TAKING RESPONSIBILITY AND THE MANIPULATION PROBLEM

Prof. MA. Roberto PARRA DORANTES
Human Development Department, Universidad del Caribe, Cancún
MEXICO,
Email: rdparra@ucaribe.edu.mx

ABSTRACT

The main objective of this paper is to raise an objection against the analysis of the notion of 'taking responsibility' offered by John Martin Fischer and Mark Ravizza while developing their complex and attractive theory of moral responsibility. In this theory, there are two main requirements an agent must meet in order to be morally responsible for an action performed by her: first, the action must issue from a moderately reasons-responsive mechanism; second, the mechanism that leads to the action must be the agent's own. This second requirement is explicitly intended to address some problems for the attribution of responsibility that arise in the context of certain cases of manipulation—what I will refer to as 'the manipulation problem'. Fischer and Ravizza argue that this requirement (mechanism ownership) is satisfied when the agent has undergone the process of taking responsibility for that mechanism, and they spell out three conditions which, according to them, are jointly sufficient for taking responsibility. I will ultimately argue that this set of conditions fails to adequately address the manipulation problem because it either rules out certain cases of manipulation by mere stipulation or fiat (on one of two possible interpretations of these conditions), or is simply too weak to effectively rule out those same cases (on the alternative interpretation).
Keywords: Responsibility; Manipulation; Free Will; Ethics;

INTRODUCTION

My main objective in this paper is to raise an objection against the analysis of the notion of 'taking responsibility' offered by John Martin Fischer and Mark Ravizza while developing their complex and attractive theory of moral responsibility (in Fischer and Ravizza 1998). A previous version of this article was presented at the Ethical Theory and Moral Practice 20th Anniversary Conference: Dimensions of Responsibility, at the University of Pavia, Italy, on June 9th 2017.

In Fischer and Ravizza's theory of moral responsibility, there are two main requirements an agent must meet in order to be morally responsible for an action performed by her: first, the action must issue from a moderately reasons-responsive mechanism; second, the mechanism that leads to the action must be the agent's own. This second requirement is explicitly intended to address some problems for the attribution of responsibility that arise in the context of certain cases of manipulation—what I will refer to as 'the manipulation problem'. Fischer and Ravizza argue that this requirement (mechanism ownership) is satisfied when the agent has undergone the process of taking responsibility for

that mechanism, and they spell out three conditions which, according to them, are jointly sufficient for taking responsibility. I will ultimately argue that this set of conditions fails to adequately address the manipulation problem because it either rules out certain cases of manipulation by mere stipulation or fiat (on one of two possible interpretations of these conditions), or is simply too weak to effectively rule out those same cases (on the alternative interpretation).

The outline of this paper is as follows: first, I will briefly explain the debate between two sorts of compatibilists (structuralists and historicists) and its relation to the topic of agent manipulation. I will then discuss an objection raised by incompatibilists such as Robert Kane to compatibilism in general having to do with cases of a specific kind of agent manipulation—what Robert Kane calls "covert nonconstraining control" (Kane 1996, p. 65). Next I will proceed to explain Fischer and Ravizza's attempted solution to the problem of manipulation through the notion of taking responsibility, and my objection to their analysis of this notion. Finally, I will discuss a possible reply they could make to this objection, and present an answer to that reply. Ultimately, I will argue that, as regards their theory, the manipulation problem remains unsolved.

1. STRUCTURALISM VS. HISTORICISM

Within the side of compatibilists there is an interesting ongoing dispute between structuralists and historicists. On the structuralist camp, philosophers such as Harry Frankfurt[1] and Gary Watson[2] argue that, for an agent to be morally responsible for an action it is sufficient that the agent possesses at the time of the action the right configuration of mental states such as desires, preferences, and values. These theories have been also labeled hierarchical, internalist, 'time slice', and 'mesh' theories. The main idea which unifies them as structuralist theories is that it is possible to determine, based on the features of the agent right at the time of action, whether the agent's action is free or not, or whether the agent is morally responsible for it. These features are sometimes called 'snapshot properties', since in principle one could observe them by only looking at the instant in which the agent acts, regardless of her past.

On the other hand, there are historical theories, according to which the features of the agent right at the moment of the action are not sufficient for determining whether the action is free or whether the agent is morally responsible for performing it, in virtue of the existence of certain elements about the agent's past which are essential to answer these questions. John Martin Fischer and Mark Ravizza, in their influential book *Responsibility and Control*, attack structuralist theories about moral responsibility in general, and develop an account of moral responsibility which intends to be explicitly historical[3].

