CRÍTICA, Revista Hispanoamericana de Filosofía. Vol. 48, No. 142 (abril 2016): 35–56

REVISITING FRANKFURT ON FREEDOM AND RESPONSIBILITY

LEONARDO DE MELLO RIBEIRO
Universidade Federal de Minas Gerais
lmribeiro@ufmg.br

SUMMARY: According to Harry Frankfurt's account of moral responsibility, an agent is morally responsible only if her reflected choices and actions are not constrained by an irresistible force —either from the first- or the third-person perspective. I shall argue here that this claim is problematic. Given some of the background assumptions of Frankfurt's discussion, there seem to be cases according to which one may be deemed responsible, although one's reflected choices and actions are constrained by an irresistible force. The conclusion is that Frankfurt should have acknowledged that freedom from an irresistible force is not a necessary condition for responsibility.

KEY WORDS: person, practical identity, irresistibility, spontaneity, first- and third-person perspectives

RESUMEN: De acuerdo con la explicación de la responsabilidad moral de Harry Frankfurt, un agente es moralmente responsable sólo si sus elecciones y acciones reflejadas no están constreñidas por una irresistible fuerza —ya sea de la perspectiva de primera o de tercera persona—. Argumentaré aquí que esta afirmación es problemática. Teniendo en cuenta algunos de los presupuestos de la discusión de Frankfurt, parece que hay casos según los cuales uno puede ser considerado responsable, aunque las elecciones y acciones reflejadas estén constreñidas por una fuerza irresistible. La conclusión es que Frankfurt debería haber admitido que la ausencia de una fuerza irresistible no es una condición necesaria para la responsabilidad.

PALABRAS CLAVE: persona, identidad práctica, irresistibilidad, espontaneidad, perspectivas de primera y tercera persona

1. *Introduction*

According to Harry Frankfurt's account, there are two senses of freedom associated with moral responsibility.[1] In one sense, "freedom" means "choices out of one's own will" or "choices out of one's personal practical identity",[2] which is then further analysed in terms of

[1] The core of these two senses appears in Frankfurt's two celebrated articles (1969, 1971).

[2] In his seminal paper (1971), Frankfurt actually claims to be providing an analysis of the concept of person and drawing its connections with freedom of the will. However, Frankfurt also indicates in that same paper that he takes a discussion about the source of our personal values as part of the self-same task (cf. 1971, p. 13, footnote.) In the course of the development of his view, Frankfurt makes it clear that he is interested in providing an analysis of a person's *practical identity* as the source of her values, and its relations to freedom and responsibility. Frankfurt uses

Frankfurt's hierarchical account of desires and reflected choices (cf. 1971). In the other sense, "freedom" means "absence of an irresistible force", which is then put forward by Frankfurt's thought-experiment about Black —a counterfactual threat to one's actual choices and actions (cf. Frankfurt 1969). Frankfurt seems to suppose that those two senses of freedom are individually necessary conditions for moral responsibility.[3]

However, this must be false. In what follows, I shall argue that neither freedom from Black's intervention nor freedom in the sense put forward by one's reflected choices is a necessary condition for responsibility. To see the point here it will be crucial for my argumentation to focus first on the notion of freedom as absence of an irresistible force and explore ways in which this might be interpreted in the context of Frankfurt's writings. There are many ways in which an agent can be constrained by an irresistible force. In the specific terms of the Frankfurtian framework, Black might have many "forms" —i.e., interpretations— and be an irresistible force in a variety of contexts. I shall argue that under one plausible interpretation of Black, an agent can choose and act out of her own will or personal practical identity, but do so constrained by an irresistible force, and still be taken to be responsible for what she does. This proposal will prove useful and pave the way for going one step further and arguing that, on the other hand, an irresistible force might be partly constitutive of an agent's personal practical identity in such a way that, at least in some cases, the agent's acting against her own reflected choices might be a manifestation of her personal practical identity and, as such, she may be deemed responsible for what she does, even if she does not acknowledge this.

My strategy will consist in exploring some connections between Frankfurt's accounts of freedom as they appear in the context of both his hierarchical account of desires and his account of moral responsibility, and highlight tensions between the two senses of freedom to which those accounts seem to give rise. Frankfurt's two

explicitly the expression "person's identity" in his 1988b (p. 175). For my purposes here, following Wolf (1990, chapter 2), I will be reading Frankfurt as taking up the less ambitious task of providing an analysis of one's *personal practical identity* and not of the concept of person in a full sense (given that, I assume, being a person in a full-blooded sense involves more than having a will out of one's own choices).

[3] Whether they should be read as sufficient conditions for responsibility is not entirely clear in Frankfurt's work. However, this should not be a problem for us here given that our task will be mainly negative —by raising doubts about those two alleged necessary conditions for responsibility.

senses of freedom are both very popular and taken by many as making clear and important intuitions about freedom and responsibility. However, a detailed comparison of them seems to go unnoticed in the philosophical literature. It is my intention here to highlight some aspects of such a comparison and, in doing this, attempt to show that Frankfurt's overall commitments seem to face serious difficulties. My ultimate aim is not to object to the details of either of Frankfurt's accounts of freedom, but to bring to light cases in which they seem to be in tension and which, as a result, may affect our understanding of moral responsibility. As we will see later, such tension may be the result of Frankfurt's commitments to a view of the mind which puts too much emphasis on the reflected, self-conscious first-person perspective of decision-making.

Thus, in the next sections we will proceed as follows. In 2. we will provide the details of Frankfurt's two senses of freedom and how they contribute to his account of moral responsibility. In 3. we will propose an interpretation of Black (following Frankfurt's writings) according to which Black is a natural force which may be manifested internally or externally, overtly or covertly in one's psychology. As a result of this, we will explore in 4. ways in which Black could be a natural force which manifests itself in one's psychology and, still, does not undermine or impair one's responsibility for acting. In 5. we take stock and pave the way for arguing in 6. that Black as a natural psychological force might be in some cases more revealing of one's personal practical identity than one's reflected choices.

