

## A dialogue on Free Will

CARLOS J. MOYA

University of Valencia  
carlos.moya@uv.es

Pages: 124 – 134

---

**Much of the recent discussion concerning the problem of free will has been centered on the compatibilism/incompatibilism dichotomy. Do you think the central role attributed to this dichotomy is well deserved? And, if so, which of the two alternatives is preferable in your opinion?**

One of the most important contributions to the problems or moral responsibility (MR) and free will (FW) in the last ten or fifteen years has been the acknowledgement of, and deepening into, the high internal complexity of those two properties or capacities. R. Adams's ground-breaking paper "Involuntary sins", Watson's "Two faces of responsibility" or Fischer and Tognazzini's "The physiognomy of responsibility", among others, bring to light that complexity. This implies that such traditional questions as "Is MR (or FW) compatible with determinism?" cannot be given a response prior to determining which concept or aspect of MR or FW is in play in such questions. If, for example, one conceives of being morally responsible as being an appropriate target of criticism and of demands of justification concerning one's attitudes, decisions, and actions, then it seems there is no incompatibility between MR and determinism.

Instead, if MR is conceived as true desert of blame or praise for one's decisions and actions, then its compatibility with determinism is much more contentious. And we can even find quite different perspectives on particular aspects of FW or MR, so that one cannot say whether a particular aspect, such as the ability to do otherwise, is compatible with determinism before fixing the way in which that aspect is conceived.

There is, then, a lot of conceptual and analytic work that can be done before entering the problems of the compatibility or otherwise between FW and/or MR, on the one hand, and determinism, on the other. Theoretical work is no doubt indispensable; however, one should be careful not to lose contact with ordinary intuitions and practices, on pain of finding oneself involved in questions with scarce or no interest for human life. Moreover, given the intricacy and internal complexity of FW and MR, it is relatively easy to engage in debates where, beyond the participants' awareness, the same terms are used with different meanings and so designate different concepts. And this equivocation may even take place inadvertently within the conceptual space of a singular author.

These remarks notwithstanding, I think that the compatibility question, in new shapes which correspond to the new conceptual investigations, is here to stay, for it gets deep into the most basic questions of metaphysics. It is, I think, an unavoidable question to deal with. We cannot speculate about human FW and MR without reflecting on how these putative properties or capacities can fit in the natural world, of which human beings are a part. Free and morally responsible agents, if such there are, must live in a world, and the basic nature and structure of this world is bound to have an impact on the question of the nature of those agents and on whether this nature is hospitable to such putative properties as FW or MR.

The debate about the compatibility question is one of the roots of scepticism concerning FW and MR. Incompatibilists argue that determinism precludes some essential conditions for FW and/or MR; compatibilists, in turn, contend that indeterminism is no less damaging for other essential conditions. The result is the suspicion that FW and/or MR are simply incompatible with the natural world as such, whether determinist or not. At this point, the debate may turn into one among naturalism and anti-naturalism. Most participants in this debate side with naturalism, so refusing to consider humanity, in Spinoza's terms, as an *imperium in imperio*, an exception to the laws (determinist or not) of the natural world. Sceptical thinkers consider the task of showing how human beings can both enjoy FW/MR and be part of the natural world as hopeless. Non-sceptical thinkers, either compatibilists or libertarians, set about fulfilling that task, with uncertain results. There is wide consensus, even among sceptics, that, in *some* senses of the terms, FW and MR are compatible with the natural world and that human beings possess them, but there is also wide consensus that FW and MR *in these senses* are

not terribly interesting qualities, and, at any rate, are not what seems to be at work, either implicitly or explicitly, in ordinary moral practices and beliefs.

To state my own position, I am both a leeway and a source incompatibilist, as well as a libertarian. And I would like to reconcile this position with naturalism, in the sense of showing how MR, in the strong, “true desert” sense, and hence the FW required for it, can be enjoyed by complex physical and biological systems as we human beings plausibly are.

**In the last three decades the discussions on the so-called “Consequence Argument” have convinced many philosophers that compatibilism is not a viable theoretical option. What is your opinion on that argument?**

There is no single argument that can be considered *the* Consequence Argument. Peter van Inwagen gave an informal formulation of it that is often quoted, but even he offered three distinct formal presentations thereof. In response to objections, other authors have given alternative formulations. By “the Consequence Argument” we may perhaps understand a group of related arguments whose aim is to show that, given determinism, nobody has or ever had any alternative possibilities of decision or action. The difference among them comes from the different ways (premises, principles, rules of inference) in which this conclusion is reached starting from the initial assumption of determinism. The general idea of the argument, in its different versions, is that, given determinism, you cannot do other than you actually do, since what you do is a logically necessary consequence of (a complete true description of) the past and the natural laws and you cannot change either.