The main argument by Fischer and Ravizza against structuralist theories is that, regardless of the precise arrangement of mental states which is considered to be sufficient for moral responsibility, it is always conceivable that a manipulator could implant in an individual exactly that combination of mental states, and to do so in a way in which, even though the conditions for moral responsibility proposed by these structuralist theories are

---

[1] Frankfurt, Harry: "Freedom of the Will and the Concept of a Person", in his *The Importance of What We Care About*, Cambridge University Press, 1988, p. 65

[2] Watson, Gary: "Free Agency", *Journal of Philosophy* 72, 1975; pp. 205-220.

[3] Fischer, John Martin and Mark Ravizza: *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, 1998, , chapters 7 and 8.

met, it seems very plausible that the agent is in fact not responsible due to the manipulation exerted on her[4]. These considerations lead Fischer and Ravizza to affirm that the notion of moral responsibility is essentially historical; they say:

> "So, for instance, it is often thought that an individual's background can have a crucial impact on his subsequent moral responsibility. (…) And, of course, having the sort of history that arguably rules out moral responsibility is consistent with exhibiting the current time-slice features constitutive of responsibility (on the various nonhistorical models of moral responsibility). (…) We believe that there are clear cases in which the relevant snapshot features are 'put in place' in ways that rule out moral responsibility. For example, these features—the selected 'mesh' or configuration of mental states—can be induced by such processes as hypnosis, brainwashing, and even direct stimulation of the brain. When the snapshot features come from certain of these sorts of histories, it is plausible to think that the individual is *not* morally"[5]

Fischer and Ravizza criticize the use Frankfurt makes of the expression "taking responsibility" while he is discussing certain kinds of manipulation. In the relevant passage, Frankfurt says that an agent can be morally responsible in spite of having been a victim of a type of unconsented-to manipulation in which both the agent's first and second-order desires have been programmed by another person. Such an agent can still be responsible, in Frankfurt's opinion, if this person comes to identify himself with his implanted second-order order desires so that those desires become second-order volitions. Frankfurt declares:

> "In virtue of a person's identification of himself with one of his own second-order desires, that desire becomes a second-order volition. And the person thereby *takes* responsibility for the pertinent first and second-order desires and for the actions to which these desires lead him"[6].

Fischer and Ravizza present their objection to Frankfurt's theory (and, *mutatis mutandis*, to structuralist theories in general) as a dilemma: either the notion of "taking responsibility" in Frankfurt's theory is exactly the same as the identification of the agent with some of her desires, or it is involves a certain kind of process. If it is nothing more than the identification of the agent with her desires, then "*it is not clear how appeals to taking responsibility provide any defense against the manipulation examples*"[7] (Fischer and Ravizza 1998, p. 201). If it is a process, then the notion of taking responsibility introduces an historical element in the theory, i.e. it turns it into an historical theory[8]. Fischer and Ravizza thus offer a theory of moral responsibility which incorporates the notion of taking responsibility as a process which is a necessary condition for an agent to be considered morally responsible for his actions. Before turning to their analysis of this notion, it is important to clarify the kind of problem posed by certain kinds of manipulation for compatibilist theories in general.

---

[4] Ibidem cap.7
[5] Ibidem, p. 187, emphasis in original.
[6] Frankfurt, Harry: "Three Concepts of Free Action: II", *Proceedings of the Aristotelian Society*, suppl. 49, 1975; pp. 119-120.
[7] Fischer, John Martin and Mark Ravizza: *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, 1998, p. 201
[8] Ibidem

2. THE MANIPULATION PROBLEM

Fischer credits Robert Kane for having shown that certain cases of manipulation present a serious problem for compatibilist theories in general, both structuralist and historical. Fischer even refers to these manipulation cases as "*compatibilists' dirty little secret*"[9]. Kane distinguishes between two types of manipulation or "control". On one hand there is "constraining" manipulation or control, which thwarts preexisting desires, values, and objectives of the agent, as happens when we force someone to do something through coercion, threats, or physical constraint. On the other hand there is "nonconstraining" manipulation or control, in which the desires, values, and objectives relevant for the performing of the action that the manipulator expects from the person are induced in her, such as when "*we manipulate them into doing what we want while making them feel that they have made up their own minds and are acting 'of their own free wills'*"[10]. Kane says that the most interesting cases of manipulation are those in which the individual is not aware of being subjected to a manipulation of this kind, what he calls "covert nonconstraining control (CNC)"[11].