2. *Frankfurt's Two Senses of Freedom*

Frankfurt's well-known account of *positive* freedom, as I will call it, comes in the form of his hierarchical account of desires, which is supposed to provide an analysis of one's personal practical identity, put forth originally in his 1971. Choosing and acting out of one's own reflected choices and higher-order desires is taken to be an expression of freedom in the sense that one has freedom of choice and, as such, as revealing one's personal practical identity. Roughly, the idea goes like this. Among the desires one has one can choose the desire to constitute one's will and, in turn, be effective in action. This is a sort of act of endorsement of the desires one has that one wants to be acted on. Thus, Frankfurt holds, one's choosing out of one's own will or personal practical identity means that one reflectively chooses to satisfy second-order desires about first-order desires one has, and

this must be related to responsibility for acting.[4] Actually, to be more precise and do justice to the details of Frankfurt's account, choosing out of one's own will means that one reflectively chooses to satisfy second-order desires in the sense that one "wants them to be one's effective desire or will" (1971, p. 10), that is, one not merely has a second-order desire about a first-order desire, but has a second-order desire to be *guided* or *motivated* by the first-order desire. To keep this distinction clear, Frankfurt introduces the concept of *second-order volition*. From now on, I shall use "second-order volition" instead of "second-order desire".

This is a *positive* and *first-person* sense of freedom. It is positive in the sense that it explicitly establishes a positive condition to be satisfied in order for there to be freedom. This should be a choice out of one's own will or personal practical identity which amounts to one's making a reflected choice about satisfying second-order volitions about first-order desires one has. By the same token, it is clear to see why it is a first-person sense of freedom. It is given or experienced from one's own self-conscious decision-making perspective.

Frankfurt's *negative* and *third-person* account of freedom comes embedded in his specific discussion about moral responsibility in his 1969. Here Frankfurt's famous strategy is to propose a thought-

[4] On the one hand, we will be simply assuming here that it is not relevant for our discussion to go through the details of the development of Frankfurt's refinements on his original hierarchical account, in particular, his specification of the notion of *identification*. On the other hand, since I am talking about one's choices as constitutive of one's personal practical identity I am assuming *identification* in a non-refined sense throughout the text. I think this is enough for our purposes here for two reasons. Firstly, as I take it, Frankfurt's main purpose in making "identification" more specific is to answer Watson's (1975) regress objection and, accordingly, the possibility that an agent might be a wanton with respect to her own higher-order volitions. However, this objection and an answer to it do not touch any of our main points here. Secondly, as will become clear, it is part of our ambition here to mount an attack on the role that the reflexivity condition plays in Frankfurt's account of freedom and responsibility. But Frankfurt's treatment of *identification* consists in gradually *inflating* his hierarchical account and its associated reflexivity condition. In a series of papers, he moves back and forth to characterize "identification" —sometimes as a further higher-order mental act, sometimes as a further condition, sometimes as both. As I see it, such inflation would make Frankfurt's account even *more vulnerable* to the objections raised in the course of this paper. So, if our point applies to a "thinner" or deflated version of Frankfurt's understanding of the nature and role that the reflexivity condition plays in his account of freedom and responsibility, I assume that it also applies (and arguably in a stronger way) to a more inflated version of it. For a useful discussion of the development of Frankfurt's views see Bratman 2003 and Buss and Overton 2002.

experiment in which a powerful "entity", Black, monitors an agent's choices and actions and is capable of intervening in the course of these. If the agent's choices do not go in accordance with his "will" (i.e., Black's), Black can change the course of things so as to make either the agent's choices or the agent's actions in accordance with his "will" (i.e., Black's). If the agent tried to do otherwise, Black would intervene and make the agent choose or act as he (Black) "wishes". Thus, Black is a sort of potentially irresistible force to an agent's psychology; a counterfactual threat to an agent's actual choices and actions.[5] Frankfurt's point is then to show that, in the absence of Black's intervention, one is free in the relevant sense necessary for ascriptions of moral responsibility, in spite of the fact that one could not have done otherwise (given Black's presence). In other words, one chooses and acts out of one's own free will —in the former sense of freedom—, but one would choose and act as one does anyway because Black "wants" it to be so (1969, pp. 835ff.). Only if Black had intervened would we consider such an agent's choices and actions *not* responsible. But, given that Black does not intervene, the agent can be responsible for what she does.[6]

At first sight, Frankfurt's overall theory exhibits conceptual order. There seems to be a straightforward connection between Frankfurt's accounts of freedom and his theory of moral responsibility. There is a sense of freedom which is guaranteed by the hierarchical account that seems to be relevant to ascriptions of moral responsibility in that it makes sense of one's own reflected choices. And this is further supplemented by lack of intervention from a force like Black's, that is, by freedom in the negative sense of absence of an irresistible force. So, irrespective of Black's presence, if an agent's actions come from her reflected choices —i.e., her higher-order perspective of

[5] Frankfurt talks, for example, of Black's being capable of generating in an agent "an irresistible inner compulsion to perform the act Black wants performed and to avoid others" as well as of Black's manipulating "the minute process of [an agent's] brain and nervous system in some more direct way, so that causal forces running in and out of his synapses and along [the agent's] nerves determine that he chooses to act and that he does act in the one way and not in any other." (1969, pp. 835–836.) We will return below to the issue of the many forms that Black might take.

