Mathieu Baril
Department of Philosophy
McGill University
Montreal

A Non-Causal Account of Agential Control

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

August 2020

Table of Contents

Abstract

In the last twenty years or so, non-causal accounts of action have gained prominence in the epistemology of agency. This is partly due to the rise of moral realism and the rejection of Davidson's account of reasons for action, the view that reasons for action are mental states and events. However, recent non-causal accounts of agency have a major flaw which has to do with their metaphysical foundation: they appear to lack an account of what it is for an agent to control her behaviour. In the present thesis, I propose to develop a non-causal account of agential control to fill that lacuna, an account that is based on Harry Frankfurt's view.

My thesis comprises two parts which correspond to two distinct contributions. The first is primarily interpretive. Frankfurt's hierarchical approach has often been interpreted in the secondary literature as an account of free agency or autonomy. Against that view, I argue that Frankfurt's hierarchical approach is best seen as an account of agency *tout court*. I present three arguments to support my conclusion: it is more contextually sensitive, more precise, and more convincing to interpret Frankfurt's view as an account of agency *tout court*.

In the second part, I use my interpretation of Frankfurt to develop and defend a non-causal account of agential control, which I call "the endorsement view." The endorsement view is a modification of Frankfurt's account. Two modifications are particularly significant. First, while Frankfurt's account is based on the concept of actual endorsement, I argue that the central notion should be that of a *disposition* to endorse. That modification allows us to avoid the problematic regress to which Frankfurt's view is subject. Second, while an endorsement involves, for Frankfurt, self-reflexivity, I argue that it should rather be conceived of as a

dialogical concept. This allows us to avoid another problem with Frankfurt's view, namely, the implausible assumption that actions lacking self-reflexivity cannot count as genuinely human.

I also show that the endorsement view is a serious contender within the metaphysical debate. I examine two serious problems faced by any account of agential control: the problem of internal deviance and the problem of the disappearing agent. I show that the main alternatives, the event-causal and the agent-causal views, fail to provide a satisfactory solution to the two problems, and that the endorsement view (a non-causal view) provides the strongest solution. Thus, the endorsement view will emerge as the strongest account of agential control.

Résumé

Depuis les vingt dernières années, les théories non-causales de l'action ont gagné en popularité dans la littérature sur l'épistémologie de l'action. Cette tendance peut être expliquée en partie par la montée du réalisme moral et le rejet de la théorie de Davidson concernant les raisons d'agir – la position selon laquelle les raisons d'agir sont des états ou des événements mentaux. Ceci dit, les théories non-causales de l'action semblent avoir une faiblesse majeure au point de vue métaphysique : elles omettent d'expliquer en quoi consiste le contrôle de l'action par un agent. Afin de remédier à cette lacune, je propose de développer une théorie non-causal du contrôle de l'action, laquelle sera basée sur les écrits de Harry Frankfurt.

Ma thèse comprend deux parties et chacune de ces parties correspond à une contribution originale. La première est principalement interprétative. L'approche hiérarchique de Frankfurt a souvent été interprétée, dans la littérature secondaire, comme une théorie de l'action libre ou de l'autonomie. Contre cette idée, je soutiens que l'approche de Frankfurt est mieux conçue comme une théorie de l'action (tout court). Je présente trois arguments pour supporter ma conclusion : je soutiens qu'il est contextuellement plus sensible, plus précis et plus convainquant d'interpréter sa théorie ainsi.

Dans la deuxième partie, j'utilise mon interprétation des écrits de Frankfurt pour développer et défendre une théorie non-causal du contrôle de l'action, une théorie que j'appelle « la théorie de l'approbation (endorsement) ». La théorie de l'approbation, telle que je la développe, est une modification de l'approche hiérarchique de Frankfurt. Deux modifications sont particulièrement significatives. Premièrement, alors que l'approche de Frankfurt est basée sur la notion d'une approbation actuelle, je soutiens que le concept central devrait être celui

d'une disposition à approuver. Cette modification nous permet d'éviter le problème de la régression auquel l'approche de Frankfurt est sujette. Deuxièmement, alors qu'une approbation implique, pour Frankfurt, la réflexion sur soi, je soutiens qu'on devrait plutôt concevoir l'approbation comme un concept dialogique. Cela nous permet d'éviter un autre problème avec l'approche de Frankfurt, à savoir l'invraisemblable position selon laquelle les actions dépourvues de réflexion sur soi ne sont pas considérées comme véritablement humaines.

Dans la deuxième partie, je démontre aussi que la théorie de l'approbation, telle que je la développe, est une concurrente solide dans le débat métaphysique concernant la nature de l'action. Pour ce faire, j'examine deux problèmes auxquels toute théorie du contrôle de l'action doit faire face : le problème de la déviance interne et le problème de la disparition de l'agent. Je démontre que les solutions de rechanges, les théories causales-événementielles et les théories causales-agentielles, n'offrent pas une solution satisfaisante à ces deux problèmes, et que la théorie de l'approbation (une théorie non-causale) offre la meilleure solution. Donc, je démontre que la théorie de l'approbation s'avère être la théorie du contrôle de l'action la plus convaincante.

Acknowledgments

The production of the present thesis was made possible with the financial support of the FQRSC. Considerable financial support was also provided by McGill's Department of Philosophy and by two research groups, GRIPP and GRIN.

I am very grateful to my two supervisors, Natalie Stoljar and Christine Tappolet. Natalie helped me structure and clarify my ideas, encouraged me to always do better and to surpass myself, and was extremely patient. I am sorry that she had to read so many painful drafts of my thesis! I am also very grateful to Christine, who helped me discover and nourish my passion for moral psychology and who continuously supported me throughout my studies. Thanks to the staff who work for the Department of Philosophy, especially to Angela Fotopoulos, Mylissa Falkner and Andrew Stoten. Many thanks to my colleagues, especially to those in my cohort, David Collins, Hakan Genc, Kosta Gligorijevic, Éliot Litalien, Andre Martin, Martina Orlandi and Charlotte Sabourin who made me feel like I belonged to a community. (A big thank you to David Collins for proofreading this thesis!) I am also very thankful to Jeremy Lane for his linguistic advice, to Jamiey Kelly for his eagerness to engage in philosophical debates and for helping me believe in the worth of my work, and to Samuel Turcotte-Bois for his *joie de vivre* and for making my life happier during this long process. And finally, I want to thank my parents, Liselle Falardeau and Rock Baril, who always encouraged me to pursue my studies, even after I was too old to be a student!

Introduction: Frankfurt and the Metaphysics of Agency

Frankfurt's hierarchical approach has attracted a lot of attention in the last few decades, especially in the literature on free agency[1] and autonomy.[2] While Frankfurt's contributions to these fields of study is undeniable, he has made a much more interesting contribution to another discipline: the philosophy of action. That contribution, however, has been obscured by the way in which his approach has been interpreted in the secondary literature.[3] The main aim of my thesis is to revitalize Frankfurt's contribution to the philosophy of action and to show how it can inform recent debates in that field of study.

Frankfurt's hierarchical approach is both simple and elegant. It is based on the idea that there are mental attitudes of different orders, desires in particular, and that these desires might be in harmony or in conflict. There are, on Frankfurt's view, first-order desires, those desires that we have to perform a certain action like the desire to run. But we may also desire to be motivated in a certain way. In that case, we have a second-order desire, that is, a desire to desire to perform a certain action. The desire to be moved by the desire to run is an example of a second-order desire. Since the object of a second-order desire is a first-order desire, these desires may be in harmony or in conflict. When I have the motivation to run, and I want to be moved in that way, my desires are in harmony. When I do not have the motivation that I want to have, my desires

[1] See, for example, Locke 1975, Watson 1975, Thalberg 1978, and Benson 1994.
[2] This is often known as the "autonomy as authenticity" view. See, for example, Friedman 1986, Christman 1989, Oshana 1998, Mackenzie and Stoljar 2000, J. S. Taylor 2005, and Mackenzie 2014.
[3] Two notable exceptions are Velleman 1992 and Mayr 2011.

are in conflict. We should add here that Frankfurt introduces further distinctions,[4] but we may leave them aside for now.

