Prolegomena 12 (1) 2013: 21–41

Moral Responsibility Beyond Classical Compatibilist and Incompatibilist Accounts

SOFIA BONICALZI
Dipartimento di Studi Umanistici, Sezione di Filosofia, Piazza Botta, 6,I – 27100 Pavia, Italia
sofia.bonicalzi@gmail.com

ORIGINAL SCIENTIFIC ARTICLE / RECEIVED: 28–12–12 ACCEPTED: 24–03–13

ABSTRACT: The concept of "moral responsibility" has almost always been defined in relation to a certain idea of metaphysical freedom and to a conception of the physical world. So, classically, for indeterminist thinkers, human beings are free and therefore responsible, if their choices are not defined by a previous state of the world but derive from an autonomous selection among a set of alternatives. Differently, for the majority of determinist philosophers (the so-called "soft compatibilists"), the only form of freedom we need has to be identified with freedom of the conduct, considered as opposite to any form of coercion. Some argue that, given the truth of determinism and the related suppression of concepts such as "guilt" and "praise", or "merit" and "demerit", morality could survive just as a utilitarian tool, even though this seems to be in conflict with our deepest feelings and practices. Considering some revisionist approaches of moral responsibility in connection with classical positions (synthetically presented in the first part of the paper), I will reconstruct some of the attempts to release responsibility from the thematisation of freedom, exploring the possibility of redefining it as an independent concept. My conclusion is that the focus on the choice-action process and on the characteristics of the "self", avoiding reference to alternative scenarios, could be a good starting point for elaborating a conception of what really counts for our moral life – even though, in the end, this could entail the abandonment of the traditional concept of responsibility itself.

KEY WORDS: Compatibilism, free will, incompatibilism, moral responsibility.

1 The Historical Background

Is there a way to characterise the concept of moral responsibility apart from issues regarding the problem of free will, one that would develop an account able to survive the changes of perspective about the status of freedom? In the present paper I want to show how the recent attempts to release the concept of moral responsibility from the thematisation of metaphysical freedom would be obliged to deal with thorny issues, even though they permit to overcome some of the problems found in traditional views. It was with the publication of “Freedom and Resentment” (P. F. Strawson 1962) and then of “Alternate Possibilities and Moral Responsibility” (Frankfurt 1969) that the reconfiguration of the debate around the concept of “moral responsibility” was made possible. In that horizon, moral responsibility – in an almost unprecedented way in the twentieth-century debate – was considered regardless of the choice between determinism and indeterminism, although within a perspective that can still be defined as a form of compatibilism.

If there are those who, like Robert Nozick in *Philosophical Explanation*s (Nozick 1981: 291), consider the issue of responsibility as a secondary one to our most profound philosophical questions, the problem of proper attribution of responsibility has always represented a challenge for the theoretical options about free will. Historically, the interaction between moral responsibility and metaphysical freedom has always been present both in compatibilist and in incompatibilist thought (see Kane 2002, 2011; De Caro 2004). In its various forms, incompatibilism proposed itself as the ideal interpreter of the pre-philosophical intuition according to which we are free if we have a set of alternatives to choose from. The main idea is that only in this case our will is not subjected to the determination of internal or external causal factors. On this premise, two scenarios seem possible: in a universe where there is room for indeterministic factors or for a form of causation irreducible to the mechanical-physical one, human beings enjoy free will; in a universe where nothing escapes the determinations of physical and mechanical properties, individuals simply could not be considered free. In this sense, the possibility of considering the agent as responsible for her actions directly depends on the availability or unavailability of a space of autonomy, where she can make a choice between genuine alternatives. So, for an indeterminist like William James (1956), the future is open to individuals who choose autonomously whether to walk along Oxford or Divinity Avenue, giving their consent, from time to time, to a different option, which is not mechanically connected to the previous state of affairs. In contrast, for a *hard determinist* as Galen Strawson, the individual subjected to causal law is unable to make an independent choice among alternatives. This is the basis of the incoherence of the idea of responsibility, for which “It is exactly as just to punish or reward people for their actions as it is to punish or reward them for the (natural) color of their hair or the (natural) shape of their faces” (G. Strawson 2002: 458). For indeterminist incompatibilists, the correct attribution of responsibility stems directly from the individual’s ability to exercise control over available alternatives. For some determinist incompatibilists, traditional concepts such as those of "guilt", "punishment" and "obligation" lose all meaning and should be abandoned. This renounce often leads to accept a utilitarian concept of moral responsibility, according to which responsibility attributions can be justified by the context, but they are not bound to an obscure idea of "merit" or "demerit". The need of order and social peace is considered as sufficient for maintaining these notions, without further considerations on the agent's merit/demerit. A partially similar idea, which immediately connects responsibility to the issues of punishment and reward, lies at the basis of most compatibilist positions. For traditional compatibilism, determinism is compatible with freedom, provided that it is not interpreted as freedom of the will, but as freedom of the conduct ("The agent could do otherwise, but only *if one had chosen to do otherwise*" (Moore 1912: 12)): in a world where causal law – which nobody could escape without a complex metaphysical loophole – obtains, agents can be considered free, no matter the factors that have influenced their behaviour. The only requirement is that these factors can not have a coercive nature.

In Moritz Schlick's classical view, only freedom of the conduct should be attributed to mankind. This is the only kind of freedom that is necessary to consider individuals as being morally responsible: I consider myself responsible because my desires correspond to the motivations that have caused my action (Schlick 1930). What about those who commit murder under the effect of a drug? Usually we judge them responsible only if they assumed drugs spontaneously, without any external intervention, while we are inclined to think that the actions of those who are affected by a mental disease (a sort of internal constraint) may be considered innocent. From a legal standpoint, the question of responsibility is directly linked to the problem of justifying punishment: it is in view of the future punishment that it is necessary to understand present or past responsibility.