Technically, the argument has at least three problems: the analysis of “can”, the validity of certain rules of inference that are used in it, and the concept of natural laws.

Concerning the former, if the conditional analysis of “can” is correct, the argument fails: though the conditionals “if I tried to change the past, I would succeed” and “if I tried to change the natural laws, I would succeed” are false, such conditionals as “if I tried to open the window, I would succeed” seem true (even if determinism is true). On this conditional reading of “can”, you cannot derive that you could not have opened the window from the fact that you can change neither the past nor the natural laws. One problem, however, as Chisholm saw, is that this proposal only postpones the problem, for the question is now whether you could have *tried* to open the window; even if the indicated conditional is true, so that you would have opened the window if you had tried to open it, you *could* not have opened the window unless you *could* have tried to open it. And we need another conditional to analyse “you could have tried to open the window”, for which the same problem will arise.

More recently, the so-called “new dispositionalists” have also argued for the compatibility between determinism and the ability to do otherwise on the basis of a new analysis of abilities and dispositions. But this proposal is not free of problems, either. The general idea is that a disposition or ability is not lost even if it cannot be exercised. But this view of the ability to do otherwise looks too weak as a conception of FW to sustain MR.

Concerning rules of inference, many criticisms have been addressed to so-called Rule Beta, which appears in van Inwagen’s formal presentations of the argument. But rules of inference other than rule Beta (e.g., Finch and Warfield, Huemer) can also play an equivalent role without falling prey to the same criticisms. And other forms of the argument (e.g. Fischer’s) have been proposed that do not face the problems of van Inwagen’s initial versions.

Another way of eschewing the conclusion is to hold a Humean, non-governing conception of natural laws as mere regularities. Under this assumption, the Argument does not show what it is intended to, since it needs a stronger conception of natural laws as given and governing the phenomena.

I tend to think that, in some version or other, the Consequence Argument does establish its incompatibilist conclusion. But even if the argument is sound, it still leaves much room for compatibilism. Determinism may well be compatible with FW understood in roughly Humean terms as the ability to do what one wants to do; with FW understood as control over one’s actual behaviour, even if this control does not extend to alternative actions. It may also be compatible with what Harry Frankfurt has called “acting freely”. And it may be compatible with MR provided that MR does not require FW understood as the ability to choose and act otherwise. Given these alternative views of freedom, which do not necessarily include a requirement of alternative possibilities, the Consequence Argument is nowadays less pressing for compatibilists than it used to be.

**Assuming that libertarianism as such is a viable position, which of the possible libertarian views (such as those centered on agent causation, indeterminist causation or no causation at all) are preferable?**

As I have said in response to another question, I am a libertarian. As for my preferences, I confess I am rather puzzled. I used to prefer indeterminist event-causation, and I still think it is an attractive position, given its consistence with (part of) modern science. But I would happily join Kane when he writes (rather surprisingly, given the customary view of his work as an indeterminist, event-causal libertarianism): “One does not have to choose between agent (or substance) causation and event causation. You can believe in both, as I do” (*Oxford Handbook of Free Will*, p. 396). Nevertheless, to make things even messier, I am finding increasingly luring the “no causation” option, the view (swept away for decades

by the causalist orthodoxy started by Davidson) that reasons do actually explain actions, but not by being their causes. However, I think that “no-causation” theorists owe us an account of the relationship between the reason-action level and the causal processes that underlie our intentional behaviour, including the way in which the former can have an impact on the latter (the so-called downward causation). Otherwise the suspicion that reasons explanations are apparent or illusory will be hard to dispel. The viability of libertarianism is connected with the possibility of downward, and especially mental, causation. This is a metaphysical issue that libertarians cannot afford to ignore, or so I think.

Libertarianism is often accused by compatibilists of being inhuman, self-righteous, rigorist and unfair, in not taking into account the myriad of circumstances that condition human beings in their deliberation, decision, and action. It is accused of defending a sort of absolute, unconditioned FW as an intrinsic quality of human beings, beyond the reach of natural causation; hence, it is charged (e.g., Waller) with providing support to unjust systems of penal justice and punishment (USA is Waller’s favourite example), which conceive of human beings as enjoying that absolute freedom which legitimates punishment no matter their advantageous or disadvantageous circumstances.