[6] More precisely, Frankfurt's point is to show through such cases that the "principle of alternate possibilities is false" (Frankfurt 1969, p. 829); cases which "make it impossible for the person to do otherwise, but that do not actually impel the person to act or in any way produce his action. A person may do something in circumstances that leave him no alternative to doing it, without these circumstances actually moving him or leading him to do it —without them playing any role, indeed, in bringing it about that he does what he does" (ibid., p. 830).

deliberation as to the satisfaction of her second-order volitions—she achieves the highest level of freedom she can, her actions are the expression of her personal practical identity, and she can be deemed responsible for what she does.

Since my point here is not exactly to assess Frankfurt's two senses of freedom independently, but rather to assess them in relation to his intuitions about moral responsibility, my focus will not be limited to cases in which Black is merely a counterfactual threat. I shall extend my point to cases in which Black *actually intervenes*. Let me now begin to explain why.

3. *Naturalizing and Internalizing Black*

Frankfurt is unspecific about what Black could be like. He says that Black could be a wide range of things, such as a human manipulator of any kind or a programmed machine or natural forces, etc. (1969, p. 836, footnote). Frankfurt also seems to hold that no matter how we interpret Black, his point would be preserved. As far as it goes, this may well be so regarding Frankfurt's attempt to falsify the principle of alternate possibilities —a question I set aside here. But it does not exhaust the possibilities that could be explored from the thought-experiment about Black regarding Frankfurt's account of *moral responsibility*. In this respect, it is no good for Frankfurt to leave unspecified how Black is to be interpreted because, depending on how we do this, we may have different intuitions about responsibility in the light of the concrete practical cases we can encounter —or so I claim.

Now, it is part of my purpose here to show that at least under one plausible interpretation of Black, we seem to be led to draw conclusions that contradict Frankfurt's own position in his 1969.[7] This is so because there can be cases of reflected choices and actions performed under an irresistible force regarding which we seem nevertheless tempted to intuitively ascribe responsibility to the agents who perform them. In this respect, there is in particular one possible interpretation of Black, mentioned by Frankfurt himself, which is illuminating. Frankfurt says that Black could be a kind of *force of nature*.[8] This sounds persuasive, I think, especially because it helps to remove a bit of the air of artificiality that may lurk around the

[7] Although, as we will see, Frankfurt seems to suggest something different in another paper (1988c).

[8] He says, more precisely, that the idea of Black could be substituted for that of natural forces "involving no will or design at all" (1969, p. 836).

thought-experiment about Black. So, I would like to suggest that we take this interpretation of Black as a natural force seriously, as a motto for what comes next. Let us then suppose from now on that Black is a kind of natural force which can take a psychological form or be manifested in one's psychology. To mark this turn in our discussion we will henceforth no longer refer to Black as "he"; we will use "it" instead.[9]

To read "Black" as a natural force which can take a psychological form seems now to give rise to a further question, namely, whether Black should be interpreted as an internal or an external force —that is, whether Black is an entirely internal force to an agent's psychology or whether Black manifests itself in an agent's psychology but is caused by an external intervention. An intuitive case which comes easily to mind to illustrate the latter option is of a neurosurgeon operating and controlling remotely one's brain, in which case Black would be the natural forces that are the causal upshot of that intentional procedure. However, there seems to be nothing amiss with the idea of Black's being a sort of irresistible force which is entirely *internal* to one's psychology. Black could well be a state, process or event that inhabits one's psychology and may be triggered on certain occasions. In this case, Black would not only be, so to speak, *naturalized*, but also its whole process of intervention (if any) would occur from inside, i.e., entirely internal to the agent's own psychology.

Now, if Black can be entirely natural and internal to one's psychology, we may well suppose that it could intervene either *overtly* or *covertly*. An overt intervention of Black should be one of which the agent is somehow aware. A covert intervention of Black should be one of which the agent has no awareness at all. Understanding how Black could intervene in these two ways is crucial for our point here. So let us look at this more closely.

It is worth noticing that in an overt intervention of Black, the scenario does not need to be one in which the agent faces the manifestation of any sort of artificial or sui generis force. Quite the opposite, since we have read Black as a psychological force, Black's "power" may be shown literally like this: as an irresistible psychological force. So, one such a case of overt intervention of Black should look like a case in which the agent cannot resist a psychological force (which is likely against or in conflict with her own reflected choices). Get-

[9] This now means that everything that will be said from now on depends on this proposed interpretation of Black and, as such, is silent about the results of the many other possible interpretations of it.

ting back to the details of Frankfurt's hierarchical account of desires may be helpful here. One famous case in Frankfurt's writings that would fit such a description of an overt intervention of Black is the case of the unwilling addict. In Frankfurtian terms, the unwilling addict is said, from her higher-order perspective of reflected choice, to desire (or will) not to desire to take drugs, but she cannot resist her desire for drugs and ends up taking them. In this sense and under the reading of Black as a psychological force, such an agent's addiction could well be an overt intervention of Black, understood as a psychologically irresistible force. Given Frankfurt's account, the agent is thus not free in her action, since the action is determined by a Blackish psychological force against her reflect choice, and the agent is aware of it to the extent that there is a mismatch between her reflect choice and her action.

On the other hand, in the case of a covert intervention of Black, the agent would not be aware of its intervention. So, in this case, things should look entirely normal from the agent's own first-person perspective; the agent is given no clue that Black is intervening. Now, this may be at first sight puzzling. It may be so because from the agent's own perspective she sees herself as entirely free. She thinks she chooses and acts out of her own reflected choices. She knowingly and willingly satisfies her second-order volitions and thinks she has full-blooded freedom in the positive sense we have seen before. However, *ex hypothesi*, as we know from a third-person perspective, her reflected choice and action are the result of Black's covert intervention. So, this is something she does constrained by an irresistible force. If this holds, it seems to be a case in which the agent is free in one sense (the positive sense) but *unfree* in the other (the negative sense) —i.e., free in the sense that she acts out of her own reflected choices, although these are constrained by an irresistible force.

Having made sense of a naturalized and internalized version of Black, we are now prepared to explore further questions regarding Frankfurt's account of moral responsibility.