Punishment is intended as an educational tool which should guide and train the motivational basis of behaviour, preventing the repetition of an act (by the same agent) and encouraging different determinations of the conduct (in the others). An expression such as "I could not have acted otherwise", therefore, has no other meaning except that, given other reasons, and compatibly with the laws of volition, I could have had a different behaviour. The aim of compatibilists such as Schlick and Hobart is to overcome their opponents, demonstrating not only that determinism is compatible with responsibility, but also that responsibility could not be established otherwise. In particular, it could not be sustained if we considered the universe as indeterministic. An indeterministic universe would lead to pure randomness and hence to the total lack of responsibility (Schlick 1930; Hobart 1934). From the practical point of view, also the connection between responsibility and punishment would fall: rewards and general measures, adopted to direct the individual conduct, would be meaningless if the agent's decisions lacked their connection with a cause on which we can operate from the outside. A refined version of this objection is represented by the so-called *Rollback Argument*, developed by van Inwagen as a counterpart of the more famous *Consequence Argument.*[1] On the basis of the two arguments, responsibility is compatible neither with indeterminism nor with determinism. In the first case, the agent would not have any kind of control over her actions, while in the latter the absence of a choice between alternatives would reduce any assignment of responsibility to zero (van Inwagen 1983). More recently, van Inwagen and others have taken a sceptical position on the possibility of defining a persuasive solution to the problem (van Inwagen 2000; McGinn 1999, 2002). During the last decades the debate on free will and moral responsibility has been strongly affected by the results of scientific analyses on brain mechanisms,[2] whose functioning seems to erode any space left to free will, given that also conscious decisions are seemingly preceded by unconscious processes, thereby supporting philosophical theories which tend towards mechanicist solutions. If our choices are the results of mechanisms we are not able to control, how we can define ourselves as free agents and what are the consequences on our responsibility theories? Even though the present state of research does not allow to glean definitive solutions, it would be naive to ignore neuro-scientific contributions to define the constitution of free will and to formulate the concept of "agency". Nevertheless, even if scientific progress were able to demonstrate the falsity of our idea of metaphysical freedom, no conclusions would be immediately available from an ethical point of view and the question about how to explain moral responsibility, in a way compatible with the scientific vision of the world, would still be open.

2 Moral Responsibility: A Revision of the Problem

It is on the basis of the unsatisfactory answers of traditional approaches and of the debate – promoted by Austin and Hart – on the conditions of excuses, that we can understand the attempt to re-establish the concept of moral responsibility without any references, at least in appearance, to

[1] Firstly presented by Ginet (1966).

[2] Consider Benjamin Libet's pioneering experiments (Libet 2002), and their subsequent versions, e.g. Soon, Brass, Heinze, Haynes (2008).

the theoretical choice between determinism and libertarianism (Austin 1956–1957; Hart 1968). Given a sympathetic, but apparently not binding, adhesion to compatibilism, this shift of perspective has allowed the formulation of moral responsibility theories that seems more responsive to our pre-philosophical intuitions than those reconcilable with classical compatibilism. This also favoured a possible rapprochement between the need for an impartial morality and a fair view of individual reasons. In some cases it also allowed an internalist interpretation of responsibility, based on the relationships between members of the moral community (Wallace 1994). The common foundation of these approaches is the attempt to offer a conception of moral responsibility able to survive the hypothetical discovery of the truth about determinism.

However, it is doubtful if these proposals, satisfactory from a regulatory standpoint, are capable of a foundation that would really avoid the preventive choice between incompatibilism and compatibilism. The forms of revisionist compatibilism – which have their roots in Strawson's and Frankfurt's contributions and which assume the separation of the question of responsibility from the thematisation of freedom – can hardly match the intuitions underlying the pre-theoretical attraction aroused by incompatibilism. At the same time they sometimes rely on strong forms of "control" that are not always innocent from a metaphysical point of view.

As is well-known, in "Freedom and Resentment", P. Strawson effectively summarises the positions generally taken about the relationship between determinism and moral responsibility, with the aim of formulating a theoretical proposal that would release the allocation of rewards and punishments from the adoption of a utilitarian morality (considered as a poor response to the pre-theoretical insights on responsibility). With respect to concepts such as "responsibility", "guilt" or "praise", the central issue would not concern the truth of the favoured theoretical option (determinism or indeterminism), but the reactive relationship established among rational agents: someone's actions provoke (or could provoke) a non-neutral attitude by someone else. The mutual and implicit recognition of the validity of these responses is diminished when we catalogue the action as not intentional or consider the agent as not responsible because she (temporarily or permanently) lacks some mental faculties.

For Strawson, the difficulty in taking an objective attitude towards individuals who "deserve" reward or punishment does not represent the theoretical proof of the falsity of determinism, but rather reveals the practical impossibility of leaving the established pattern of our ordinary conduct. The legitimacy of the deterministic thesis, for Strawson, has little or nothing to do with the only form of rationality in question, which can be linked to the acceptance or the rejection of such practices. The net of reactive feelings, which opposes the establishment of an objective attitude, would not even need an external rational explanation, like the utilitarian one, but would be formed essentially by itself, thanks to its bound with the growth of the moral community.