Two main interpretations of Frankfurt's hierarchical approach have emerged. The first is about a certain kind of freedom or free agency. According to this interpretation, first-order desires, just like external objects, may constrain us. And they constrain us when they are in conflict with our second-order desires. A person may want to be motivated by a desire to refrain from taking drugs, but she may also struggle with her compulsive desire to shoot herself. And she might finally succumb to her desire to take drugs. That person is, so to speak, the slave of desire. On the other hand, when the person's first and second-order desires are in harmony, when the person has the motivation that she wants to have, her mental life seems to flow naturally. The person encounters no obstacle in the mental realm. In a case like that, she experiences a certain kind of freedom.

When interpreted in this way, Frankfurt's hierarchical approach has been seen as a major contribution to our understanding of free agency. Before Frankfurt, free agency had been associated with the notion of alternate-possibilities freedom. It was assumed that a person acts freely only when she could have acted otherwise, that is, only when she has the freedom to do otherwise. What Frankfurt introduced is a compelling alternative to that account of alternate-possibilities freedom, what is sometimes called the "sourcehood" view.[5] On the sourcehood

[4] In particular, Frankfurt draws a distinction between first-order desires and the will (a specific kind of first-order desire) and between second-order desires and second-order volitions (a kind of second-order desire). We shall discuss these concepts in Chapter 1.
[5] See O'Connor and Franklin 2020.

view, what matters is the source of one's action, not the counterfactual possibilities. Following Frankfurt, the sourcehood view has had many followers.[6]

The second interpretation is about autonomy. According to that view, when a person has the desire that she wants to have, we can say that she *really wants* to perform the action in question. Her first-order desire is said to better represent who she really is, compared with a desire that has not been endorsed at the second level. In other words, a first-order desire that a person wants to have is a desire that corresponds to that person's conception of herself, to her ideal self. And when the person acts on such desire, she can be said to be autonomous or for her actions to be determined by her true or authentic self.

When understood in this way, Frankfurt's approach has been seen as a major contribution to our understanding of autonomy. Indeed, Frankfurt is often seen as a pioneer in the study of autonomy, alongside Dworkin.[7] The two philosophers are seen as initiating a certain kind of approach to autonomy, the so-called "content-neutral" view.[8] The content-neutral approach is based on the idea that a person may be autonomous regardless of the content of her desires and her conception of the good life. This view has been seen as quite attractive to many, including some feminist philosophers, because autonomy, on the content-neutral approach, does not require that a person endorses specific values such as independence and individualism: autonomy appears to be compatible with relation of dependence and care, for example.[9]

My interpretation of Frankfurt differs from these two interpretations. That being said, I do not want to deny that these two interpretations are accurate. In fact, I believe that we can find

---

[6] See, for example, Watson 1975, Stump 1988, Wolf 1990, and Fischer and Ravizza 1998.
[7] See Dworkin 1970.
[8] For the notion of a content-neutral approach, see Christman 1991.
[9] On this, see Mackenzie and Stoljar 2000.

textual evidence for both of them. Nor do I want to deny that Frankfurt has made an important contribution to the study of free agency and autonomy. My position is rather that Frankfurt's hierarchical approach can *also* be seen as an account of agency. That is because Frankfurt gives a special meaning to the concept of free agency and autonomy. For him, the two notions can be cashed out in terms of activity and passivity: a constraint is a passive happening and heteronomy is a form of mental passivity. The heteronomous person is, in Frankfurt's words, a "passive bystander."[10] That being said, I shall argue that his approach is *best seen* as an account of agency, and that is so for three distinct reasons.

The first reason why Frankfurt's hierarchical approach is best interpreted as an account of agency is that it is *more contextually sensitive* to see his approach as such. Frankfurt's work is often interpreted out of context, especially in the literature on autonomy where Frankfurt is perceived as a pioneer. But Frankfurt was not writing in a vacuum. Rather, Frankfurt established an explicit, and sometimes an implicit, dialogue with different philosophers, two of the most important being central figures in the philosophy of action, namely, Davidson and Chisholm.

The second reason is that it is *more precise* to interpret his approach as an account of agency. As we shall see, Frankfurt's account of free agency and autonomy is based on a certain account of constraint and heteronomy. Constraints are passive happenings and heteronomy is a form of mental passivity. That account is prone to misunderstanding, since it does not correspond to our intuitive notions of constraint and heteronomy. For that reason, interpreting Frankfurt's account as an account of agency allows us to avoid misunderstandings and to acquire a sharper grasp of his view.

---

[10] See for example Frankfurt 1971, p. 22 and Frankfurt 1975, p. 54.

The third reason why Frankfurt's approach is best seen as an account of agency is because his theory is much *more convincing* when it is interpreted as such. Frankfurt's hierarchical approach, when understood as a theory of free agency or autonomy, has been widely criticized. One critique that I shall examine is the view that his approach is over-inclusive. The problem, on this view, is that Frankfurt considers as free or autonomous many actions that should not count as such. I believe that this critique is accurate. That being said, once we reformulate Frankfurt's hierarchical approach as an account of agency (*tout court*), the critique loses its strength and Frankfurt's view appears much stronger.

After showing that Frankfurt's hierarchical approach is best interpreted as an account of agency, I will develop a Frankfurtian account of agency that addresses recent debates in the philosophy of action. The aim, then, is to revitalize Frankfurt's contribution to the philosophy of action.

In the last forty years or so, the philosophy of action has been dominated by what is called "the causal theory of action" or "the standard account," a view that is often attributed to Davidson.[11] The standard account is based on two distinct positions.[12] The first concerns the *metaphysics of agency*. It answers the what-question, namely, "what is an action?" According to the causal theory of action, an action can be defined as a behaviour that is caused by an intention and a desire-belief pair. The second position is *epistemological* or *explanatory* and concerns the nature of action explanation. It answers the why-question: "why did a person do φ?" It is generally agreed that to explain an action we need to identify the person's reason for acting. The controversial issue concerns the nature of these reasons. According to the standard account, the

---

[11] Davidson 1963.
[12] See, for example, Sandis 2009, p. 2; Aguilar and Buckareff 2010, p. 9; Schlosser 2011, p. 13; Mele 2013, p. 162; and D'Oro and Sandis 2013, p. 22.

reason for which a person acts is the cause of her action, namely, the desire-belief pair that causes her behaviour.

Davidson's view on agency provoked two main debates, which correspond to these two aspects of his theory. The first main debate concerns the metaphysics of agency. The standard account has been seen as quite attractive because it is an attempt to "naturalize" agency: the standard account uses the ontology of the natural scientific worldview to define agency. More specifically, it is based on a conception of the world according to which every event is caused by previous events, what is called "the natural event-causal order." The main advantage of this view is that it does not create an ontological divide between the human realm of action and the realm of natural phenomena.

While this view has (at least) one very attractive feature, it is not obvious that it succeeds. The standard account is subject to some problems,[13] two of the most serious being the problem of internal deviance[14] and the problem of the disappearing agent.[15] These two metaphysical issues can be cashed out as two parts of a debate over the nature of *agential control*. According to the standard account, we exercise agential control over our bodily motions when our desires and beliefs cause them: desires and beliefs bear agential control. The problem of internal deviance shows that this account of control is flawed or insufficient because it allows cases of causal deviance in which control is lost. The problem of the disappearing agent, on the other hand, shows that the standard account of control is not an account of *agential* control because no agent is involved. That is why the metaphysical debate in question can be cashed out as a debate

[13] See for example Aguilar and Buckareff 2010.
[14] On the problem of internal deviance, see for example Chisholm 1964, Taylor 1966, Davidson 1973, Frankfurt 1978, Peacocke 1979, and Bishop 1989.
[15] On the problem of the disappearing agent, see Melden 1961, Chisholm 1964, Taylor 1966, Nagel 1986, Velleman 1992, Bratman 2001, Enç 2003, Hornsby 2004, Schroeter 2004, Schlosser 2011, and Steward 2013.

over the nature of agential control. This debate is still very much alive today, and philosophers have recently attempted to solve the two issues.

The second debate is epistemological and concerns the nature of reasons for action or *reasonology*. Here again, the standard account has been seen as quite attractive. The main reason for this is that the standard account models the explanation of human action on the explanation of natural phenomena, creating a kind of explanatory monism. To explain natural phenomena we typically identify their causes. For example, to explain why a bridge collapsed we will look for its cause (say, an earthquake). Human actions, on the other hand, are explained by identifying the reasons for which they are performed, a process called rationalization.[16] But according to Davidson, reasons are causes, and thus rationalization is a species of causal explanation.