But I think that this view of libertarianism is a caricature. Or, if it is not, it is a label for a political ideology, a sort of “Tea Party” radical individualism and neo-liberalism, rather than a serious philosophical position. Libertarianism need not be a rigorist position. In fact, since it is more demanding concerning the conditions of FW and MR, and highly sensitive to all the conditioning, responsibility-undermining factors that affect human beings, libertarianism is likely to be much more cautious in ascribing FW and MR than compatibilism, given that the latter is prepared to ascribe those qualities on less restrictive conditions, and even in the case of decisions that are fully determined by the past and natural laws.

I think that a renewed and refreshing defence of libertarianism is likely to come from action theory, especially from a reflection on the nature of reasons and practical reasoning. An idea that goes around my head for some time and that I project to argue for is that, against what is commonly assumed, practical reasoning, the process of considering reasons and deliberating about what to do on the basis of them, cannot be subject to determinism. I suspect that the concept of practical reasoning is incompatible with that of determinism. But these are high waters.

**During the last years, a growing number of philosophers and scientists have advocated sceptical, eliminativist, pessimistic, or illusionistic views on free will. What do you think of these kinds of views?**

I am decidedly against them.

Philosophical rejection of MR on the basis of several arguments clashes with

lay intuitions about real cases where questions of responsibility and desert are involved. I can think of the case of the Scott sisters, two black girls condemned to life imprisonment for orchestrating a minor robbery, who were released in 2010 after 16 years in prison, not because judges saw the injustice and disproportion of the punishment, but because it was too expensive for the American taxpayers to have them in prison: one of them had a nephropathy and needed dialysis; they were released on condition that one sister donated one kidney to the other. Another case is that of a woman in Requena (Valencia): she found a credit card and bought some food and nappies for her little children; after paying a severe fine, she was sentenced to two-year imprisonment; after strong social pressure, she was finally reprieved by the judge. Another recent case: after full rehabilitation, a young man had to be imprisoned because several years ago he sold a small quantity of cocaine.

Now compare these (and many other) cases with some recent scandals of embezzlement in Spain, like that of Luis Barcenas, who was the treasurer of the Popular Party until he was accused of embezzlement and other legal offences; the Spanish justice discovered that he had gathered a fortune of about 38 million euros (last news increases that figure), presumably through unclear and illicit means. Madoff is another example, but there are many more all around the world. When bank executives sold as inversion products packets that included subprime mortgages, so starting a process that would ruin millions of people, they acted in a morally wrong way, and I think they knew that. These people were not cultural or moral idiots. They (well, many of them, at least) were people with a good education. They had the opportunity to build their own personality to a good degree. They knew that doing those things was morally wrong, and, from a legal point of view, a criminal offence. Nevertheless, they did it. Why? They were not irrational. They had quite good reasons to act that way, only these reasons were not moral: say, richness and the power that goes with it. They were prompted no doubt by some human passions, like ambition, avarice, and so on. But these passions were not irresistible. Their conduct is perfectly understandable. But it is morally wrong. They should not have acted that way. They did also partly because they thought they would not be caught; but many of them have been (some have got away with their offences). They could have avoided acting that way, but they decided not to.

To hold, in view of an even superficial comparison between these cases, between, say, the Scott sisters and Barcenas, that neither (nobody, in fact) actually deserves or fails to deserve blame or praise, sounds like academic snobbism. Here, like in other fields, if a philosophical argument presumably shows that there is no deep difference between those cases, that there is no true praise or blameworthiness, there are important reasons to think that something must have gone wrong with the argument, rather than with the belief in the MR of the agents.

Now, if you agree that the sentence dictated against the Scott sisters was too

severe and out of proportion with their crime, then you believe that they did not *deserve* the treatment they received from the State legal system, represented by the judge. You believe there is such a thing as deserving a certain treatment in virtue of something one has done. And if you believe this, then you believe there is something in virtue of which a treatment is or is not deserved, something related to the connection between the agent and her act (and its consequences). We may call this connection ‘responsibility’. A theory of responsibility is an attempt to specify the nature and features of such a connection between an agent and her act that makes her deserve praise or blame (and eventually prize or punishment).

If you deny that there is, or even that there can be, such a connection, I think, naïve as this view may sound, that you are not entitled to the judgment that sentencing the Scott sisters to life imprisonment, and the punishment inflicted to them as a result, were unjust, in the sense that they did not deserve them.

I am aware that the preceding considerations are unlikely to move the sceptics. They know these things. My only purpose has been to remind them of what their scepticism commits them to.

