility*. Cambridge: The MIT Press, 2011.

WEGNER, D. M. *The Illusion of Conscious Will*. Cambridge: The MIT Press, 2002.

Received/Recebido: 02/03/21
Approved/Aprovado: 02/04/21