In the following decades, Strawson's proposal has been subjected to a number of objections. The most obvious and penetrating one refers to the idea that if determinism proves true, it would seem rational to abandon this type of behaviour, whose validity would be compromised (G. Strawson 1986). Is it really possible to maintain these practices because of their social function or their relationship with our daily life, even knowing that they are simply illusory, or instead the need for truth cannot be suppressed? Even if we could not take truth as a parameter for evaluating reactive attitudes, it seems equally difficult to understand how it would be possible to overcome those feelings that, from an external view, would seem inappropriate. In this sense, how can we escape their mere acceptation and avoid subjecting them to criticism if they appear to be morally questionable? We can easily imagine, as Fischer and Ravizza (1993) did, a community in which some individuals, who suffered from a mental disease, are constantly blamed and punished, as guilty of provoking resentment (the reactive feeling) in the other members of the community. The suspect is that (once eliminated the link between the attribution of responsibility, arising from a judgment, and the moral sentiment) the mere consideration of excusing conditions could not be sufficient to isolate the appropriate reactive attitudes from the inappropriate ones. How can we combine the normative element, implied in the assignment of responsibility, with the need to immediately match our practices with the presence of reactive attitudes, not supported by a previous judgment? How does the presence of reactive attitudes characterise human relationships, building an intersubjective concept of moral responsibility based on mutual recognition of specific abilities? These seem to me like the most relevant obscurities of Strawson's account, whose most interesting specificity resides in the fact that reactive sentiments would not need the support of a previous theoretical judgment. Moral feelings emerge from themselves, in a completely independent way. For Strawson, the basis of this type of behaviour seems to lie precisely in its non-rational character: it is from our practices that we can understand the meaning of the concept of responsibility.

Paradoxically, a critical attitude towards the claim to keep reactive attitudes unchanged, even in front of the discovery of the validity of determinism, characterises both libertarian thinkers (who see determinism as a threat for the existence of the moral community) and some (*hard*) determinists, such as Ted Honderich (1993), who deny the possibility to maintain the same attitude towards existence, once we accepted the truth of determinism. Moreover, the Strawsonian position is also subjected to another kind of objection, because it does not seem able to provide a satisfactory definition of responsibility in an absolute sense, independently from the perception of the agent.

From another point of view, also H. Frankfurt questions the connection between freedom and responsibility in "Alternate Possibilities and Moral Responsibility", where he tries to refute the *Principle of Alternate Possibilities* (PAP): that is one of the basic assumption of incompatibilist freedom, according to which "X is morally responsible only if X was able to do otherwise because he was able to choose otherwise". The principle would be false because there are situations in which, although it is impossible to act otherwise, the agent can still be held responsible. The credibility of the principle consists in its undue overlap with the idea that coercion and moral responsibility are not compatible. However, in the absence of coercion, the only circumstance under which it is impossible for the agent to do otherwise might not be sufficient by itself to eliminate moral responsibility.

Frankfurt's procedure is not aimed at illuminating the bond between freedom and alternative possibilities, but rather at defining what elements are necessary and sufficient to define the agent as morally responsible. He develops his argument through a series of counterexamples, variously accepted or refuted by his commentators. The basic form of these cases is quite known: consider an agent (Jones) who, for personal reasons, decides to do a certain action. Then he receives a terrible threat that requires him to do exactly the same act. If Jones chooses to perform the action in question, we would tend to think of him as morally responsible, although it is not true that he could act otherwise. In such circumstances, the problem of responsibility attribution does not seem to depend on the possibility of making alternative choices, but on the type of the relationship that exists between the agent's original decision and the suffered imposition.

This is evident in a famous manipulative scenario: Jones is remotely controlled by Dr. Black, through a system that allows him not to reveal his power over Jones's conduct. Dr. Black wants Jones to perform the action A, but will intervene only if his victim tries to do B. If Jones chooses autonomously to do A, we would tend to think that the whole moral responsibility falls on him, regardless of his inability to make alternative choices. Once Jones has done this, both the possible explanations of his conduct (he acted spontaneously or as a result of Black's imposition) do not depend on the presence of alternatives in order to determine whether Jones is morally responsible or not. Hence Frankfurt concludes that the *Principle of alternate possibilities* is wrong, because there are conceivable circumstances in which the impossibility to do otherwise plays no role in the determination of moral responsibility.

However, the tendency to moderate criticism against those who justify themselves, saying that they could not act otherwise, belongs to the sphere of ordinary reactive attitudes and to the supposed charm of incompatibilism. In Frankfurt's view, that justification would be acceptable if the agent had performed an action in the presence of a form of coercion, only if this plays an effective role in the succession of events. Regardless of the fact that desires are themselves caused by something, if the action results from the agent's will, it is possible to consider the subject as a responsible agent. This form of responsibility does not need to consider the presence of alternative possibilities and so it is compatible with determinism. The shift from the *Principle of alternate possibilities* to the focus on the characteristics of the choice-action process, which separates the concept of responsibility from the idea of metaphysical freedom, would not satisfy the libertarians who believe that freedom of will (and not only of conduct) is a necessary condition for moral responsibility. It is not by chance that libertarianism prefers other paths, continuing to connect responsibility with metaphysical freedom, as in the case of Robert Kane's *Ultimate Control*, which involves the insertion of an indeterministic element between choice and its causes (Kane 1985, 1996). The cases presented by Frankfurt are suitable to give an account of the concept of moral responsibility only in relation to theories that have already embraced the compatibility between determinism and freedom, but they seem inadequate to answer the needs of a libertarian conception: if we consider Jones responsible only when he acts without the intervention of Black, we implicitly assume as valid the idea of responsibility that derives from compatibilist conceptions of freedom (understood as the mere absence of obstacles). Since this is not the commonly accepted definition of freedom, a libertarian could bypass Frankfurt's argument, simply affirming that Jones' responsibility derives from the fact that he could have, at some point (even before the actual choice), taken a different route (metaphysical freedom is the theoretical basis of moral responsibility), because the course of events that led him in front of a certain choice is not (completely) causally determined.