4. *Responsibility under Irresistible Force*

Drawing on what we have said so far we can formulate the following two possibilities of Frankfurtian scenarios:

> I. One may be free from one's own perspective (by satisfying one's second-order volitions) and free from the intervention of

> Black. That is to say, one acts out of one's own reflected choices (i.e., satisfies one's second-order volitions), and this is already in accordance with Black. Black does not intervene. This is, then, freedom both from the first-person perspective and from the third-person perspective.
>
> II. One may be *unfree* from one's own perspective (by not satisfying one's second-order volitions) and *unfree* from the intervention of Black. That is to say, one does not act out of one's own reflected choices (i.e., satisfy one's second-order volitions) because one's second-order volitions are not in accordance with Black. Black intervenes. This is, then, lack of freedom both from the first-person perspective and from the third-person perspective.

The first scenario refers to cases in which the entire process of deliberation seems to "go well": choice and action do not come apart. There is no mismatch between one's choices and actions, and Black does not intervene in these because the agent's will and Black's power are convergent. On the contrary, in the second scenario, things do not seem to "go well": choice and action come apart. There is a mismatch between one's choices and actions, and this is due to Black's intervention.

Now, given our previous considerations, it should be clear that there is more to be said about cases like II. Arguably, the intervention in the second scenario is, according to our former characterization, an *overt* manifestation of Black. As we have seen, an example of this might be a case of an unwilling addict. In one plausible description of one such a case, the agent is aware of her addiction as a psychological force that compels her to take drugs. She cannot resist it despite making reflected choices to the contrary. This compelling force, as we have put the point, might well be Black.

Now, the truth is that, having our previous considerations in mind, it seems that there is still room left for figuring out even more refined Frankfurtian scenarios. This is so because an overt manifestation of Black would not exhaust our possibilities here. We have also seen that Black could well intervene *covertly*. If so, there might well be a further possible scenario like this:

> III. One may be free from one's own perspective (by satisfying one's second-order volitions) but *unfree* from the intervention of Black. That is to say, one acts out of one's own reflected

> choices (i.e. satisfies one's second-order volitions) but does so due to Black's *covert* intervention. This is, then, freedom from the first-person perspective but lack of freedom from the third-person perspective.

As we have said before, this is a puzzling possibility because as far as deliberation goes from the first-person perspective, it seems to go well. There is no mismatch between one's choices and actions, in spite of the fact that Black intervenes. But, plausibly, many of these cases should reveal lack of responsibility of the agent due to the intervening irresistible force.

For example, if we suppose a case of a neurosurgeon operating externally on one's brain, it seems that we would have a straightforward answer. Arguably, such an external constraint would not be constitutive of the agent's own psychology, own will or personal practical identity. The agent's psychology in this sort of case would be the result of an external intervention that produces forces which in turn determine one's choices and actions. So, our third-person knowledge of the irresistible force in this sort of scenario would seem to be decisive with respect to our judgements and ascriptions of responsibility with respect to such an agent. Not only this. Arguably, the agent herself, if she had such knowledge, would also judge that her choices and actions are not genuinely hers given that they stem from a source which is entirely external to her.

Similarly, it seems plausible to suppose that a wide range of cases of addiction would warrant the same verdict, even though the irresistible force in these cases would be entirely internal to the agent's psychology. An agent may well willingly endorse her addiction from her first-person perspective and, yet, not be deemed responsible for acting as she does. This may be so because her addiction may affect her first-person reflected choices in such a way that our third-person knowledge of this as the result of the intervention of an irresistible force (like her addiction) could justify our not taking her as responsible. It might be the case that her addiction affects her reflected choices from her first-person perspective in such a way that she could not actually—that is, under the conditions in which she finds herself—choose differently from the way she does. However, it might not be *practically unconceivable* that the same agent, if provided with more information—both cognitive and conative—about herself and her situation as to the constraint that her addiction has on her decision-making processes, ended up judging that her reflected choices are not really hers given that her addiction is not something

she would choose to maintain if she could. Or, more precisely, she might see her addiction as incompatible with what she takes to be constitutive of herself.

So, in cases like these two we have just considered we would seem to be justified in *not* deeming the agent responsible for her choices and actions due to the intervention of an irresistible force like Black, just as Frankfurt suggested in his 1969. However, given that we have made sense of a stronger sense of an internalized and naturalized version of Black, it seems to be an open question whether Black might be part of the agent's personal practical identity in the sense that Black would not simply be internal to the agent's psychology but would also be (at least partly) constitutive of the agent's deepest values or commitments.[10] So, if this proposal makes sense, even if we can ascertain that there is a psychological force that constrains one's choices, it would not be straightforwardly clear on some occasions how we should decide whether such an irresistible force is part of one's personal practical identity. In other words, the question I wish to ask now is what grounds we have to say that the agent is *never* responsible for her actions in scenarios like III —as, I am assuming, this sort of conclusion about the situation should be at first sight endorsed by Frankfurt in his 1969.