While this account of action explanation appears to be quite attractive, it has also been widely criticized. Davidson's account of reasons for action is often characterized as a form of "psychologism": on his view, reasons for action are psychological states like beliefs and desires. But, in the late twentieth century, moral realism gained prominence and the view that reasons are psychological states lost in plausibility.[17] One popular critique of psychologism is Dancy's claim that the reasons for which we act are the kind of things that can be good reasons,[18] and good reasons are typically construed as facts, states of affairs, or true propositions.[19] Thus, the reasons for which we act are not psychological entities, because psychological entities are not facts, states of affairs, or true propositions.

---

[16] Davidson 1963.
[17] See D'Oro and Sandis 2013, p. 28.
[18] See Dancy 1995, 2000. See also Alvarez 2017.
[19] See Dancy 1993 and Raz 1999.

In that context, different approaches have been developed to solve the metaphysical and epistemological difficulties to which the standard account is subject. As we saw, the standard account is based on the event-causal framework. Some philosophers have attempted to refine and develop the event-causal view to solve the difficulties to which the standard account is subject.[20] Others have opted for an alternative causal approach: the agent-causal view.[21] Finally, some have abandoned causal approaches entirely and opted for the non-causal approach. Before Davidson, the non-causal approach was the dominant view, associated with post-Wittgensteinians like Ryle, Anscombe, Melden and Von Wright.[22] In the last few decades many people have attempted to revive the non-causal view such as Wilson, Sehon, and Dancy,[23] among many others.[24]

The non-causal view has fared quite well with respect to the second debate, the epistemological debate concerning the nature of reasons for action. We may distinguish two broad categories of non-causal approaches to action explanation: the teleological approach and the non-teleological approach. Take for example Dancy's non-teleological view.[25] On his view, reasons for action are facts and facts do not cause actions.[26] For teleologists, however, reasons are typically construed as states of affairs towards which we aim.[27] And, just like anti-causal non-teleologists, anti-causal teleologists claim that reasons are not causes. One advantage of these views is that they are based on a non-psychologistic account of reasons for action – the

---

[20] See Peacocke 1979, Brand 1984, Bratman 1987, Dretske 1988, Bishop 1989, Velleman 1992, Mele 1992 and 2003, Enç 2003, Schlosser 2007 and 2011, Aguilar 2010 and 2012, and Wu 2016.
[21] See Chisholm 1964, Taylor 1966, Alvarez and Hyman 1998, O'Connor 2000, Clarke 2003, Lowe 2008, and Mayr 2011.
[22] See Ryle 1949, Anscombe 1957, Melden 1961, and Von Wright 1971. On this post-Wittgenstanian movement, see Alvarez 2013, D'Oro and Sandis 2013, and Schumann 2019.
[23] Wilson 1989; Sehon 1994, 2005, 2016; Dancy 2000.
[24] See also, Ginet 1990; Tanney 1995, 2005, 2009; Hacker 1996, 2009; Rundle 1997; McCann 1998; Hutto 1999; Schroeder 2001; Schueler 2003, 2009, 2019; D'Oro 2007; Ruben 2009; Candish and Damnjanovic 2013.
[25] Dancy 2000.
[26] Cf. Mellor 1995.
[27] See, for example, Sehon 2005 and 2016.

psychologistic account being, as we saw, more and more frequently rejected. For that reason, the non-causal view has enjoyed a good reputation in the last few decades such that some have talked about "an anti-causal *fin de siècle* movement"[28] or mentioned that "a teleologist backlash against causalism" has emerged in the last twenty years.[29]

That being said, the reputation of the non-causal view is quite different in the metaphysical literature. In addition to the teleological and non-teleological anti-causalist accounts of action explanation, which often entail a certain metaphysical conception of agency,[30] another non-causal approach has been identified: what we may call "uncaused volitionism," the view that all actions involve an uncaused act of will.[31] Within the metaphysical debate, however, the non-causal approach has not been taken seriously[32] and has often been quickly dismissed. We can take the following two cases as examples.

In his recent entry in the *Stanford Encyclopedia of Philosophy*, Schlosser discusses the three main metaphysical frameworks for agency and claims that "non-causal theories… are widely rejected."[33] What he calls the non-causal approach is in fact what I have called "uncaused volitionism." The reason why uncaused volitionism is generally dismissed, on Schlosser's view, is that it cannot explain agential control: "Volitionist theories… do not explain what an agent's exercise of control consists in." Schlosser's point is that volitions, according to that view,

---

[28] See D'oro and Sandis 2013, p. 8 and Schumann 2019, p. 21.
[29] See D'oro and Sandis 2013 and Schumann 2019, p. 23.
[30] These accounts generally assume that the concept of acting for a reason or intentional action is more fundamental that the concept of acting, since unintentional actions depend on the performance of intentional actions.
[31] See Ginet 1990 and McCann 1998.
[32] The most serious contender to the event-causal approach is taken to be the agent-causal approach. See Aguilar and Buckareff 2010.
[33] Schlosser 2019.

randomly occur to the agent and for that reason are not under the agent's control: they cannot bear agential control.[34]

Similarly, in his recent book *Understanding Human Agency* (2011), which focuses on the metaphysics of agency, Mayr quickly dismisses the non-causal view. For Mayr, the non-causal view is what he calls "intentionalism," another name for the anti-causal teleological view previously discussed. The problem with intentionalism, for Mayr, is the same problem that Schlosser attributes to uncaused volitionism: the intentionalist approach "cannot satisfactorily deal with the element of the agent's control which is necessary for actions."[35] The point is that the intentionalist view seems to lack an account of how agents control their behaviour.

I take Schlosser's and Mayr's position to represent the current opinion on the non-causal approach in the metaphysical literature. And, overall, that opinion seems accurate: non-causalists have generally neglected the notion of agential control. But that certainly does not mean that we cannot develop a compelling non-causal account of agential control. The main aim of the present thesis is to show that it is possible to develop such an account of agential control, and to show, thereby, that the non-causal approach to agency is a serious contender within the metaphysical debate.

It is in that context that Frankfurt's contribution to the philosophy of agency should be revitalized, for his work contains a very interesting non-causal account of agential control, one that has been obscured in the secondary literature. In my discussion of the non-causal approach, I will focus on one particular version of Frankfurt's account of agency, the one that is the most

---

[34] On this point, see also O'Connor 2000, pp. 25–26 and Clarke 2003, pp. 17–24.
[35] Mayr 2011, p. 36.

compelling,[36] which I call "the endorsement view." Frankfurt's non-causal view is distinct from the three other versions of the non-causal account. It can be characterized as a *structural* approach.[37] The main idea of the structural view is that we should look, not at the way in which a behaviour is caused (or uncaused) or at the reason that explains it. Rather, we should look at the structure of a behaviour. On the endorsement view, an action involves a complex structure composed of a bodily and/or mental motion plus the agent's endorsement.

The endorsement view is based on a certain account of agential control. It is based on the idea that when we endorse our bodily or mental motion, we exercise a kind of non-causal control, which I shall call "standby control." Second, it is based on the idea that an endorsement "plays the role of the agent," as we might say. Thus, to endorse our bodily or mental motion is to exercise *agential* control. As we shall see, this account of agential control appears to be the most compelling because it provides the strongest solution to the problem of internal deviance and the problem of the disappearing agent.

That being said, Frankfurt's endorsement view is subject to some difficulties, some of which are well-known in the secondary literature.[38] In my development of the endorsement view, I will highlight some of these difficulties and propose a series of modifications. Two modifications are particularly significant. The first concerns the regress that Frankfurt's view generates. An endorsement is for Frankfurt a kind of decision, which is itself a reducible mental act. But, as I will show, this view generates a problematic regress. To fix that issue and to

---

[36] In my thesis, I identify two version of Frankfurt's approach. The first is an event-causal account of agency. It is developed principally in "Freedom of the Will and the Concept of a Person" (1971). The second it a non-causal account, which I call "the endorsement view." It is mainly developed in "The Problem of Action" (1978). The endorsement view is, in my opinion, the most compelling and the most innovative.