He offers as examples of CNC the cases of social engineering found in the utopian societies described in the novels *A Brave New World* by Aldous Huxley and *Walden Two* by B.F. Skinner. In these societies, people undergo since a very early age a thorough process of conditioning which leads them to have only the kind of desires, objectives, and values that are approved by the scientists in charge of the manipulation, in such a way that, during the whole of their lives, these individuals "can do whatever they choose", but only because "*they have been conditioned since childhood to want and choose only what they can have or do*"[12].

It is clear that this kind of manipulation can lead to situations in which the victims are completely satisfied with themselves and with their own motivational states, in a way that they can wholeheartedly identify with their first and second-order desires, as is required by Frankfurt's and other similar structuralist theories. However, Kane's point goes well beyond an attack to structuralist theories; according to him, "*CNC control poses problems for all compatibilist views of free agency, and not only for hierarchical views*"[13]. The general challenge compatibilism faces is "*to locate the relevant difference between the two that makes one of them (CNC control) objectionable and the other (mere determination) not.*"[14] According to Kane, it is "pretty obvious" what compatibilist must do in order to accomplish this:

> "They must emphasize the distinction between CNC control by other purposeful agents and mere determination by natural causes (without purposeful control by other agents); and they must argue that while CNC control takes away freedom in a significant sense, mere determination by natural causes does not do so"[15].

In other words, compatibilists must find a *principled distinction* between cases of causal determination which is due to certain kinds of manipulation and mere causal determinism.

---

[9] Fischer, John Martin: "Responsibility, History, and Manipulation", *The Journal of Ethics*, vol. 4, no. 4, Free Will and Responsibility: Three Recent Views, 2000; p.390
[10] Kane, Robert: *The Significance of Free Will*, Oxford University Press, 1996, p. 65
[11] Ibidem
[12] Ibidem p. 95
[13] Ibidem p. 68
[14] Ibidem p. 68
[15] Ibidem p. 68.

James Stacey Taylor, writing about the concept of autonomy, makes a similar point which in my opinion applies equally well to the topic of moral responsibility; he says:

> "An acceptable analysis of autonomy should not merely list the ways in which it is intuitively plausible that a person will suffer from a lack of autonomy with respect to her effective first-order desires, but must also provide an account of *why* a person's autonomy would be thus undermined, so that influences on a person's behavior that do not seem to undermine her autonomy (e.g. advice) can be differentiated from those that do (e.g. deception)"[16].

Some authors, like Alfred Mele and Gerald Dworkin, have realized the kind of problems posed by manipulation cases for freedom and responsibility theories but have refrained from attempting to solve it. In their work, they simply block this problem by stipulating, as a necessary condition for either freedom or responsibility, the absence of certain kinds of manipulation, without fully spelling out the details of what kinds of manipulation are excluded or of the reasons why these kinds of manipulation rule out freedom or responsibility. Thus Mele claims that "*an agent who performs an overt action A does not freely A if (...) he expresses unsheddable values in A-ing [and] those values were very recently produced in a way that bypassed his capacities for control over his mental life by value engineering to which he did not consent and are seriously at odds with autonomously acquired values of his that were erased in the process*"[17]. Dworkin, in turn, claims that a person is autonomous only if she possesses "*procedural independence with respect to her motivations*", which means that her desire to be moved to act by them has not been produced by "*manipulation, deception, the withholding of information, and so on*"[18]. Dworkin does not offer an analysis of "procedural independence", but realizes the difficulties such an analysis would imply:

> "Spelling out the conditions of procedural independence involves distinguishing those ways of influencing people's reflective and critical faculties which subvert them from those which promote and improve them. It involves distinguishing those influences such as hypnotic suggestion, manipulation, coercive persuasion, subliminal influence, and so forth, and doing so in a non *ad hoc* fashion"[19].

Fischer and Ravizza, as we will see, do attempt to solve the problem of manipulation while developing their theory of moral responsibility, and they do so by requiring that the 'mechanism' which leads the agent to act be the agent's own. A mechanism is an agent's own, according to them, only if the agent has undergone a process of 'taking responsibility' for it[20]. Let's now turn to their analysis of this notion.

3. TAKING RESPONSIBILITY

With the objections raised by Kane regarding the problem of manipulation in mind, Fischer and Ravizza offer a theory of moral responsibility that includes an element which they take to be essentially historical: the notion of "taking responsibility". According to the

---

[16]Taylor, James Stacey: "Introduction", in James Stacey Taylor (ed.), *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, Cambridge University Press, 2005. p. 7, emphasis in original

[17] Mele, Alfred: *Free Will and Luck*, Oxford University Press, 2006, p. 170

[18] Dworkin, Gerald: *The Theory and Practice of Autonomy*, Cambridge University Press, 1988, p. 16

[19] Ibidem, p.18

[20] Fischer, John Martin and Mark Ravizza: *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, 1998, p. 2017

analysis of this notion presented in chapter 8 of *Responsibility and Control*, the process of taking responsibility for a certain mechanism that issues in action contains three elements:

1) The individual "must see himself as an agent"[21].