One way to try to give support to this latter hypothesis would go like this. Suppose that we are facing a case in which Black covertly intervenes and compels the agent to perform a certain action; but, contrary to those two other cases, in this case there is apparently nothing about Black as a natural irresistible force which would, at first sight, preclude us from taking as an open question whether such a natural force would be constitutive of the agent's personal practical identity. And let us suppose further that the agent herself (from her own first-person perspective) could accept it. That is to say, the agent herself could accept that such a force might well be constitutive of her personal practical identity in that she would not care if —arguably from a more informed perspective of herself and her

[10] From now on, I will omit the qualification "at least partly" when talking about Black's being possibly constitutive of an agent's personal practical identity. Besides, it is also worth noticing that whenever I am talking about values or the agent's evaluations I am suggesting neither that values are objective nor that evaluations are cognitive. Personal values for Frankfurt are somehow constituted by the attitudes and mental acts which are part of one's hierarchical perspective. As should be clear, this is far from saying that evaluations should be understood as involving evaluative belief, cognition or judgment as a response to something like an objective evaluative reality.

situation— she discovered that there is something, say, sub-personal (that we are calling "Black") operating on her choices which might be accessible only third-personally. Even in the light of such knowledge, the agent could wholeheartedly hold that she is sure of her choices, irrespective of any truth about Black. That is to say, irrespective of Black's intervention, this fact would not change her mind. So, one explanation for this might be that she could not conceive of herself from a *practical point of view* without making the choices that she has actually made. Examples here abound. In this sense, Black could be constitutive of one's emotional patterns, of one's ingrained subjective dispositions, of one's deeply entrenched habits, of one's strong-willed dispositions, of one's internalized moral education, of one's overwhelming tastes, and so on.

Now, if all that holds, it seems that there could be a scenario in which Black, as an irresistible force, could be taken to be entirely internal to one's psychology to the extent that the person could be deemed morally responsible for what she chooses and does. In other words, in spite of the fact that Black intervenes and, as a result, that the agent chooses and acts constrained by an irresistible force, the Blackish force at stake might be constitutive of the agent's personal practical identity so that we would be entitled to hold the agent responsible. To illustrate the point, let us consider an example.

Take Allan Gibbard's case of "a civil servant who firmly rejects all offers of bribes" and who might "fear that if he dwelt on all that he is forgoing, he would yield to temptation". Let us adapt it to our debate and purposes. The idea would then be that going through a process of reflection to discover the civil servant's deep motivations could

> involve vivid awareness of the social consequences of bribery and its personal dangers. If the personal danger is minimal, though, the civil servant may well suspect that vivid realization of the social consequences of bribery would little avail against vivid realization of the pleasures accepting bribes would open to him. (Gibbard 1992, pp. 20–21)

The point I would like to raise in the light of this case is now the following one. Let us suppose that the civil servant's evaluation of his situation initially bends him towards accepting bribes. Now, although he may feel disposed to do so, this temptation may well be precluded by the intervention of a psychological force that keeps him straight. Such a psychological force may simply make it look entirely wrong to him to accept bribes. This could well be a kind of

wholehearted reaction to his practical situation. Perhaps it would just be something deeply entrenched in his personal life that he probably acquired without noticing in the course of his so many individual and social interactions with the world around him. It may be the result of his emotional patterns, his deeply entrenched habits, his strong-willed dispositions, his internalized moral education, or even an overwhelming taste. Such a range of possible psychological forces, I maintain, may well be Black.

Thus, it seems that we could be entitled to hold the civil servant responsible for not accepting bribes in one such scenario, regardless of the fact that the thing that precluded him from choosing and acting otherwise could have been just a sort of irresistible force, entirely internal to his psychology and constitutive of his personal practical identity —which we are calling here, following Frankfurt, "Black".[11]

5. *Taking Stock and One Step Further*

If our last suggestion from the previous section is tenable, it is not a necessary condition for responsibility that an agent's actions be free from an irresistible force in the third-person sense of freedom advanced by Frankfurt's thought-experiment about Black. So, in this sense, we have argued against the main conclusion about responsibility that Frankfurt tries to draw from his discussion about Black in his 1969.

Surprisingly enough, given what Frankfurt says in his 1988c he could (or even should) partly agree with our point so far. There Frankfurt considers the possibility of a subject's being provided "with a stable character or program" (1988c, p. 53) by an external manipulator[12] so that "the subsequent mental and physical responses of the subject to his external and internal environments are determined by this program rather than by further intervention on the part of the [external manipulator]". However, Frankfurt contends that in such a case there are no "compelling reasons either against

[11] Some might suggest affinities between the kind of idea I am trying to put forward here and Frankfurt's later notion of *volitional necessity* (which appears for the first time in 1988d and 1988e but is better developed in 1999b). There may be some truth in this suggestion. A detailed comparison between the idea of an irresistible force as I am trying to convey it here and Frankfurt's later notion of necessity would lead us astray from the scope of this paper. However, it is worth adding a comment on this and I will come back to it later.

[12] Actually, Frankfurt thinks of a neurologist in a way more or less similar to our previous case of the neurosurgeon.

allowing that the subject may act freely or against regarding him as capable of being morally responsible for what he does." (1988c, p. 53) So, it seems that Frankfurt himself could agree that freedom from an irresistible force in the third-person sense of freedom is not a necessary condition for responsibility as there may be cases in which it seems plausible to say that the agent's responsibility is not impaired or undermined by the fact that some of her choices are the result of a concomitant irresistible force. The point here becomes clearer when we consider Frankfurt's further comments on the same case. He says that what is at stake "is not so much a matter of the causal origins of the states of affairs in question [i.e., of the contingent psychology the agent turns out to have], but [the agent's] activity or passivity with respect to those states of affairs" (1988c, p. 54).

This is so because "to the extent that a person identifies himself with the springs of his actions, he takes responsibility for those actions and acquires moral responsibility for them; moreover, the questions of how the actions and his identifications with their springs are caused are irrelevant to the questions of whether he performs the actions freely or is morally responsible for performing them" (ibid.).

So, "the fact that [the external manipulator] causes his subject to have and to identify with certain second-order desires does not, then, affect the moral significance of the subject's acquisition of the second-order volitions with which he is thereby endowed" (ibid.).

Thus, as these passages indicate, Frankfurt seems to hold that what is *really* a necessary condition for responsibility is that the agent acts out of his own first-person, self-conscious perspective of his reflected choices and satisfies some of his second-order volitions, no matter how they were acquired —in particular, according to the terms of our discussion, no matter whether they are the result of the intervention of an irresistible force. So, given what he says in his 1988c, what seems to play the decisive role in Frankfurt's account of responsibility is freedom from irresistible force in the first-person sense of freedom, and not freedom from an irresistible force in the third-person sense of freedom.