[37] Frankfurt's view can also be characterized as a teleological view (see Frankfurt 1978), but this is not the aspect on which I will focus.

[38] The regress objection is one very well-known objection. See, for example, Watson 1975, Friedman 1986, Christman 1989, Meyers 1989, Velleman 1992, J. S. Taylor 2005 and Mayr 2011.

generate an unproblematic regress, I argue that we need a *dispositional* account of endorsement. The second significant modification concerns Frankfurt's account of the agent, which is based on the notion of self-reflexivity. I will show that this view has an implausible implication for moral responsibility. To avoid that issue, I argue in favour of a *dialogical* account of the agent. Thus, my account of agency will be, unlike Frankfurt's, a dispositional *and* dialogical one. Again, these are probably the most significant differences between my account and Frankfurt's account, although they are not the only ones.

While the main aim of the present thesis is to show that the non-causal approach is a serious contender within the metaphysical debate, I will also show how the endorsement view can inform the epistemological debate. As we saw, one popular account of action explanation is the anti-causal teleological view. On that view, reasons for action are typically construed as states of affairs towards which we aim. And it is also argued that states of affairs do not cause actions. One difficulty with this view is known as "Davidson's challenge." The challenge consists in explaining the distinction between justifying and motivating reasons without appealing to causation. Many anti-causalists have attempted to respond to that challenge, including Wilson[39] and Sehon.[40] But Mele, who has been one of the main detractors of the non-causal view in the last two decades,[41] has criticized these attempts on metaphysical grounds. As he says, "because teleologists have not offered an acceptable account of what it is to *act*, or to 'direct' one's bodily motions, they have not offered an acceptable account of what it is to act for

[39] See Wilson 1989.
[40] See Sehon 1994, 2005, and 2016.
[41] See Mele 2000, 2003, 2010, and 2019.

the sake of a particular goal."[42] Another aim of my thesis is to show how the endorsement view can respond to Mele's critique, and thus to strengthen the case for the non-causal approach.

The present thesis comprises two parts. Part I is primarily interpretative. Its main aim is to show that Frankfurt's hierarchical approach is best seen as an account of agency. The first part is divided into three chapters. In Chapter 1, I put forward a new interpretation of Frankfurt's hierarchical approach, the view that his hierarchical approach is a volitional account of agency, and I show that my interpretation is more contextually sensitive than the alternatives. In Chapter 2, I examine the view that Frankfurt's hierarchical approach is an account of free agency or autonomy, and I show that my interpretation is more precise. In Chapter 3, I argue that my interpretation is more convincing because it avoids a common critique of Frankfurt's view: the claim that his account is over-inclusive.

In Part II, I develop a Frankfurtian non-causal account of agency, what I call "the endorsement view," and I argue that this view is stronger than the main alternatives and thus that it is a serious contender within the metaphysical debate. The second part is divided into four chapters. In Chapter 4, I argue that the endorsement view provides the strongest account of control and thus the strongest solution to the problem of internal deviance. In Chapter 5, I examine the problem of the disappearing agent and I argue, *contra* Schlosser,[43] that it is a real issue for the standard account. In Chapter 6, I show that the endorsement view, as I develop it, provides the strongest solutions to the problem of the disappearing agent, and thus the strongest account of *agential* control. Finally, in Chapter 7, I argue that the endorsement view can provide

[42] Mele 2000, p. 287.
[43] Schlosser 2011.

resources for anti-causal teleological accounts of action explanation and that it can respond to Mele's critique.

Part I: Rethinking Frankfurt's Hierarchical Approach

The first part of this thesis is primarily interpretative. Its main goal is to develop and defend a new interpretation of Frankfurt's hierarchical approach that views it as a volitional account of agency.[44] To defend that view I present three main arguments, with each of these arguments corresponding to one chapter.

In Chapter 1, I show that Frankfurt's hierarchical approach can be interpreted as a volitional account of agency and I argue that it is *more contextually sensitive* to interpret his approach as such. More specifically, I argue that, in 1971, Frankfurt introduced his two-level framework to develop an event-causal account of decision, which is based on the notion of second-order desire. I also explain that this was part of a larger project which consists in showing, *contra* Chisholm, that moral responsibility does not require the freedom to *decide* otherwise.

In Chapter 2, I examine the main alternatives to my interpretation, the views that Frankfurt's approach is an account of free agency or autonomy, and argue that my interpretation is *more precise*. I show that Frankfurt construes constraints as passive happenings and heteronomy as a form of mental passivity, and I argue that this view is not an intuitive one and that it is, for that reason, prone to misunderstanding. I conclude that it is favourable to interpret

[44] Velleman (1992) and Mayr (2011) suggested that Frankfurt's approach might be more about agency than autonomy or free agency, but they do not argue that his approach is a *volitional* account of agency. In that sense, my interpretation is truly original. What I mean by a volitional account is not what Schlosser means by it. As we saw in the introduction, for Schlosser the volitional approach is based on the concept of an uncaused act of will (Schlosser 2019). Here, by a volitional account of agency I simply mean an account that requires the performance of an act of will.

Frankfurt's approach in agential terms as it allows us to better grasp what is going on in his work.

In Chapter 3, I argue that Frankfurt's approach is *more convincing* when we interpret it as an account of agency. To do so, I examine a common critique of his approach, the view that it is over-inclusive because it counts as free and autonomous too many actions that should not count as such. I argue that this critique is correct but that it loses its strength once we reformulate Frankfurt's hierarchical approach as an account of agency.

The reader should be aware that I focus, in these first three chapters, on Frankfurt's hierarchical approach as it is developed in "Freedom of the Will and the Concept of a Person" (1971). I argue that Frankfurt's account of agency is, in that paper, an *event-causal* one. That being said, Frankfurt has made some important modifications to his account in the few years following that paper. In particular, he seems to abandon the event-causal approach and to adopt a *non-causal* view in his 1978 paper, "The Problem of Action."ot While I focus on Frankfurt's event-causal account of agency in Part I, I will focus on his non-causal account in Part II. This non-causal account is the one that I find the most compelling.

Chapter 1: Frankfurt's Hierarchical Approach as an Account of Agency

Frankfurt's hierarchical approach has often been interpreted, in the secondary literature, as an account of free agency or autonomy. In this first chapter, I develop another interpretation of Frankfurt's two-level approach. More specifically, I argue that it can be seen as a volitional account of agency.

I also provide a first reason to support the view that Frankfurt's approach is *best* seen as an account of agency: I show that it is *more contextually sensitive* to interpret his view as such. As we saw in the introduction, philosophers who are interested in free agency and autonomy often take Frankfurt's work out of context. In this chapter, I propose a contextual analysis of Frankfurt's most influential paper, his "Freedom of the Will and the Concept of a Person" (1971). I show, through that contextual analysis, that Frankfurt was concerned about issues within the philosophy of action and that his hierarchical framework is a response to them.

This first chapter is divided into two sections. In the first section, I provide an exposition of the context in which Frankfurt's 1971 paper should be located. That context is a debate between Moore, Chisholm, and Frankfurt. I explain that Chisholm rejected Moore's compatibilist account of the ability to do otherwise, and that his rejection led him to claim that moral responsibility requires, not so much the ability to do otherwise, but the ability to *decide* otherwise. Then, I show that Frankfurt's 1969 rejection of the principle of alternate possibility fails to repudiate Chisholm's view: it fails to show that moral responsibility does not require the freedom to decide otherwise.

In the second section, I argue that Frankfurt's 1971 paper can be seen as an attempt to fill in that gap. I argue that Frankfurt wants to show, by means of his example of the willing addict,

that moral responsibility does not require the freedom to decide otherwise. But I also explain that before doing this, Frankfurt revised Chisholm's account of decision, which is a substance-causal account. More specifically, I argue that Frankfurt proposed an event-causal account of decision based on the notion of second-order desires. Thus, I show that Frankfurt's hierarchical framework was developed to provide an account of decision, which is part of a larger account of agency.

1.1. The context

In this first section, I provide an exposition of the context in which Frankfurt's 1971 paper should be located. The debate is a prominent one in the free will literature, and it concerns the notion that a person "could have done otherwise." One of the most influential accounts of the notion is Moore's conditional analysis, so I shall start with his view.