3 New Paradigms for Moral Responsibility

Thanks to this new horizon of meaning, however, future compatibilists will be allowed to overcome the rigid conceptual grid imposed by the conditional interpretation of freedom formulated by George E. Moore at the beginning of the nineteenth century (see Fischer 1999; McKenna and Russell 2008). The elements characterising this new paradigm are also linked to the subsequent essay by H. Frankfurt's, "Freedom of the Will and the Concept of a Person" (1971). In the attempt at preserving the value of choice in a deterministic context, the article proposes the possibility of a distinction between freedom of will and freedom of conduct, independently from the preferred metaphysical substrate. The deepest question relating to the debate on free will (what the value of the individual choice is in a world where universal causation obtains), would not find its answer through a special metaphysical theory, but through the analysis of the specific structure of human will.

For Frankfurt, the peculiar aspect of human beings, who can enjoy freedom and can be considered as responsible agents, is the ability to form *second-order volitions*, which require the presence of a developed rationality and reflect the agent's concerns about the orientation of her own will. Freedom of the will can be represented as the agent's freedom of "wanting what she wants", the ability to coordinate first order desires in relation to higher order volitions, regardless of whether they are determined or not. We could not consider free the individual who was neutral in front of the conflicts of his own will, like in the case of the drug addict who takes drugs without worrying about the fact that the impulses driving his will are the ones he would like to have. Similarly, what is important so as to define responsibility is exclusively that the agent has been moved by her own will, regardless of the presence of alternative possibilities.

The reversal of perspective produced by Strawson's and Frankfurt's essays leads to a shift of attention from the concept of absolute freedom to issues related to interpersonal relationships structures and to human attitudes. The objections formulated by Gary Watson – underlining the presumed arbitrariness of the concept of *second-order volition* and proposing instead a distinction between faculty of desiring and faculty of reasoning (as separate and different motivational sources) – are also built on top of Frankfurt's theoretical framework. Also in this case, the concepts of "choice" and "responsibility" are defined on the basis of a reflection on the constitution of the subject-agent (Watson 1975). Responsibility is not defined in relation to an individual act of choice, but to a historical scenario, which allows to associate moral assessments of merit/demerit (unfamiliar for classical compatibilism) to the conduct.

According to the new paradigms opened by Strawson and Frankfurt, various approaches (built on the ideas of "control" and of "choice as something depending on the subject") can be enumerated (even though, quite often, their outcomes are not completely alternative). My purpose is not to consider all the possible perspectives but, more modestly, to briefly analyse the metaphysical approach provided by J. M. Fischer and the normative proposal developed by T. Scanlon (in comparison with the internalist account provided by R. J. Wallace) The first approach treats responsibility as a historical phenomenon (responsibility attribution depends on the structure of the choice-action process); the second is more concentrated on the characteristics of the moral agent in an interpersonal context. Both these views try to make some concessions to the first opponent (classic libertarian incompatibilism), proposing a conception of responsibility which is built in a compatibilist scenario, but claim to be independent from the question of the validity of determinism, pretending to be able to resist any discovery about determinism's truth/falsity. The concessions towards incompatibilism often go in the direction of the attribution of a merit/demerit to the agent. This is an element which hardly finds some space in consequentialist (compatibilist) ethical theories representing, on the contrary, one of the cornerstones of the libertarian thought.

Among the supporters of the first approach, which proposes a new metaphysics of responsibility, John Martin Fischer (1995, 2011), through the formulation of the concept of *guidance control,* suggests an intriguing externalist account (*historicism*) of moral responsibility. The concept of *guidance control* represents the basis of responsibility attribution together with a form of reason-responsiveness (the mechanism of choice is responsive to reasons). Responsibility requires both the ability to respond, through personal behaviour, to reasons (among which we find the moral reasons), and the recognition of the agent as the holder of the mechanism of choice from which the action flows. Both requirements are related to the actual action sequence, avoiding the reference to alternative scenarios but, differently from Frankfurt's focus on second-order volitions, here responsibility is not a time-slice notion depending on the harmony between mental faculties (Fischer 2000: 388). The version proposed by Fischer is a form of merit-based view which associates the concept of moral responsibility to *guidance control*, the type of control that is independent from the availability of alternative courses of action and that involves the agent's ability to "do what she is doing". The possibility of making alternative choices is a different type of agency power or a different kind of control over actions (*regulative control*), not decisive for the definition of responsibility. Imagine, according to a variation of the Frankfurtian counter-examples, to be driving a car. We would consider two possible circumstances: in the first scenario, we decide to turn right and we complete the designed turn. Similarly, if we had the intention to turn left, we could do it, turning the steering wheel in the opposite direction. We therefore possess two different types of control over our vehicle, which Fischer identifies as *guidance control* and *regulative control*. The first term means that we are able to do what we currently do while, with the other, we describe our power to act in a different way. In the second scenario, while we are driving, we do not realize that something in the guidance system breaks in a particular way. If we turned right all would take place in the normal manner but, if we decided to proceed to the left, the car would move to the right.