However, whether Frankfurt's proposal as it appears in those passages is entirely satisfactory remains to be seen. In particular, it all depends on how much emphasis we would think it appropriate to put on the first-person perspective —or, more precisely, in the terms of our debate, on how much emphasis to put on the first-person sense of freedom. And this is now the question that remains to be explored here.

At first sight, considering some aspects of what we have suggested above about scenarios like II and III, we may well have given the impression that we are pointing in the same direction as Frankfurt's. However, when we look at the details of what we have said, we realize that it is not really so. We have actually given a "mixed" response to the question about the role of the first-person perspective. More specifically, taking into account what we said about the last case of the previous section —namely, the case of the civil servant we borrowed from Gibbard—, we suggested that even though an agent might become aware that a certain choice of hers was constrained by a Blackish irresistible force, she might not in the least care about it as she could not conceive of herself from a practical point of view without making that same choice. This, indeed, suggests that she somehow endorses her response to a given situation in spite of the fact that her response itself was the result of an irresistible force. However, although there seems to be a sort of endorsement from the first-person perspective, there is nothing special about the first-person perspective as there is no clear *explanatory distinction* between the endorsement itself and the operation of the irresistible force —which, as such, might be accessible from the third-person perspective. And this is so because, on the one hand, contrary to the neurosurgeon case, such an irresistible force is *internal* to the agent herself and, on the other hand —contrary to the willing addict case considered before— such internal force might be constitutive of the agent's personal practical identity.

Now, everything seems to turn on how to further characterize the agent's personal practical identity. After all, if our argument is not to be taken as question-begging we should say more about the relations of internality (to an agent's psychology) which are constitutive of an agent's personal practical identity and those which are not. In particular, we should be able to explain what is the role of the first-person in determining that relation of internality, that is, in determining which internal forces can be constitutive of the agent's personal practical identity. And we have now arrived at the crux of our discussion. If the first-person perspective is decisive in determining the relation of internality at stake, then we should agree with Frankfurt's words in those passages quoted above and accept that that's all there is to say. However, if we think that what Frankfurt says about that point is not entirely satisfactory, then we should try to make a case for an alternative proposal. Let us explore this in more detail.

If we reject a decisive role for the first-person perspective in determining the relation of internality we are talking about, what would we

have instead? We would seem to leave open the possibility that the relation of internality at stake be characterized in such a way that the agent (from her first-person perspective) might not be aware of it or acknowledge it. Granted, we should perhaps admit that, under a more informed perspective of herself and her conditions of decision-making, the first-person perspective of an agent should be decisive —or at least strongly relevant— as to what is constitutive of her personal practical identity. However, given that our ordinary situations of decision-making are far less than ideal, we should expect that agents can very often fail to recognize what is constitutive of their personal practical identities —and such a failure could, in dramatic cases, actually last a whole life. So, this means that, at least in some special cases, an agent might persistently fail to recognize what is constitutive of her personal practical identity from her own first-person perspective, although some (third-personally) well-situated observers (or a more informed counterpart of the agent herself) might ascertain that a given piece of choice and action is really constitutive of the agent's personal practical identity.

The relevant question would then turn out to be whether the agent could be held responsible for her actions in some such circumstances. If the answer is "yes", then it seems that we could call into question even freedom in the first-person sense, in the way advanced by Frankfurt, as a necessary condition for responsibility. However, as should be clear from what we have said previously, to argue about this point we will need to get back to scenarios like II above, where we find a mismatch between one's reflected choices and actions.

When considering scenarios like II, we seem to have assimilated Black's overt interventions to cases in which the agent is not responsible for her actions because in such cases there is a mismatch between the agent's own first-person, self-conscious perspective and her actions. However, given what we have just suggested, the fact that there is a mismatch between the agent's choices and actions should not perhaps always be taken as decisive with respect to conferring responsibility on the agent. It may be decisive in some cases but not in others.

But how could this be so? A first indication as to how to attempt to answer this question could be put in the form of another question: given that our intuitions supported the view that covert interventions of Black can be, in some cases, revealing of an agent's personal practical identity, what would preclude us from saying also that, at least in some cases, an *overt* manifestation of Black could be revealing

of an agent's personal practical identity? As should be clear, this would be a scenario like II except for the fact that we would be tempted to say that the agent is responsible for her actions regardless of there being a mismatch between her reflected choices and actions. Could there be cases like that?

Frankfurt himself would seem to reject such a proposal. As we have seen, his emphasis on the first-person perspective seems to preclude it. However, we may perhaps question this emphasis (and the associated intuition) by considering particular, concrete cases which seem to bend us towards another direction. If we are successful in showing that we have good reasons to accept that even overt manifestations of Black might be revealing of an agent's personal practical identity in some cases, despite absence of acknowledgement from the first-person perspective of the agent herself, we will be better positioned to hold that not even freedom in the first-person sense advanced by Frankfurt is necessary for responsibility.

6. *The Possibility of Spontaneous Personal Practical Identities*

So, how could scenarios like II make room for some cases according to which the agent might be deemed responsible? As is common in philosophical argumentation we may be better positioned to show a point or articulate an idea by providing an example. Here I suspect that there is one popular example in the literature on moral philosophy that can do the job for the purposes of our discussion. This is Huckleberry Finn's case.

Huckleberry Finn's famous story is about the conflict between his endorsed morality (which favours slavery) and his feelings of sympathy for his friend Jim (a runaway slave). In the end, as Jonathan Bennett puts it in his paper, sympathy wins over bad morality (Cf. 1974, p. 126). Huckleberry Finn helps Jim to escape and does not turn him in. Thus, in the specific terms of our debate and purposes here, we might say that Huck Finn seems to act against his reflected choice (or second-order volitions). Still, as we have suggested above, depending on further details about himself and his situation, we could well be entitled to deem him responsible for what he has done. Although this is not easily recognized by Huck Finn, he may be acting against his first-personally endorsed principles and reflected choices and, still, be acting in a way which is truly expressive of his personal practical identity.