Moore has proposed a conditional analysis of the ability to do otherwise.[45] In short, Moore suggests that when we say that a person "could have done otherwise," what we mean is that the person "would have done otherwise *if* she had willed or chosen to do so." Whether Moore's conditional analysis really captures the meaning of the expression is a difficult question—one that has been extensively debated. But we shall leave semantical issues aside here.

Moore's view has been seen as quite attractive because it appears to provide a compatibilist analysis of freedom understood as the ability to do otherwise. Indeed, if we follow

[45] See Moore 1911, chap. 6.

Moore, it seems that freedom is compatible with determinism. Imagine that someone's action is completely determined, that it is completely determined by the mental situation in which she is, by the specific desires and beliefs that she has at a certain time, and by the choice she has made. Now we might nevertheless say that if she had had different desires and beliefs and if she had made a different decision, she would have done otherwise. Thus, on Moore's reading, we might say that even if her action is completely determined, the person nevertheless has the ability to do otherwise.

Moreover, Moore's analysis has some important consequences for our understanding of moral responsibility. If Moore's analysis is correct, then it seems that there is no incompatibility between moral responsibility and determinism. In other words, even if we live in a fully determined world it would still be true to say of an agent that she could have done otherwise, that is, that she has a power to do or not do something. And since this power is often seen as a necessary condition for moral responsibility, it would follow that she can be morally responsible for her actions.

Moore's analysis has attracted a lot of attention and many philosophers have criticized his view,[46] including Chisholm.[47] Chisholm points out the importance of thinking about the nature of agency for any analysis of the conditions for moral responsibility. In particular, Chisholm defends volitionism, the view that all actions entail an act of will (a decision or a choice) by which an agent "sets out to do" something. The act of running, for example, involves the decision to run and the bodily motion that accompanies it. Chisholm uses a pair of medieval

[46] For a discussion of Moore's view, see in particular Austin 1956, Lehrer 1966, and Aune 1967.
[47] Chisholm 1964, pp. 14-16.

terms to present his view.[48] On Chisholm's account, an act of will is an *actus elicitus*, while a corporeal action is an *actus imperatus*. On his view, an *actus imperatus* necessarily involves an *actus elicitus*.

To defend the volitional account of agency, philosophers often appeal to the case of the paralyzed person. Consider what a paralyzed person is doing when she tries to raise her arm. What she is doing is often qualified as a mental act, an act of will, or an endeavour to move her body.[49] For an able bodied person, the bodily movement in question might follow. But this bodily movement need not be present. Even without the relevant bodily movement, an act of will still counts as an action. On the other hand, a bodily movement does not count as an action unless it is preceded by this act of will.

The volitional account is also subject to some difficulties; one of the most pressing is the regress objection. This objection has been well-known since Ryle,[50] but Hobbes[51] also formulated it a few centuries before. The objection goes more or less as follows. If an action is defined as a bodily motion that is caused by an act of will, and if an act of will is itself an action, then it must involve another act of will. But this other act of will must also involve another act of will, and this goes on *ad infinitum*. We will discuss the regress objection in much more detail in Chapter 6, so for now we may leave it aside.

As we just saw, Chisholm points out the importance of thinking about the nature of agency in our analysis of the ability to do otherwise. The question now is: what consequences

---

[48] Aquinas, for example, drew a distinction between the acts that the will *elicits* and the acts that the will *commands* (Aquinas, ST II 1, q. 8.). Those acts that the will elicits are mental acts such as choices or decisions. They are directly produced by the will. On the other hand, those acts that the will commands include, for example, walking. They are indirectly produced by the will, in conjunction with other powers such as the motive power.
[49] That is why volitionism is sometimes called the "trying" view of action, that is, the view that all actions necessarily involve a trying. On that, see Proust 2010, pp. 209-217.
[50] See Ryle 1949.
[51] See Hobbes 1640.

does Chisholm's volitional account of agency have for Moore's conditional analysis? Remember that for Moore, the statement that a person "could have done otherwise" means that she "would have done otherwise if she had chosen or willed to do so." Now Chisholm points to a contradiction that Moore's analysis generates. The statement that a person "would have done otherwise if she had chosen or willed to do so" is compatible with the deterministic claim that she "could not have chosen or willed otherwise." But if she could not have chosen or willed otherwise, then, according to volitionism, she could not have done otherwise. This contradicts the notion that we were proposing to analyze in the first place. Thus, it seems like Moore's conditional analysis of the claim that a person "could have done otherwise" is flawed. For that reason, Chisholm argues that we should reject Moore's view.

This leads Chisholm to develop his own account of the conditions for moral responsibility. What is crucial for moral responsibility, on Chisholm's view, is the freedom to will or to decide otherwise, what he calls "freedom of the man." Freedom of the man is the freedom to perform or not perform an *actus elicitus*. As Chisholm claims, the *actus elicitus* is at the center of the metaphysical problem of freedom and the question of moral responsibility: "the metaphysical problem of freedom does not concern the *actus imperatus*; it does not concern the question whether we are free to accomplish whatever it is that we will or set out to do; it concerns the *actus elicitus*, the question whether we are free to will or set out to do those things that we do will or set out to do."[52]

This is the point at which Frankfurt comes into the debate. Frankfurt is well-known for his critique of the principle of alternate possibilities, a principle on which Chisholm's account of

[52] Chisholm 1964, p. 32.

moral responsibility rests. But we need to take a closer look at Frankfurt's argument and better understand how it repudiates Chisholm's view.

In his paper "Alternate Possibilities and Moral Responsibility" (1969), Frankfurt famously repudiated what he calls the principle of alternate possibilities, namely, the view that moral responsibility requires that a person "could have done otherwise." To show that the principle of alternate possibilities is false, Frankfurt develops a few counterexamples—which are now known as "Frankfurt-style examples."[53] One of them is of particular interest here, namely, the case of Jones_4. Frankfurt imagines that Jones_4 does not have alternate possibilities because Black stands in a specific relation to him:

> Suppose someone – Black, let us say – wants Jones_4 to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones_4 is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones_4 is going to decide to do something other than what he wants him to do. If it does become clear that Jones_4 is going to decide to do something else, Black takes effective steps to ensure that Jones_4 decides to do, and that he does do, what he wants him to do. Whatever Jones_4's initial preferences and inclinations, then, Black will have his way.[54]

Now Frankfurt asks us to further imagine that Jones_4 in fact decides and does exactly what Black wants him to decide to do and to do, but for his own reasons. In fact, he does not even know that Black exists, that he controls him in the way described above, and that he limits his possibilities. Frankfurt concludes from this that Jones_4 is morally responsible for what he does, even though he does not have alternate possibilities.

One important point in Frankfurt's example is that Black controls what Jones_4 *decides to do*. Indeed, in his example, Black can tell what Jones_4 will decide to do before Jones_4 makes the

---

[53] These examples are in fact a reinterpretation of a Lockean argument. For this, see Locke's *An Essay Concerning Human Understanding*, Chap. XXI, Section 10.
[54] Frankfurt 1969, p. 6.

decision, and he can intervene and force $Jones_4$ to make the decision he wants him to make. Thus, $Jones_4$ does not have the freedom to *decide otherwise*. This is a crucial point to repudiate a view such as Chisholm's, because for Chisholm the crucial kind of the freedom is precisely the freedom to decide otherwise.

But it is also concerning this element that Frankfurt's counterexample reveals a weakness. As Widerker claims, "Frankfurt's attack on the principle of alternative possibilities does not work for decisions."[55] According to many,[56] Widerker's critique is one of the most influential early critiques of Frankfurt, one that is also often attributed to Ginet and Kane.[57] The critique can be expressed as a dilemma concerning the connection between whatever indicates that $Jones_4$ will make a certain decision and $Jones_4$'s decision. In Frankfurt's example, Black knows what $Jones_4$ will decide to do before $Jones_4$ makes the decision. Thus, there must be an indicator that reveals what $Jones_4$ will decide. The first horn of the dilemma is that there is a deterministic relation between the indicator and the decision. The second horn is that there is an indeterministic relation between the indicator and the decision. It is argued that both cases are problematic, and thus, that Frankfurt's critique of the principle of alternate possibilities has failed to show that moral responsibility does not require the ability to *decide* otherwise.