**A very recent debate concerns the nature of our pre-philosophical views regarding free will. However, some surveys seem to suggest that we naturally tend towards compatibilism, others that we naturally tend towards incompatibilism. What do you think is the value of this kind of “experimental philosophy” in regard to the issue of free will?**

I must say I initially was rather sceptic about this sort of enterprise, maybe because of a philosophical upbringing deeply influenced by the Wittgensteinian sharp distinction between empirical and conceptual issues. Within this tradition, which includes the Oxford school of analysis of ordinary language, philosophy’s proper object is conceptual analysis, not empirical investigation. After this tradition lost much of its initial lure, philosophers were much more prone to engage in interdisciplinary studies together with empirical scientists and to find empirical discoveries highly relevant for philosophical problems. This was also my own case. But even on this perspective, the task of philosophy was taken to be different, more aprioristic and analytic, than that of sciences, so that the idea of an “experimental philosophy”, the view that philosophy itself, rather than empirical sciences, might engage in empirical investigation, sounded quite strange. However, my initial doubts concerning this philosophical enterprise were dispelled after reading some good samples of it by very smart people. I was especially impressed by Knobe’s study about the folk concept of intentional action and the asymmetry with which lay people apply this concept depending on apparently (from a philosophical perspective) irrelevant factors, such as the moral quality of certain side effects of voluntary behaviour. On the basis of this and other studies, I now feel

deep respect towards experimental philosophy.

Concerning more specifically philosophical problems about FW and MR, I think that a valuable contribution of experimental philosophy is the questioning of a view widely taken for granted among philosophers, according to which lay people are natural incompatibilists. What the studies seem to show is that people's intuitions vary from incompatibilism to compatibilism, depending on the scenarios they are confronted with in the inquiry. As a consequence, experimental philosophy has also fostered, among philosophers, humility and cautiousness as to the value of their 'spontaneous' intuitive judgments about real or artificially constructed scenarios. Philosophers' intuitions need not coincide with those of lay people, and may be much more permeated by philosophical reflection than they tend to think. Experimental philosophy has had salutary effects against philosophical arrogance.

A non-experimental, more 'classical' philosophical task connected with experimental philosophy concerns the interpretation of the results of field studies. How to understand and explain these results, which assumptions, unnoticed to the participants, underlie their responses to the questions posed, are by themselves highly interesting theoretical tasks. As happens with empirical sciences, there are initially several hypotheses that can explain the results, and choosing among them calls for additional and more sophisticated experimental work, which in turn leads to more sophisticated theoretical interpretations.

Experimental philosophy, however, has to face serious methodological problems, in that it has to work with subjects who have their own conceptual frames, which may differ in important ways from those of the philosophical experimenters. These may design the tests from the point of view of philosophical problems and distinctions whose point ordinary people might not be sensitive to. The choice of terminology used in the tests is also important. The term "intentional" as applied to action is more frequent in philosophy than in ordinary parlance, which would rather use "intended", "on purpose" or "voluntary". And this discrepancy may affect the subjects' responses and their experimental value.

**What do you think the relationship is between free will and moral responsibility? With regard to this, do you think that the famous Frankfurt scenarios are crucial for assessing the issue?**

I used to assume, with many others, that FW was a necessary condition of MR. But I presently consider that this view should be subjected to careful scrutiny. What has led me to be more reluctant is reflection on some ordinary ascriptions of MR that do not seem to presuppose FW. In his "Involuntary sins", Adams pointed out, rightly, that we sometimes hold people responsible for certain emotional reactions, attitudes, and traits of character, without any of these things being freely



chosen by those agents. More recently, Thomas Scanlon and Angela Smith have also followed related paths.

Maybe MR for attitudes, emotions and character features could be accommodated within a view of free will as a necessary condition of MR by classifying the MR that is ascribed in these cases as being of the aretaic or attributability variety, rather than of the accountability variety (in Watson's terms); what we would be doing in those cases is a sort of moral appraisal of the agents as good or bad, but not properly as praise – or blameworthy. It might then be held that MR in the accountability sense does require FW, and that this was the sense implicitly assumed in taking MR to presuppose FW. Attitudes, emotions and character traits are not actions and hence are no subject to the will; but, correspondingly, what is ascribed to agents by virtue of them is not full-blooded MR, but a sort of positive or negative moral mark.

However, this possible line of defence of the traditional approach, which takes FW to be a requirement of MR, is in my view seriously threatened by the existence of negligent actions. Negligent actions may give rise to quite unfortunate consequences, and agents can be, and often are, held morally responsible for them and for some of their consequences, in Watson's 'accountability' sense. Nevertheless, negligent actions are neither freely chosen nor freely performed by their agents. Some of the cases Sher designs in his book *Who Knew? Responsibility without Awareness*, are also of this kind.