In this case we would lose the second kind of control (*regulative control*), but not the first (*guidance control*), the one that – in agreement with Frankfurt's results, but without an inconvenient reference to mysterious external agents (and so avoiding perplexities about the location of the mechanism of control) – is useful for the determination of moral responsibility. How does Fischer intend to release freedom from the validity of PAP? His starting point is the idea that, even in a deterministic context, the individual can be identified as the mechanism from which the action flows. He may be held responsible if he is not subjected to external mandatory pressures. However, even in this case, we could speak of freedom and responsibility, in a Hobbesian and a Lockean sense, only from a compatibilist perspective. These concepts, in fact, are based on the distinction between actions imposed by coercive mechanisms and actions dictated by a will whose impulses are not under the control of the agent. The idea of "being the owner of the mechanism" has a different meaning depending on whether you prefer a compatibilist or an incompatibilist conception: absence of coercion in the first case, freedom of choice in the second. Moreover, as Fischer himself suggests, demonstrating the falsity of PAP is not enough: at the real core of incompatibilist thought there is not the possibility to do otherwise, but the idea that choices are not determined by an internal or external source that is beyond the agent's control. Even though in a particular circumstance we could not do otherwise, we would be free – in an incompatibilist sense – if our action was the result of an autonomous process of choice. Nonetheless, an idea of (at least) "moderate control" (being able to respond and interact with reasons, without claiming an absolute control over the process of choice), together with the focus on actual-sequence scenario, seems to represent a promising ingredient for a tenable conception of what moral responsibility could be.

The theoretical proposals of the second line, within which I will consider the positions of Thomas M. Scanlon (1986, 1998), are partially different. Here, the attribution of responsibility derives from an evaluative judgment of the agent's conduct in an interpersonal context. This proposal keeps a Strawsonian inspiration, but dismisses his rigid anticognitivist perspective. Scanlon, considering some Strawsonian issues in the light of Hart's view, builds a theory of moral responsibility that distinguishes *responsibility as attributability* (the agent is responsible if it is reasonable to consider her action as a basis for a moral evaluation; the question concerns the rightness of blaming or praising the agent for a certain action) from *substantive responsibility* (determined by obligations imposed by voluntary choices and by the position assumed in the social context) (see also Watson 1996). Responsibility and moral evaluation are defined from a theoretical perspective of contractarian type, which is able both to offer substantiality to the Strawsonian intersubjective requirement and to untie moral judgment from the presence of reactive attitudes (Scanlon 1998: 248)

In Scanlon's view, morality is understood as a system of co-deliberation. Moral reasoning would correspond to an attempt at developing common standards which could represent the basis for further deliberations and for the formulation of criticism. In an intersubjective system, there is a sense in which it is reasonable to ask people to justify their actions and by which it is permissible to make moral judgment, taking into account the appropriate excusing conditions. The two forms of responsibility belong to different moral sources, but both would not be endangered by the eventual truth of determinism. Nonetheless, if this is certainly true conceiving responsibility just as a social phenomenon (which derives from an intersubjective agreement and from the necessity to accept a certain burden as a consequence of a wrong action) difficulties can arise from the attempt at preserving a more pregnant concept of moral responsibility. For Scanlon, moral error and responsibility arise when, despite the awareness of the existence of reasons, we act in a way that is different from that commonly established or we fail to consider its consequences. Scanlon formulates the proposal of a moral contractualism, in which common principles are applied impartially to anyone who covers a certain position and share certain contextual features. Moral fault coincides with the failure of the possibility of justifying an action according to principles that the subject is unable to reject.

Responsibility founds its basis on the analysis of human specificity, which is built not on the Frankfurtian concept of "volition"/ "desire", but on the concept of "reason" (Scanlon 1998: 18). We are responsible because we are capable of *judgment-sensitive attitudes*: beliefs, intentions and judgments we can be asked to justify, regardless of whether they are linked to desiderative feelings. The choice of principles does not require an external point of view, but it is located within the given context, leading to a normative theory that preserves the deontological concept of merit with an idea of responsibility compatible with the truth of the causal thesis. Maintaining something similar to the classical idea of merit, together with the idea that the agent can modify her attitudes, seems to represent the only weapon against a deflationary conception of moral responsibility, which seems to reduce the assignment of responsibility to non moral evaluation (Smart 1961).

A difficulty can be related to the attempt at distinguishing what specifically characterises moral blame from what can be defined just as a non-compliance to reasons. Not all the standards we reject (and for which we consider ourselves as obliged to justify ourselves in consideration of the setting-up of an harmonious community) are subjected to moral blame. The shift from what is "reasonable" to what is "moral", from rational criticism to moral blame, does not seem immediate if we want to preserve the idea of "merit". Where is the shift from the inability to recognise and respond properly to reasons to the attribution of a moral fallacy to the agent located? In this sense, can the compatibilist distinction between free actions and actions produced by a coercive mechanism be considered as sufficient or would we require a substantial analysis of the content of moral judgments? The problem is discussed by Scanlon, who presents moral wrong as different from any other kind of violation, as a special case of a more general rational criticism: in this case what is violated by wrongdoers is not simply a generic value adopted by someone, but the special value of people "as rational creatures" (Scanlon 1998: 272).

In Scanlon's view determinism does not represent a menace for *attributability* and for substantive responsibility, because both are defined in relation to the values of our choices, whose importance is not undermined by the truth of the *causal thesis*. If we knew that the outcomes of our conduct are defined by process beyond us, we would still have reasons "for preferring principles that make what happens to us depend on the ways we respond when presented with alternatives" (Scanlon 1998: 255). The general thesis consists in the idea that we do not need to appeal to the voluntariness of our choice in order to explain its significance for our moral life. The conditions under which the choice is taken are the element according to which it is possible to reach a conclusion about responsibility attribution. The difference between a man who makes an unfavourable deal because he ignores the presence of a good alternative, and a man who makes an unfavourable deal because this alternative is precluded to him does not rest in the fact that the former made the choice willingly, but in the conditions that characterise the situation of choice.