2) The individual "must accept that he is a fair target of the reactive attitudes as a result of how he exercises this agency in certain contexts."[22]

3) The individual's view of himself specified in the first two conditions must "be based, in an appropriate way, on the evidence."[23]

The idea of a mechanism that issues in action is left relatively unanalyzed, but is supposed to include both practical reason as well as 'nonreflective mechanisms of various kinds'[24], among which Fischer and Ravizza mention 'irresistible urges'[25],'direct manipulation of the brain by someone else'[26]and 'nonreflective habits'[27]. About this point, Bratman criticizes (rightly, in my opinion) the fact that there seems to be nothing in the three conditions that specifically refers to a particular mechanism, and so, he says, "*this account of taking responsibility seems at best to get at a general idea of seeing oneself as a responsible agent*"[28]. However, I believe the conditions presented by Fischer and Ravizza for taking responsibility are susceptible to significantly more important criticisms. The third condition requires that the agent's beliefs referred to in the first two conditions (that the agent views himself as an agent and considers himself as a fair target for reactive attitudes such as moral praise and blame) be based, in an appropriate way, on the evidence.

About this condition, Fischer and Ravizza say it is intended to imply that an individual who has been covertly manipulated into having the relevant beliefs about himself (and thus satisfy the first two conditions) "has *not* formed his view of himself in an appropriate way"[29]. They admit that they do not have an analysis about how the notion of a belief based on evidence in an appropriate way must be understood in this context, and that therefore their defense of compatibilism is not decisive[30], but they claim that at least they succeed in showing that it is plausible to think that taking responsibility is compatible with causal determinism[31].

In my opinion, it is hard to see how the notion of taking responsibility is able to do all the work Fischer and Ravizza assign to it in their account of moral responsibility, especially in view of the specific conditions they offer as sufficient for an agent to take responsibility. For example, I do not believe it is clear that these conditions guarantee that their theory is historical instead of structuralist.

Bratman detects this problem too; he writes: "*Fischer and Ravizza insist that taking responsibility is a genuinely historical notion. But this seems wrong. For all that has been said, it seems that I can newly come, at the time of the action, to take responsibility, in the*

---

[21] Ibidem, p. 210
[22] Ibidem, p. 211
[23] Ibidem, p. 213
[24] Ibidem, p. 86
[25] Ibidem, p. 84
[26] Ibidem, p. 84
[27] Ibidem, p. 86
[28] Bratman, Michael: "Fischer and Ravizza on Moral Responsibility and History", *Philosophy and Phenomenological Research*, vol. 61, no. 2, 2000; p.454
[29] Fischer, John Martin and Mark Ravizza: *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, 1998, p. 236, emphasis in original
[30] Ibidem , p. 236
[31] Ibidem, p. 239

sense spelled by the three conditions"[32]. After all, it is at least conceivable that an agent may acquire the relevant beliefs instantaneously exactly at the time of action. However, I will focus here on an objection which I think goes to the center of Fischer and Ravizza's proposal. Imagine a case of unconsented-to covert manipulation in which the manipulator makes sure to completely erase any trace or evidence of the manipulation, in a way such that the person being manipulated has absolutely no way of knowing that her actions are the product of manipulation. Or suppose that the manipulator, much as Descartes' evil genius, has the capacity to make the person never even think about the possibility of being manipulated, so that the manipulation remains invincibly concealed from the victim. In any such cases, all the evidence available to the person would point to her not being manipulated. The person can then have the beliefs about herself specified by the first two conditions for taking responsibility; moreover, these beliefs would be based in all the evidence available to the agent. However, it seems highly unlikely that the agent is morally responsible for those actions. A reply that could be expected from Fischer and Ravizza to this kind of example is that the agent's relevant beliefs in the example are not *appropriately* based on the evidence. After all, as mentioned earlier, they explicitly intend this condition to rule out these cases. They could say, for example, that in order for a belief to be based on evidence in an appropriate way it is necessary that the evidentiary statements supporting that belief be always true.

Nevertheless, I think such an answer will not work. Solving the problem of manipulation requires, as mentioned before, that the compatibilist find a *principled* distinction between cases of manipulation and cases of mere causal determination. If the third condition for taking responsibility implies that the relevant beliefs of the person must be true, then Fischer and Ravizza are attempting to solve this problem by stipulation or fiat, since it seems very implausible that an individual could be an actual agent and an actual appropriate target of reactive attitudes if she is at the same time being subject to covert manipulation.