On the first horn, there is a deterministic relation between the indicator and the decision. There are a few problems with this option. First, it seems to assume determinism, which is not something that libertarians would concede to Frankfurt. Another problem is that the presence of Black does not add anything to the case, since $Jones_4$ lacks the ability to decide otherwise whether Black is present or not. Thus, Black's presence seems to be superfluous. The second

---

[55] Widerker 1995, p. 253.
[56] See Talbert 2019 and O'Connor and Franklin 2020.
[57] Ginet 1996 and Kane 1996.

horn consists in claiming that there is an indeterministic relation between the indicator and the decision. Suppose that everything indicates that Jones$_4$ will decide to run, and that this is exactly what Black wants Jones$_4$ to do. In this case, Black will not intervene. But if there is an indeterministic relation between the indicator and Jones$_4$'s decision, then Jones$_4$ will preserve the ability to decide otherwise: he could decide *not* to run at the last second, even if nothing indicates that he will decide *not* to run. Thus, the case does not show that Jones$_4$ lacks the ability to decide otherwise. Since each of the two horns generates serious issues, it is argued that Frankfurt's case has failed to show that moral responsibility does not requires the ability to *decide* otherwise.

It seems improbable that Frankfurt thought about that critique, exactly in these terms, in the two years separating his two most influential papers ("Alternate Possibilities and Moral Responsibility" [1969] and "Freedom of the Will and the Concept of a Person" [1971]). What is probable, however, is that Frankfurt thought that there was something wrong with his critique of the principle of alternate possibilities when applied to decisions. In other words, it seems plausible to suppose that Frankfurt thought, just like Widerker and many others, that his critique of the principle of alternate possibilities failed for decisions. Consequently, it is also plausible to suppose that he attempted to fix the issue and to develop a stronger case against the requirement of freedom to decide otherwise.

In the next section, I shall argue that Frankfurt did two things in his 1971 paper to develop a stronger critique of the view that moral responsibility requires the freedom to decide otherwise. First, he developed an account of decision, and of agency more generally, which was lacking in his previous paper. Second, he developed another kind of case, one that does not involve a mind-reading intervener, to show that moral responsibility does not require the ability to *decide* otherwise: the case of the willing addict. One point of clarification is in order here. I

shall not attempt to show how the case of the willing addict can solve the issue that Widerker (and others) have highlighted. That would be beyond the scope of the chapter. The aim of the next section is rather to show that Frankfurt's hierarchical approach was introduced to provide an account of decision, or more generally an account of agency.

1.2. Frankfurt's 1971 critique of the principle of alternate possibilities

In the last section, we saw that Frankfurt's 1969 argument seems to fail to repudiate Chisholm's account of moral responsibility, the view that moral responsibility requires the freedom to decide otherwise. In this section, I argue that Frankfurt's 1971 paper can be seen as an attempt to fill that lacuna. In particular, I show that Frankfurt argues, by means of his example of the willing addict, that moral responsibility does not require the freedom to decide otherwise. That point, however, only becomes clear when we understand why and how Frankfurt has revised Chisholm's account of decision. I explain that he has revised Chisholm's account of decision to avoid what I call here "the problem of substance-causation,"[58] and I explain that, in order to avoid that issue, he has developed an event-causal account of decision based on the notion of second-order desires.

1.2.1. Chisholm and the problem of substance causation

[58] The substance causal approach will be discussed in more detail in Chapters 4 and 6.

As we saw in the first section, Chisholm develops a volitional account of agency. On that view, a bodily action (an *actus imperatus*) necessarily involves an act of will or a decision (an *actus elicitus*) by which the agent sets out to perform the bodily action in question. Take, for example, the action of running. On Chisholm's view, the action of running necessarily includes the decision to run. A person cannot perform the act of running without making the decision to run.

Chisholm's account of agency has also been characterized as an agent-causal or substance-causal one. The idea here is that ultimately the cause of any actions is the agent, conceived of as an irreducible substance: "We should say that at least one of the events that are involved in the act is caused, not by other events, but by something else instead. And this something else can only be the agent."[59] What Chisholm suggests here is that an *actus elicitus*, or a decision, is directly caused by the agent and that this agent is not reducible to some event.

In Chapters 4 and 6, I will take a closer look at the substance-causal approach, but for now some preliminary considerations will suffice. At first sight, Chisholm's view might appear plausible since we use substance-causal expressions in our common talk about actions. Consider this example: "John killed the Queen." A common way to rephrase this statement is to say that "John" caused "the killing of the Queen." When formulated as such, we have an instance of substance-causation: John, an agent, is said to cause an event, the killing of the Queen.

When we take a closer look at substance-causal statements, however, they appear to be metaphorical ways of talking. Many philosophers would insist that it is not John as such that caused the killing of the Queen, but something happening to John (an event). To make the point clear, consider the following case: a bomb caused the collapse of the bridge. We have here a

---

[59] Chisholm 1964, p. 28.

substance-causal statement. A substance, the bomb, is said to cause the collapse of the bridge, an event. It seems, however, that we should not take this statement literally. Indeed, it seems like the bomb itself did not cause the collapse of the bridge. What caused the collapse of the bridge is something happening to the bomb, namely, the explosion of the bomb (an event). Thus, many philosophers would insist that we have to rephrase substance-causal statements in event-causal terms, and that includes agential statements. When we say that "John killed the Queen," what we really mean, on that view, is that something happening to John (an event such as John's desiring or intending to kill the Queen) caused the killing of the Queen.

For that reason, Chisholm's substance causal view has been regarded with general suspicion and many thinkers have rejected it. In the next section, I shall argue that Frankfurt develops an event-causal version of Chisholm's volitional account of agency to avoid the problem of substance-causation.

1.2.2. Frankfurt's volitional account of agency

In this section, I argue that Frankfurt has revised Chisholm's volitional account of agency and that in order to do so he has developed two expressions that roughly correspond to Chisholm's *actus imperatus* and *actus elicitus*. These two notions are, respectively, a "doing that one wants to do" and a "willing that one wants to will."[60] Because these two notions are event-causal notions, they allow Frankfurt to avoid the problem of substance causation.

[60] Frankfurt 1971, p. 20.

Let us first examine some more basic concepts that Frankfurt put forward in his 1971 paper. Frankfurt distinguishes first and second-order desires. First-order desires are motivations to perform certain actions such as running, walking, and talking, while second-order desires are motivations to have (or not to have) a certain first-order desire. For example, if someone desires to run, then she has a first-order desire. If she has a desire to desire to run (she would like to have that kind of motivation), then she has a second-order desire.

Frankfurt also identifies a specific kind of first-order desire, namely the will or first-order volitions. The will is an effective first-order desire, that is, a desire "that moves (or will or would move) a person all the way to action."[61] For example, if a person is running we can say that her first-order desire to run is her will because the desire to run is effective in moving her to actually run. Similarly, Frankfurt also identifies a specific kind of second-order desire, namely, second-order volitions. Second-order volitions have a specific object: they are a desire to have a certain desire as one's will. For example, if a person not only desires to have a desire to run, but also desires that such first-order desire moves her all the way to action (in order words, she wants this first-order desire to be her will), then we can say that she has a second-order volition. With these concepts in mind, we can now move on to Frankfurt's revision of Chisholm's volitional account of agency.

Frankfurt uses a first expression that is roughly equivalent to Chisholm's *actus imperatus*, that of a "doing that one wants to do." My suggestion here is that we should see this "doing that one wants to do" as designating, roughly, a bodily action. It is common to see voluntariness as a necessary condition for bodily action: a bodily event is commonly seen as a bodily action when it is wanted, desired, intended or willed. Thus, a bodily doing that is wanted

---

[61] Frankfurt 1971, p. 14.

or willed by the agent just is a bodily action. This is in fact a widespread view in the history of the philosophy of action.[62]

Notice that this is merely a *rough* equivalence. There are two reasons why it is a rough equivalence. First, it is generally agreed that voluntariness is insufficient for agency: there must be, in addition, some kind of link between the wanting and the doing. For example, according to the causal theory of action, that link has to be a causal one: a bodily event is a bodily action if it is the object of a desire or an intention, but also if it is *caused* by that desire or intention. Thus, a complete account of agency needs to clarify the relation between the wanting and the doing in question. In his 1971 paper, Frankfurt does not explicitly stipulate what is the required link between the bodily doing and the wanting,[63] but he suggests that this link is a causal one.[64] Notice that Frankfurt rejected the event-causal approach in a later paper, his "The Problem of Action" (1978). We shall discuss this issue in Chapter 4.