Also in this case, the attention is shifted from the presence of alternative possibilities to the actual scenario, in order to define the nature of a particular choice in relation to its causal history (the background conditions). The suggestion consists of the idea that while the *value of choice* account is not undermined by the truth of determinism, an approach that tries instead to defend the willingness principle is inevitably menaced by the lack of a strong idea of freedom of will. In my view, even though it is true, as Scanlon suggests, that choices maintain their value independently from the possibility to do otherwise, it would be more difficult to defend a substantial idea of deserved merit/blame, given the absence of the form of freedom defended by indeterminists. The focus is not only on the future action, on the educational property of reproaches and warnings, but also on the moral evaluation of an action, even though moral sanction is specifically directed at modifying *judgment-sensitive attitudes*.

As mentioned before, the classical problem regarding the possibility of evaluating an agent whose conduct is determined by something beyond her control is solved by Scanlon with reference to the concept of "reason". When the agent perceives an inner struggle between opposite tendencies, she has still to solve a dilemma about which the reason to follow would be. Also in this case we do not have to separate (moral) reasons from "desires" or "impulses" that could not be referred to the agent: if there are no reasons to suspend the judgment, she can be asked to justify her conduct and moral assessment. The difference is between this thesis (the idea that people can be the target of moral assessment because they are rational beings) and what Scanlon refutes and calls the *desert thesis* (the retributivist idea according to which wrongdoers should suffer for what they have done), incompatible with the truth of determinism. Scanlon's effort consists of the attempt at distinguishing his thesis from the *desert* one, which seems morally questionable if we accept the truth of determinism. Even if determinism were true, we could still maintain a notion of weak self governance (together with the idea that causal control is not sufficient for moral responsibility in the sense required for moral blame: we do not consider mechanical tools as morally responsible, even though their inner mechanism causes a certain outcome), according to which people are responsible because they are able to exercise a form of control over their behaviour. Also in this case, as in Fischer's account, it is the idea of weak control that seems less questionable, and a consideration like this seems to be at the basis of every account of moral responsibility pretending to be compatible with the truth of determinism or of the *causal thesis*. Scanlon briefly discusses cases such as those presented by Frankfurt or Fischer (e.g. actions caused by a hypnotist who produced a determined reaction through a mechanism) and concludes that the element which entails a lack of responsibility could not be traced simply in the presence of causal factors, but in the interruption of the physiological chain between *judgment-sensitive attitudes* and the outcomes of our conduct. In general, the simple presence of causal factors does not represent a menace for responsibility attribution. Nevertheless, if we do not have the kind of freedom portrayed by incompatibilism, how is it possible to blame people who do not modify their own *judgment-sensitive attitudes* in a way compatible with what they owe to one another? The problem is not only that the agents could not be able to modify their habits, but also that blame can be considered as an inappropriate reaction towards people who are causally determined. If it is true to say that wrongdoers could not complain about a burden that they have to accept as a consequence of their conduct, it is not so clear that the simple ability to understand reasons could be considered as sufficient for moral assessment.

Within a similar context, some intriguing suggestions come also from the internalist account proposed by R. J. Wallace, who does not focus his attention on the characteristics of the choice-action process or on the natural properties of the agent's practices, giving instead a specific importance to her normative competence. Wallace restricts the list of reactive attitudes, proposing a smaller catalogue, which includes only resentment, indignation and guilt that accompany the actions we consider morally wrong (no particular feeling is aroused by the simple observation that the agent is doing his duty). The attribution of responsibility can arise if we are able to consider the subject as a potential target of a specific moral evaluation. This is based, first, on the Frankfurtian idea of the agent as a self-reflective and autonomous "self", capable of building a network of "commitments sufficiently structured to constitute what we might call a 'conception of the good'" (Wallace 1994: 53). The attribution of responsibility is then linked to the formation of expectations, supported by reasons that we are ready to accept as a basis for deliberations, criticism and normative discussions. Considering the agent as responsible means nothing more than feeling her as tied to obligations. In this sense, Strawsonian reactive attitudes play a secondary role, as they are subsequent to responsibility judgment. These are the feelings that we would judge legitimate towards an individual who we consider *morally accountable* (and who accepts these moral obligations), in the case in which moral expectations have been disregarded. Moral blame is a form of moral judgment that goes beyond the mere description of the subject's actions. It derives its strength from the attitude that she expresses or can express. In this case, how can we escape the risk of arbitrariness – similar to the one that hung over Strawsonian reactive attitudes – linked to the relationship between judgments and moral sentiments? In Wallace's view, the problem can be solved if we consider moral sentiments not as pure expression of the subjectivity of the judge, but as attitudes linked to the moral obligations we accept.

Nevertheless, the final question about the necessity of sustaining a conception of moral responsibility that would try to preserve our ordinary intuitions about merit and demerit attribution (which seems something more than a result of the mere description of the agent's properties) reminds of a Spinozian suggestion, according to which the feeling of freedom implicit in our practices is the result of our ignorance about causes:

how much credibility the form of our ordinary responsibility attribution should receive?

4 Conclusions

The attempts at overcoming the consideration about freedom in order to establish a new concept of moral responsibility are not uncontroversial. The deepest query on these revisionist approaches, based on the attribution – in a Davidsonian sense (Davidson 1973) – of a degree of rationality to human action, are related once again to the possibility of finding a validity outside a preventive adhesion to compatibilism: the only option that allows to treat responsibility separately from freedom. The Strawsonian prospect – especially if deprived of the intersubjective and social foundation provided by some normative approaches – seems to primarily involve an unresolved tension between the need for truth and the adherence to reactive attitudes. Even if these did not have a rational justification, depending on natural human dispositions, and if the *causal thesis* were true, it may be legitimate to abandon such attitudes (in particular those connected with the attribution of a form of "merit"/"demerit" to the agent). They would be somehow irrational, as suggested by the supporters of *hard determinism*, or feared by those who find the requirement of Kantian autonomy necessary for holding people responsible but doubt that human beings possess this characteristic.