This objection can be posed in the form of a dilemma: either Fischer and Ravizza's notion of taking responsibility requires that it be false that the agent is being covertly manipulated or it does not. If it does, then their theory is susceptible to the accusation of trying to solve the problem of manipulation in an *ad hoc* fashion. If it does not, then their theory does not address the problem of manipulation at all, since it seems that one can easily imagine cases of manipulation in which the three proposed requirements for taking responsibility are met, but where the agent cannot be rightly considered morally responsible because of the manipulation.

Bratman characterizes Fischer and Ravizza's argument for supporting the conclusion that the concept of moral responsibility is 'globally historical'—meaning that the agent's past is important for moral responsibility not only because of the time required for deliberation previous to action, but also because the way in which the motivations of the agent came to exist are important—in the following way: "*the argument is that no matter how we characterize the agent's 'locally historical' properties at the time of action, this characterization ensures responsibility only if these locally historical features are free of the kind of pedigree illustrated by the manipulation cases*"[33].

---

[32] Bratman, Michael: "Fischer and Ravizza on Moral Responsibility and History", *Philosophy and Phenomenological Research*, vol. 61, no. 2, 2000; p.454
[33] Ibidem 455.

If this description of their view is correct, then it seems that the kind of historical element that the Fischer-Ravizza account of moral responsibility truly needs, according to their own arguments, can only be provided by a *negative* condition: the agent's motivations must be free of a certain kind of manipulation in order for her to be morally responsible. But then they are right back at the starting point of the manipulation problem, for a principled distinction between manipulation and causal determinism has not been found.

CONCLUSION

In this paper I have shown that solving the problem of manipulation requires that the compatibilist find a principled distinction between cases of manipulation and cases of mere causal determination. The set of necessary and sufficient conditions for the process of taking responsibility in the compatibilist theory of moral responsibility presented by Fischer and Ravizza fails to adequately address the manipulation problem because it either rules out certain cases of manipulation by mere stipulation or fiat, or is simply too weak to effectively rule out those same cases. This objection can be posed in the form of a dilemma: if Fischer and Ravizza's notion of taking responsibility requires that it be false that the agent is being covertly manipulated, then their theory is guilty of trying to solve the problem of manipulation in an *ad hoc* fashion. If it does not, then their theory does not address the problem of manipulation at all, since we can easily imagine cases of manipulation in which the three proposed requirements for taking responsibility are met, but where the agent cannot be rightly considered morally responsible because of the manipulation. Either way, the manipulation problem remains unsolved.

BIBLIOGRAPHY:

[1] Bratman, Michael: "Fischer and Ravizza on Moral Responsibility and History", *Philosophy and Phenomenological Research*, vol. 61, no. 2, 2000; pp. 453-458.

[2] Dworkin, Gerald: *The Theory and Practice of Autonomy*, Cambridge University Press, 1988.

[3] Fischer, John Martin and Mark Ravizza: *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, 1998.

[4] Fischer, John Martin: "Frankfurt-Style Compatibilism", in Sarah Buss and Lee Overton (eds.), *Contours of Agency: Essays on Themes from Harry Frankfurt*, MIT Press, 2002; pp. 1-26.

[5] Fischer, John Martin: "Responsibility, History, and Manipulation", *The Journal of Ethics*, vol. 4, no. 4, Free Will and Responsibility: Three Recent Views, 2000; pp. 385-391.

[6] Frankfurt, Harry: "Freedom of the Will and the Concept of a Person", in his *The Importance of What We Care About*, Cambridge University Press, 1988.

[7] Frankfurt, Harry: "Three Concepts of Free Action: II", *Proceedings of the Aristotelian Society*, suppl. 49, 1975; pp. 113-125.

[8] Judisch, Neal: "Responsibility, Manipulation, and Ownership: Reflections on the Fischer/Ravizza Program", *Philosophical Explorations*, vol. 8, no. 2, 2005; pp. 115-130.

[9] Kane, Robert: *The Significance of Free Will*, Oxford University Press, 1996.

[10] Mele, Alfred: *Free Will and Luck*, Oxford University Press, 2006.

[11] Taylor, James Stacey: "Introduction", in James Stacey Taylor (ed.), *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, Cambridge University Press, 2005; pp. 1-29.

[12] Watson, Gary: "Free Agency", *Journal of Philosophy* 72, 1975; pp. 205-220.