There is a second reason why the expression "doing what one wants to do" and "*actus imperatus*" are roughly equivalent: this is because an *actus imperatus* necessarily involves an *actus elicits* (an act of will or a decision). This is the main idea on which the volitional account is based. And, as we shall see shortly, Frankfurt's notion of an *actus elicitus* involves the second level of desires. Thus, an *actus imperatus*, on Frankfurt's view, must also involve second-order desires. To make all of that clearer, let us take a look at Frankfurt's account of an *actus elicitus*.

---

[62] See, for example, Davis 2010, p. 32.
[63] Frankfurt even claims that "both the doing and the wanting, *and the appropriate relation between them* as well, require elucidation." (Frankfurt 1971, p. 20, my emphasis)
[64] Frankfurt talks about the "overdetermination" of a person's first-order desire (Frankfurt 1971, p. 25), which suggests that a first-order desire can be caused not only by, say, a physical addiction but also by a second-order desire. And since what is going on at the second level is analogous to what is going on at the first level, we might assume that the relevant link between a bodily doing and a first-order volition is a causal one as well.

My second suggestion is that Frankfurt's expression of a "willing that one wants to will" is roughly equivalent to an *actus elicitus*. As we saw, voluntariness is often seen as a condition for agency. Accordingly, a "willing that one wants to will" simply designates an "active willing" or an "act of will." If we examine some other passages of the text, we can understand this notion of an act of will a bit more. Elsewhere in his 1971 text, Frankfurt says that a person can "make a first-order desire his will" or that he can "constitute his will" in a certain way.[65] The idea here is that we can constitute our will by means of a second-order volition: with a second-order volition, we can add to the strength of our first-order desire and make it strong enough to become effective. But notice that this is a kind of mental act whereby we do something to our will.

This mental act of will is what Frankfurt also calls an *act of identification*, which he defines, elsewhere, as a *decision*.[66] And Frankfurt also claims, consistently, that a decision is a certain kind of mental act whereby a person "makes up her mind."[67] Thus, a decision is for him an act whereby a person does something to her will, an act whereby a person constitutes her will the way she wants it to be.

The suggestion here is that "willing what one wants to will" roughly designates a *decision*. Again, this is a rough equivalence. As we saw, it is generally agreed that voluntariness is insufficient for agency. There must be, in addition, some link between the second-order volition and the mental event in question. As we saw a few paragraphs before, Frankfurt seems to endorse the event-causal framework in his 1971 paper. Thus, we may stipulate that, for Frankfurt, a decision is a mental event that is *caused* by a second-order volition.[68]

---

[65] Frankfurt 1971, p. 24.
[66] Frankfurt 1976.
[67] Frankfurt 1987.
[68] Again, Frankfurt is not explicit about that point. The main evidence that suggests that view is Frankfurt's discussion of the overdetermination of a first-order desire (See Frankfurt 1971, p. 25).

We just saw that Frankfurt employs two expressions that are roughly equivalent to the notions of an *actus imperatus* and an *actus elicitus*, and thus that he proposes a new version of the volitional account of agency. On that view, every *actus imperatus* necessarily involves an *actus elicitus*. If my interpretation is correct, it means that for Frankfurt all bodily actions involve an act of identification or a decision. That position is in fact clearly expressed by Frankfurt when he claims that "a person is active with respect to what he does when what he does is the outcome of his identification of himself with the desire that moves him in doing it."[69] In that quote, Frankfurt claims that a person is active with respect to what he does (i.e. he is acting) when his doing is caused by an act of identification. Thus, on his view, all bodily actions involve an act of identification or a decision.

We should notice here how Frankfurt's volitional account of agency avoids the problem of substance causation, a problem that affects Chisholm's view. According to Frankfurt, a decision is a mental event that is caused by a second-order volition (another event) and not by the agent herself (an irreducible substance). One might say that, for Frankfurt, second-order volitions *replace the agent* or *play the role of the agent*.[70] Thus, Frankfurt's account of decision is based on the event-causal approach rather than the substance-causal approach. This, I would suggest, is one of Frankfurt's main contribution in his 1971 paper. In other words, what is original about Frankfurt's 1971 view is his event-causal account of decisions, which is based on the concept of second-order volition.

---

[69] Frankfurt 1975, p. 54.
[70] For a similar suggestion, see Velleman 1992. I discuss this idea in more detail in Chapters 5 and 6.

1.2.3. Frankfurt's dual account of freedom

In this section, I discuss Frankfurt's dual account of freedom, which comprises freedom of action and freedom of the will. I argue that, for him, freedom of action is the freedom to perform a bodily action (an *actus imperatus*) and freedom of the will is the freedom to perform a decision (an *actus elicitus*). Moreover, I argue that Frankfurt's concept of freedom merely designates the notion of alternate-possibilities freedom, and thus that it is far from original. What is original about Frankfurt's view is, as we just saw, his account of agency or his interpretation of the notions of an *actus imperatus* and an *actus elicitus*.

Let us start with freedom of action. Frankfurt claims that "freedom of action is (roughly, at least) the freedom to do what one wants to do."[71] As we saw in the previous section, "to do what one wants to do" designates, roughly, a bodily action or an *actus imperatus*. Thus, we can say that, for Frankfurt, freedom of action is the freedom to perform a bodily action or an *actus imperatus*. Frankfurt claims that freedom of the will is analogous to freedom of action, and he defines freedom of the will as the freedom to will what one wants to will.[72] As we also saw in the previous section, "to will what one wants to will" designates, roughly, a decision or an *actus elicitus*. Freedom of the will is thus, for Frankfurt, the freedom to perform a decision or an *actus elicitus*.

From the previous discussion, we do not know much about the concept of freedom itself, since Frankfurt's definitions are always partly circular: he defines both freedom of action and freedom of the will as the *freedom* to perform a certain act. In another passage, however,

---

[71] Frankfurt 1971, p. 20.
[72] See Frankfurt 1971, p. 20

Frankfurt makes clear that freedom means for him "alternate possibilities." Let us take a closer look at the passage in question. Frankfurt claims that

> (1) A person's will is free only if he is free to have the will he wants. (2) This means that, with regard to any of his first-order desires, he is free either to make that desire his will or to make some other first-order desire his will instead. (3) Whatever his will, then, the will of the person whose will is free could have been otherwise; he could have done otherwise than to constitute his will as he did.[73]

In the first sentence, Frankfurt reiterates his definition of freedom of the will, the definition that we just examined. But, instead of using the expression "to will what one wants to will," he uses the expression "to have the will one wants" which is equivalent. In the second and third sentences, Frankfurt explains what his definition means. In these sentences, it becomes clear that freedom means, for him, *alternate possibilities*.[74] Indeed, he claims that the person who has freedom of the will can "make that desire his will or… make some other first-order desire his will instead." Thus, the person who has freedom of the will has alternate possibilities with respect to the constitution of his will: he can constitute his will in one way or another. This is confirmed in the third sentence where Frankfurt says that the person who has freedom of the will "could have done otherwise than to constitute his will as he did."

Since the act whereby a person constitutes her will is a decision, we can conclude that freedom of the will is, for Frankfurt, the *freedom to decide otherwise*. And since freedom of action is analogous to freedom of the will, freedom of action is just the *freedom to do otherwise*. Thus, Frankfurt's dual account of freedom is just an account of the freedom to do and decide

---

[73] Frankfurt 1971, p. 24.
[74] It appears, therefore, that Frankfurt does not depart importantly from the traditional notion of free will, what Fischer calls "alternative-possibilities freedom" (Fischer 2005, xxviii).

otherwise. As such, his notion of freedom is far from original. What is original, again, is the way he develops the related notions of "doing" and "deciding."

Before moving on to the next section, we should notice here that Frankfurt's dual account of freedom perfectly echoes Chisholm's own view. Indeed, Chisholm draws a distinction between "freedom of the man," which is the freedom to decide otherwise, and "freedom of the will," which is the freedom to do otherwise.[75] The main difference between Frankfurt's and Chisholm's views does not concern the concept of freedom itself, but, as I have reiterated, the way they account for the doing (*actus imperatus*) and the deciding (*actus elicitus*). While Chisholm's account of agency is a substance-causal one, Frankfurt proposes an event-causal account.