However, even if we don't judge that negligent actions are freely chosen, we still assume that agents had control of some sort over them, even of the "regulative" or "alternative possibilities" kind. We tend to think (perhaps wrongly, at least in some cases) that it was within their reach to have avoided them. This suggests that the widely extended view that equates free will with control might be wrong. MR for negligent actions does not require free will or free choice, but it may require some sort of control nonetheless. Equally, if we think that agents were able to avoid acting negligently but that did not freely choose to so act, this also suggests that equating free will or free choice with the ability to avoid what one does or to do otherwise is problematic as well. The latter would be a broader notion than the former, which would be included in it, but not vice versa.

Given the existence of negligent actions, which are not freely chosen or performed, we could perhaps think of FW, in conjunction with other conditions, as sufficient, though not necessary, for MR.

I think that negligent actions are a more powerful consideration against the necessity of FW for MR than Frankfurt cases, fascinating as the latter may be. In Frankfurt cases, unlike cases of negligent behaviour, the agent performs the pertinent act consciously and voluntarily, and, in this sense, freely; hence they do not question FW or freedom, in some sense of the word, as necessary for MR. How-

ever, at several places, I have argued that no Frankfurt case so far designed has reached its end, for none is such that in it the agent is clearly morally responsible for something she did while having no morally significant alternative to doing it.

**Given the evidence coming from neuroscience and genetics, during the last few years a growing number of scholars have been arguing that the idea that we deserve blame for our bad deeds (and punishment for the worst of them) is ungrounded and should be abandoned. What is your opinion of this view?**

As a libertarian, my opinion goes against it, but the issue is important and challenging. I find this question very closely related to question number 4 and some of the remarks I made in answer to it are also pertinent to the present one. In fact, sceptical or eliminationist attitudes toward MR, understood as true desert, are quite often fuelled, at least partly, by scientific perspectives on human beings as provided by basic physics, but also by genetics and neuroscience.

Why do we tend to think that the evidence coming from the sciences points against the existence of MR as true desert? Maybe part of the answer is that blame and eventually punishment (but also praise) are justified only if persons, as conscious beings, are the true authors and sources of their conscious decisions, and these decisions are the true causes of their overt actions. But this view of persons and their decisions is threatened by the suspicion that our conscious processes of deliberation and decision-making might only be a superficial and illusory appearance of hidden forces and subpersonal mechanisms that do the real work behind our consciousness and control. Spinoza saw this already when he said that the appearance (the imagination) of freedom arises both from our consciousness of our desires and appetites and from our complete ignorance of their causes. As Dennett put it, the process of decision-making quite often takes place “at central headquarters”, beyond our conscious reach. Hence, if it is not us, as conscious beings, who really make decisions, but our brain and its subpersonal mechanisms and modules, which operate fully behind our consciousness, then it is not justified to blame or praise us for what we do, for we are not the true authors and sources of our actions.

However, between the almost exclusive emphasis on consciousness and conscious reasoning which has characterized the main stream of Western philosophical tradition and a complete devaluation of them in favour of unconscious forces, motives and mechanisms, there should be some reasonable middle ways. And part of them would be to correct the no doubt exaggerated stress on conscious reasoning with an acknowledgement of the importance, for who we are, of sub-conscious and subpersonal factors, which include psychological dispositions, abilities and psychophysical features, conditioned by our genetic endowment. If we faced the world fully deprived of emotional tendencies, behavioural dispositions and character traits, we probably would lack the minimal ground and stability that

is plausibly needed to gradually become a morally responsible agent. As for subpersonal information processing mechanisms, if we did not rely on them for most of our daily choices and tasks, we would not be able to use our ability for conscious deliberation in an efficient way and for more difficult matters. My suggestion is that we look at these subconscious factors not just as hindrances but also as possibility conditions of free and responsible agency. An additional remark is that an increasing knowledge of the subconscious factors that underlie and influence our conscious reasoning and decision-making may allow us to diminish their influence and increase our autonomy and self-mastery. Take this example. Some marketing studies show that an important factor in customers' choice of products in a supermarket is their situation in the shelves, though, if asked, customers will give quite other reasons. However, it seems that there is an important difference between someone who knows about those marketing techniques and someone who does not. The former is better positioned to get rid of these subconscious influences and choose more autonomously. This is the old idea of the Enlightenment: the importance of knowledge and reflection as a ground for freedom.

FW and MR are not primarily biological, neurological or genetic faculties. They are connected with the ability to deliberate and to consider reasons, an ability that requires language, which in turn requires sociality. FW and MR are properties of human beings as social beings, not biological beings, though, of course, certain biological traits are necessary for having those properties.

Let me end by saying that, metaphysically speaking, showing how downward, mental causation is possible is a requirement to overcoming the scientific challenge to FW and MR.