Huck Finn's case has given rise to revisionary discussions about the role of reflected choices in practical reasoning and in the constitu-

tion of one's values. More specifically, some philosophers have argued that we should downplay the emphasis on the higher-order first-person perspective of reflected choices when talking about character, values, responsibility, and practical rationality.[13] Notably, Arpaly (2000) has offered good insights into this debate. Although Arpaly is talking about practical rationality we could certainly take a cue from her main ideas about this and apply them to our discussion about responsibility. In particular, we could borrow from her the idea that

> the right way is not always via deliberation. If we were only to call people rational when their actions were caused by deliberation, we would have to call people rational considerably less often than we do, and if we were to deny that people act for reasons whenever their actions are not the result of deliberation, then we would find that it is uncomfortably rare for people to act for reasons. (2000, p. 506)[14]

Similarly, we might claim that if we were to call people responsible only when their actions stemmed from reflected choices, we would meet with responsible people less often than we do; if we were to deny that people act responsibly whenever their actions are not the result of reflected choices, it might turn out to be rare to find responsible people. People may be deemed responsible for some actions even when decision-making or deliberation points them in another direction. This is so because we may sometimes have a better indication of an agent's personal practical identity through her first-order desires and spontaneous actions than through her higher-order or reflected choices. And such a personal practical identity can manifest itself overwhelmingly despite provoking a mismatch between the agent's higher-order perspective of choice and the agent's

[13] See Williams 1994 (p. 45), Williams 1995 (especially essay 2), Rorty 1988 (especially essays 12 and 13), and Arpaly 2000 for general points about agency and the possibility of interpreting instances of mismatches in an agent's psychology as not irrational or against the agent's own values, character or deepest commitments. Similar points applied to hierarchical theories are explored by Thalberg 1989.

[14] To be more precise, this means that, although Arpaly is putting the point in terms of *practical (ir)rationality*, and we have nowhere in our discussion of Frankfurt's account suggested that it involves a presupposition of practical rationality, Arpaly's insight could also apply to Frankfurtian scenarios in the sense of evincing a *mismatch* between first-order and higher-order perspectives of an agent. We seem to be able to make sense of this point without needing to presuppose any notion of practical rationality. So, we should read "practical mismatch" where "practical irrationality" appears in Arpaly's passage, and something like "convergence" (between first- and higher-order perspectives, where one acts out of one's reflected choices) where "rational" appears, to refer to cases that apply to Frankfurt's account.

lower-order perspective of desire and action. In other words, there may be cases in which one's first-order desires and spontaneous actions are more representative of one's personal practical identity than one's reflected higher-order perspective. Again, as Arpaly says about a case that can be nicely adapted to our Huck Finn's case:

> His visceral reluctance to abide by his decision [or reflected choice], which he himself perceives as weakness or laziness, was (let us imagine) in fact the result or the embodiment of an awareness, inaccessible at the moment to his deliberation, of all the things that are, given his beliefs and desires, overwhelmingly wrong with [turning in Jim]. Far from being the result of fatigue, major depression, or some general lack of self-control [...] [Huck Finn's] lack of [or weak] motivation [of a higher-order level] was a response to the badness of his decision or, rather, to the same factors which make his decision bad. (2000, p. 503)

Now, in the light of this insight, we can perceive what seems to be an overall difficulty with Frankfurt's hierarchical account of desires and the role he confers on this for his account of responsibility. It seems to preclude agents from having personal practical identities which are *not* discovered through the agent's higher-order perspective of reflected choices but which are, instead, discovered only through the concrete practical situations that they face —in their immediacy, spontaneity and particular appeal. Alas, this is very often the way that our personal practical identities— our deepest commitments, character and values —are revealed. It is not rare that we recognize our personal practical identities only through a sort of retrospective method of reinterpretation of our actions. And when this happens, taking responsibility retrospectively for what we have done may be the first sensible thing to come to mind.

Having said this, we are now better prepared to understand why we could take some cases of a scenario like II as cases in which the agent might be deemed responsible. This should actually be no surprise. After all, we should not expect to be able to confer responsibility on an agent's action only if the agent herself explicitly and actually endorses her action. It seems to be a common phenomenon in our lives that we frequently are not coherent in our choices in the light of our further beliefs and desires, that we recurrently desire and act in such a way that may be revealing of our deepest commitments (even though we do not immediately recognize it to be so), that we may find it difficult to openly accept that we are a particular sort of person (especially when we think we would rather be a different

one), that whenever an occasion for acting arises we fail to put into practice our previously reflected choices —thus making it possible that our reflected choices are not really expressive of our deepest commitments—, etc. These are all ways in which the agent may *not* be entirely *active* (or perceive herself as such), as it were, as to what counts as constitutive of her personal practical identity and, as such, of her responsibility for acting —contrary to Frankfurt's overall point of view.[15]

To argue now about which way to go so as to offer a complete theory of responsibility is not part of my purpose here. My task was mainly negative: I have tried to show that Frankfurt's two conditions for responsibility as provided by his two senses of freedom should not be taken as necessary. All in all, Frankfurt's hierarchical theory and overall commitments seem to make a common mistake in practical philosophy. They portray the mind as needing to be luminous and the reflective (higher-order) perspective of the agent as the incorrigible locus of the agent's personal practical identity. But this should be at most an empirical generalization about ourselves and our practical profiles.[16]