Nevertheless, even though maintaining a form of moral appraisal is extremely problematic given the truth of determinism, we do not seem able to accept neither an utilitarian nor a completely deflationary conception of responsibility. At the same time, it is unlikely that human beings may be at ease if they think that the whole of their practices are simply not well founded. Probably what people want to preserve from deconstruction is not the idea of responsibility itself, but the general value of the morality system and interpersonal exchanges. In order to solve the dilemma, we could substitute the attempt at saving merit and demerit attribution with something really able to survive determinism's truth, even though this implies the abandonment of the traditional concept of responsibility.

An interesting option is provided by some hard determinist positions that (eliminating also responsibility attribution and moral appraisal) try to save and explain the feelings that structure our existence.[3] In particular, Derk Pereboom, like Fischer or Scanlon, maintains an actual-sequence approach, focusing his attention on the action process and not on the presence

[3] See Pereboom (1995), (2001: 112–126); Smilansky (2000); G. Strawson (1994), (1986), (2002).

of alternative possibilities. Without embracing a consequentialist account or assuming a merit-based view, he defines traditional moral responsibility as a form of illusion (Pereboom 2001): if responsibility, as it is commonly intended, requires a complete control over the factors that determine our choices, human beings are not responsible in a deterministic universe (and not even in an indeterministic one), no matter if their choices are not produced in a manipulative scenario. Pereboom presents his critics to Frankfurt's conclusion with his famous *Four-Case Argument for Incompatibilism*,[4] which employs a generalisation strategy in order to show that no real differences exist between a manipulative scenario (where Plum, an individual created by a neuroscientist, commits a murder under the control of his creator, who can manipulate Plum's reasoning process locally – as in case 1 – or in a remote past – as in case 2), a situation in which an ordinary being commits a murder after a certain type of education or training, and a scenario in which physical determinism obtains and an ordinary being commits a murder under the effect of the causal law. In all these circumstances, given also Plum's general ability to recognize the strength of moral reasons, responsibility should be excluded simply because the murderer's conduct is caused by factors which are beyond his control, no

[4] Pereboom (1995), (2001: 117). Nevertheless, Pereboom's analysis is not uncontroversial. I will briefly consider the objection proposed by Mele in his "A critique of Pereboom's 'Four-case argument'". Mele observes that when an action is produced by a mechanical system, determinism does not really play an essential role in excluding moral responsibility. Even though the program which controls Plum's conduct produced its effects in an indeterministic way, Plum could not be considered responsible, since, in these cases (1 and 2), what really excludes moral responsibility is not the fact that determinism obtains, but Plum's inability to direct his own conduct. Something different seems to happen in cases in which Plum's conduct is defined by previous education or training or Plum is an ordinary man in a clearly deterministic world. In these scenarios, determinism and indeterminism could not be considered as interchangeable factors undermining moral responsibility, since only if we accept that Plum is not able to reject his education and modify his conduct (as in cases of coercion), we could not consider him morally responsible. Traditionally, compatibilism distinguishes between causation (compatible with moral responsibility) and coercion (which excludes moral responsibility). In Mele's view, the generalisation strategy adopted by Pereboom, who tries to show that, in every case, determinism is what excludes moral responsibility would fail, because in cases of direct manipulation determinism is not such an essential factor (Mele 2005: 75–80). I think Mele is right in saying that determinism is not essential in undermining moral responsibility in cases of direct manipulation (indeterminism could play the same role), but I believe that this is not enough to destroy the force of Pereboom's intuition. Even though we reformulate case 1 and case 2 in an indeterministic manner, the final solution does not change: moral responsibility does not find a foundation in any of the cases discussed by Pereboom. In every case, people could not be considered morally responsible in a traditional sense, because acting on the basis of factors which are beyond our control (deterministically or indeterminstically produced) undermines the presence of moral responsibility.

matter if this lack is produced by manipulation, by education or by the inner structure of our world.

While an indeterministic universe would lead to pure randomness, thus excluding the existence of moral responsibility, the possible truth of universal determinism would also oblige to reject incompatibilism in the form of *agent causation* (which would be contradicted by scientific outcomes). In that case, compatibilism would not represent the only alternative in order to save the value of morality (as an utilitarian instrument). In a deterministic universe we may be obliged to reject *strong accountability* (the idea that the agent deserves praise/blame for the action performed), because our conduct would then be oriented by factors which are beyond ourselves, but we may still maintain what Pereboom calls *weak accountability* (Pereboom 2002–2003). This form of responsibility attribution avoids taking into any account a conception of the agent as praiseworthy or blameworthy, demanding only the individual's ability to be moderately responsive to reasons and, so, to exercise something similar to a form of control over the causal history of her action, independently from the presence of alternative possibilities. Rejecting the attribution of merit/demerit does not entail a danger for interpersonal relationship, rejected in favour of objective attitudes, because most of our social practices (such as parental or adult love) are founded on feelings that are not connected with what is required by strong accountability (love for children does not lie in the idea of parent's voluntary choice; love among adults involves something different from an authentic choice) and would not be touched by the loss of reactive attitudes. The concept of *self-disclosure* (i.e. the idea according to which action belongs to the agent, expressing her identity and moral values, independently from the truth of determinism) is central to this. People who follow moral values can be appreciated for themselves, also if we avoid considering them as praiseworthy, just because their lifestyle is the expression of a self that displays their moral characteristics. The main suggestion is that the reactive attitudes could be abandoned in favour of analogous feelings, really not endangered by discoveries about determinism. Is a conception of punishment acceptable if it renounces to consider the wrongdoers as blameworthy? Once separated "blame" from "wrong", it is still possible to help people do the right thing with warnings and notifications or, in the worst cases, to isolate dangerous people (as they had serious illnesses) in order to protect the community. Renouncing strong accountability does not necessary entail the acceptance of an aesthetic consideration of ethics. In the end, the fundamental relationships that structure our existence are not based on the presupposition that people around us can be considered as strongly accountable. If universal determinism obtained, I think that this approach, better than those that try to reconcile determinism and merit, could represent a good option, giving consistent and acceptable answers to issues like the meaning of life and the status of interpersonal relationships.