[15] Now, could Frankfurt's later introduced notion of *volitional necessity* make sense of all this? There are two reasons why it is not at all clear that it can. Firstly, part of the objections raised against Frankfurt here are an attack on the nature and role that reflection plays in Frankfurt's hierarchical account as an adequate way of capturing one's personal practical identity. But, if volitional necessities also involve such an attack, it is no longer clear that Frankfurt could maintain his hierarchical account and its reflexivity condition intact. Indeed, at some points in 1988d, 1988e and 1999b, Frankfurt seems to signal a departure from the hierarchical account and its reflexivity condition (especially at those moments in which he rejects that choices or decisions could make sense of *identification* and suggests instead that this notion could be better characterized as something with respect to which the agent is *passive*). But, secondly, Frankfurt never officially abandons the hierarchical account and its reflexivity condition, to the extent that, in his 1999a, he adds as a further condition, "satisfaction" (which clearly presupposes reflexivity), and in many passages of other papers of his 1999b he insists on the reflexivity condition. Thus, if, on the other hand, volitional necessities can be fully accommodated into the hierarchical account and the reflexivity condition, Frankfurt's later writings would still be subjected to the objections I raise here. Be that as it may, the reader may take my purposes here as having a limited scope: as a criticism directed at the "first" Frankfurt (where the hierarchical account and the reflexivity condition clearly play a decisive role in his account of personal practical identity) but as taking no stand concerning the "second" Frankfurt (of 1988d, 1988e and 1999b).

[16] To make the point clearer: Frankfurt does not require that our minds (including our practical selves) be entirely transparent to us —*nor do I*. My point here is exactly that our minds are *not* entirely transparent to us. But Frankfurt *does* require, on the other hand, that the conditions for freedom and responsibility be subjected to

7. *Conclusion*

We have seen that, depending on how we interpret Black, we may have results which go against Frankfurt's official theory of moral responsibility as it appears in his 1969. According to the interpretation we have considered, Black is a natural force which may be internal or external to one's psychology and manifest itself overtly or covertly. We have seen that, as soon as we choose this interpretation, we can conceive of a series of cases which seem to highlight a tension between Frankfurt's two senses of freedom, as put forward by his account of an agent's personal practical identity (in terms of his hierarchical account of desires) and his theory of moral responsibility.

In one sort of case, we have seen that an agent may be constrained by an irresistible force but, even so, be deemed responsible for her choices and actions. In such a case, a psychologically irresistible force like Black is so internalized to one's psychology that it has become one's own.

In another sort of case, we have seen that an agent can be responsible for acting against her higher-order reflected choices. As we have suggested, this seems to be due in part to the fact that the agent's personal practical identity may be manifested also in first-order desires and spontaneous actions rather than exclusively in her reflected choices (from her higher-order perspective).

If all this holds, we have made a case for rejecting Frankfurt's view according to which choosing and acting out of his two specified senses of freedom provide necessary conditions for moral responsibility.[17]

REFERENCES

Arpaly, N., 2000, "On Acting Rationally against One's Best Judgment", *Ethics*, vol. 110, no. 3, pp. 488–513.

Bennett, J., 1974, "The Conscience of Huckleberry Finn", *Philosophy*, vol. 49, no. 188, pp. 123–134.

Bratman, M.E., 2003, "A Desire of One's Own", *The Journal of Philosophy*, vol. 100, no. 5, pp. 221–242.

a transparency clause, given his hierarchical account and its associated reflexivity condition. So, the agent's choices and second-order volitions must be something that is transparent to her, and that, for Frankfurt, is all there is to be said about the locus of her personal practical identity. The scenarios we have seen here question exactly this.

[17] I thank George Botterill for many useful discussions on this topic and two anonymous referees for *Crítica* for pressing me on several points and helping me clarify and improve an earlier version of this paper.

Buss, S. and L. Overton (eds.), 2002, *Contours of Agency*: *Essays on Themes from Harry Frankfurt*, MIT, Cambridge, Mass.
Frankfurt, H., 1999a, "The Faintest Passion", in Frankfurt 1999b, pp. 95–107.
———, 1999b, *Necessity, Volition, and Love*, Cambridge University Press, Cambridge.
———, 1988, *The Importance of What We Care About*, Cambridge University Press, Cambridge.
———, 1988a, "Identification and Externality", in Frankfurt 1988, pp. 58–68.
———, 1988b, "Identification and Wholeheartedness", in Frankfurt 1988, pp. 159–176.
———, 1988c, "Three Concepts of Free Action", in Frankfurt 1988, pp. 47–57.
———, 1988d, "The Importance of What We Care About", in Frankfurt 1988, pp. 80–94.
———, 1988e, "Rationality and the Unthinkable", in Frankfurt 1988, pp. 177–190.
———, 1971, "Freedom of the Will and the Concept of a Person", *The Journal of Philosophy*, vol. 68, no. 1, pp. 5–20.
———, 1969, "Alternate Possibilities and Moral Responsibility", *The Journal of Philosophy*, vol. 66, no. 23, pp. 829–839.
Gibbard, A., 1992, *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Harvard University Press, Cambridge, Mass.
Rorty, A.O., 1988, *Mind in Action: Essays in the Philosophy of Mind*, Beacon Press, Boston.
Thalberg, I., 1989, "Hierarchical Analyses of Unfree Action", in J. Christman (ed.), *The Inner Citadel: Essays on Individual Autonomy*, Oxford University Press, Oxford.
Watson, G., 1975, "Free Agency", *The Journal of Philosophy*, vol. 72, no. 8, pp. 205–220.
Williams, B., 1995, *Making Sense of Humanity and Other Essays*, Cambridge University Press, Cambridge.
———, 1994, *Shame and Necessity*, University of California Press, Berkeley.
Wolf, S., 1990, *Freedom within Reason*, Oxford University Press, Oxford.

Received: June 2, 2015; revised: October 9, 2015; accepted: November 19, 2015.