References

Austin, J. 1956–7. “A Plea for Excuses”, *Proceedings of the Aristotelian Society* 57, 1–30.

Davidson, D. 1973. “Radical Interpretation”, *Dialectica* 27, 314–28.

De Caro, M. 2004. *Il libero arbitrio. Un'introduzione* (Roma-Bari: Laterza).

Fischer, J. M. 2011. *Deep Control: Essays on Free Will and Value* (New York, Oxford: Oxford University Press).

——. 2000. “Responsibility, History and Manipulation”, *The Journal of Ethics* 4: *Free Will and Moral Responsibility: Three Recent Views* (Dec. 2000), 385–391.

——. 1999. “Recent Work on Moral Responsibility”, in *Ethics* 110 (October), 93–139.

——. 1995. *The Metaphysics of Free Will: an Essay on Control* (Oxford: Blackwell).

——. and Ravizza, M. (eds.) 1993. *Perspectives on Moral Responsibility* (Ithaca: Cornell University Press).

Frankfurt, H. 1971. “Freedom of the Will and the Concept of a Person”, *The Journal of Philosophy* 68, 14, 5–20.

——. 1969. “Alternate Possibilities and Moral Responsibility”, *The Journal of Philosophy* 66, 828–839.

Ginet, C. 1966. “Might We Have No Choice?”, in K. Lehrer (ed.), *Freedom and Determinism* (New York: Random House), 87–104.

Hart, H. L. A. 1968. *Punishment and Responsibility: Essays in the Philosophy of Law* (New York, Oxford: Oxford University Press).

Hobart, R. E. 1934. “Free Will as Involving Determination and Inconceivable Without It”, *Mind*, New Series, 43, 1–27.

Honderich, T. 1993. *How Free Are You? The Determinism Problem* (New York, Oxford: Oxford University Press).

Van Inwagen, P. 2000. “Free Will Remains a Mystery: The Eighth Philosophical Perspectives Lecture”, *Noûs* 34, *Supplement: Philosophical Perspectives*, 14, *Action and Freedom*, 1–19.

——. 1983. *An Essay on Free Will* (New York, Oxford: Oxford University Press).

James, W. 1884. “The Dilemma of Determinism”, *Unitarian Review*, September, 1884, in *The Will to Believe: And Other Essays in Popular Philosophy* (Dover: Courier Dover Publications), 1956, 145–183.

Kane, R. (ed.) 2011. *The Oxford Handbook of Free Will* (New York, Oxford: Oxford University Press), 2nd ed.

——. (ed.) 2002. *The Oxford Handbook of Free Will* (New York, Oxford: Oxford University Press).

——. 1996. *The Significance of Free Will* (New York, Oxford: Oxford University Press).

——. 1985. *Free Will and Values* (New York: SUNY Press).

Libet, B. 2002. “Do we have free will?”, in Kane (2002: 551–564).

Mele, A., “A critique of Pereboom’s ‘Four-case argument’ for Incompatibilism”, *Analysis* 65, 75–80.

McGinn, C. 2007. *The Making of a Philosopher* (New York: Scribner).

——. 1999. *The Mysterious Flame* (New York: Basic Books).

McKenna, M. and P. Russell (eds.) 2008. *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson’s “Freedom and Resentment”* (Burlington, VT.: Ashgate Publishing).

Moore, G. E. 1912. *Ethics* (London: Williams and Norgate).

Pereboom, D. 2005. “Defending Hard Incompatibilism”, *Midwest Studies* 29, 228–47.

——. 2002–2003. “Meaning in Life Without Free Will”, *Philosophic Exchange* 33, 19–34.

——. 2001. *Living without Free Will* (Cambridge: Cambridge University Press).

——. 1995. “Determinism *Al Dente*”, *Noûs* 29, 21–45.

Nozick, R. 1981. *Philosophical Explanations* (Cambridge, MA: Harvard University Press).

Scanlon, T. 1998. *What We Owe to Each Other* (Cambridge, MA: Harvard University Press).

——. 1986. “The Significance of Choice”, in *The Tanner Lectures On Human Values* 7, 149–216.

Schlick, M. 1930. *Fragen der Ethik* (Wien: J. Springer).

Smart, J. J. C. 1961. “Praise and Blame”, *Mind* LXX, 291–306.

Smilansky, S. 2000. *Free Will and Illusion* (Oxford: Oxford University Press).

Soon, C., M. Brass, H. Heinze, J. Haynes. 2008. “Unconscious determinants of free decisions in the human brain”, *Nature Neuroscience* 11 (5), 543–545.

Strawson, P. F. 1962. “Freedom and Resentment”, *Proceedings of the British Academy* 48, 1–25.

Strawson, G. 2002. “The Bounds of Freedom”, in Kane (2002: 441–460).

——. 1994. “The Impossibility of Moral Responsibility”, in *Philosophical Studies* 75, 5–24.

——. 1986. *Freedom and Belief* (Oxford: Oxford University Press).

Wallace, R. J. 1994. *Responsibility and the Moral Sentiments* (Cambridge, MA: Harvard University Press).

Watson, G. 1996. “Two Faces of Responsibility”, *Philosophical Topics* 24, 227–248.

——. 1975. “Free Agency”, *The Journal of Philosophy* 72, 205–20.