1.2.4. The case of the willing addict

In the previous section, we saw that freedom of the will is for Frankfurt the freedom to decide otherwise. In this section, I argue that Frankfurt's main argument in his 1971 paper, and his second major contribution, is to show that moral responsibility does not require the freedom to decide otherwise. This is a repudiation of Chisholm's account of moral responsibility, for whom moral responsibility requires this kind of freedom.

To show that moral responsibility does not require the freedom to decide otherwise, Frankfurt presents the case of the willing addict. The willing addict wants to take drugs and, since he is an addict, his desire to take drugs is compulsive or irresistible. But the willing addict

[75] Chisholm 1964, p. 32.

also wants to be an addict: he wants his desire to take drugs to be effective and to lead to action. Thus, the willing addict's second-order volition is in harmony with his first-order volition to take drugs.

If what I have said is correct, then we can say that the willing addict *decides* in favour of his desire to take drugs, that is, he makes his desire to take drugs his will or he constitutes his will in such a way as to end up taking drugs. But the willing addict does not have the freedom to decide otherwise. And the reason for that is because he could not have decided otherwise or constituted his will otherwise. Since his desire to take drugs is irresistible, he could not have made any other desire his will. This is clearly expressed by Frankfurt when he says that "[t]he willing addict's will is not free, for his desire to take drug will be effective regardless of whether or not he wants this desire to constitute his will."[76]

What should we think about the moral responsibility of the willing addict now? Frankfurt argues that he may be morally responsible for what he does, even if he does not have the freedom to decide otherwise. The willing addict may make the decision in question for his own reasons (and not because he is addicted). The fact that he is addicted may play no role in his behaviour. This is especially true if he does not know that he is addicted. Frankfurt claims that, for that reason, we tend to see him as morally responsible for his behaviour. Thus, with his example of the willing addict, Frankfurt wants to show that moral responsibility does not require the freedom to decide otherwise, and thus, he wants to repudiate Chisholm's account of moral responsibility.

---

[76] Frankfurt 1971, pp. 24-25.

1.3. Conclusion

In this chapter, we saw that Frankfurt argues, in his 1971 paper, that moral responsibility does not require the freedom to decide otherwise, and that this argument is a repudiation of Chisholm's account of moral responsibility. We also saw that, in order to make his argument, Frankfurt proposed a revision of Chisholm's volitional account of agency and, in particular, of his concept of an *actus elicitus* (or his concept of a decision). Frankfurt's original proposition is that a decision is a kind of mental act that involves a second-order desire, an act whereby a person constitutes her will in a certain way.

Having this understanding of the debate between Chisholm and Frankfurt, we can now see why Frankfurt's hierarchical approach can be seen as a theory of agency: because it is an attempt to provide an event-causal account of Chisholm's distinction between an *actus imperatus* and an *actus elicitus*. Moreover, this interpretation appears to be superior to the alternative interpretations that view Frankfurt's hierarchical approach as an account of free agency or autonomy because it is contextually sensitive: it locates Frankfurt's development of his two-level framework in a debate that dominated the literature at the time.

Chapter 2: Hierarchy as Free Agency and Autonomy

In the previous chapter, we saw that Frankfurt's hierarchical approach can be interpreted as a volitional account of agency. I also provided a first reason to favour my reading of Frankfurt over the main alternatives: I showed that it is more contextually sensitive. In the present chapter, I discuss in more detail the main alternatives to my reading of Frankfurt, which are the views that his hierarchical approach is an account of free agency[77] or autonomy.[78] I provide a second reason to favour my reading over these interpretations: I argue that my interpretation is *more precise* than the main alternatives.

To begin with, I concede that it is textually justified to interpret Frankfurt's two-level approach as an account of free agency or autonomy. But I also show that Frankfurt's account of free agency and autonomy is based on the idea that a constraint is a *passive happening*, and that heteronomy is a kind of mental passivity, where I explain that this is not an intuitive view. I deduce from this that Frankfurt's view is prone to misunderstandings and that a reformulation of his approach in agential terms is justified: this reformulation eliminates confusions and provides a clearer grasp of his hierarchical approach. Thus, I show that it is more precise to interpret his view as an account of agency, which is the second reason to favour my reading.

This second chapter is divided into three sections, with each section corresponding to one of the main interpretations of Frankfurt's approach. In the first section, I discuss Locke's and

---

[77] See, for example, Locke 1975, Watson 1975, Thalberg 1978, and Benson 1994.
[78] See, for example, Friedman 1986, Christman 1989, Oshana 1998, Mackenzie and Stoljar 2000, J. S. Taylor 2005, and Mackenzie 2014.

Thalberg's interpretation: the view that Frankfurt's two-level approach is a single account of free agency. I reject that first interpretation because it lacks consistency.

In the second section, I discuss the view that Frankfurt's approach is a dual account of free agency. I argue that this interpretation is correct because it is supported by textual evidence. But I also show that Frankfurt equates free agency with agency (*tout court*), and I explain that this is so because Frankfurt's account of free agency is based on the idea that a constraint is a passive happening. I conclude that his view is prone to misunderstandings and that a reformulation of his approach in agential terms eliminates ambiguities and allows us to gain a sharper understanding of his view.

In the third section, I discuss the view that Frankfurt's hierarchical approach is an account of autonomy. I concede that this interpretation is accurate because it is supported by textual evidence. But I also argue that Frankfurt equates autonomy with a form of mental agency, just as he equates free agency with agency. I conclude that his account of autonomy is also prone to confusions, and that a reformulation of his view in agential terms is justified because it eliminates these confusions and provides a more perspicuous understanding of his hierarchical approach.

2.1. The Locke-Thalberg interpretation

In this first section, I examine the view that Frankfurt's hierarchical approach is a single account of free agency, a view that was put forward by D. Locke and Thalberg. I argue that we should reject that first interpretation because it lacks consistency.

The first interpretation with which we shall begin is based on the idea that Frankfurt draws a distinction between "acting" and "acting freely." On that view, "acting" consists in "doing what one wants to do" while "acting freely" involves, in addition, a second level: acting freely occurs when a person acts, and when the will that moves her into action is the will that she wants to have. On this view, the "free" aspect comes from our higher-order identification with the lower level.

This view was suggested by G. Dworkin in an early paper, his "Acting Freely" (1970). It is also a view that Locke and Thalberg attribute to Frankfurt. In "Three Concepts of Free Action," Locke examines a first concept of free action, the view that "to act freely is to act as you want to."[79] An obvious issue with that view, as Locke points out, is that all "intentional actions are willing [and hence free] actions." For example, "the behaviour of the bridegroom threatened with the shotgun… will to the extent that it is intentional, be willing and hence free."[80] Locke suggests that to avoid this difficulty, Dworkin and Frankfurt claimed that a free action is not simply a doing that we want to do, but also involve a desire that we want to have:

> Thus (following Dworkin, "Acting Freely") we might say that a man acts willingly, and so freely, when he acts from wants he wants to have, or reasons he does not mind having. More formally (following Frankfurt, "Freedom of the Will"), we might distinguish between first- and second-level wants or desires. A man who has a want without wanting that want, *e.g.*, a smoker trying to kill his desire for cigarettes, has the first-level want but not the second-level want; a man who wants a want he does not have, *i.e.*, wishes he felt some desire he does not feel, has the second-level want but not the first-level want. And the suggestion is that a man really acts as he wants to, acts willingly [and hence freely], when he has both the first- and the second-level wants.[81]

---

[79] Locke 1975, p. 96.
[80] Locke 1975, p. 97.
[81] Locke 1975, pp. 97-98.

Likewise, Thalberg believes that Dworkin's and Frankfurt's approaches are similar with respect to the distinction between acting and acting freely. On his view, they both use the second order as a way to account for the "free" aspect of "acting freely." In the following quote, Thalberg suggests that constraints happen, for both Dworkin and Frankfurt, at the second-level: "Dworkin and Frankfurt… suppose that what a constrained person 'doesn't want' is for some desire or other to move him."[82]

A follower of the Locker-Thalberg view might object to my interpretation and argue that her interpretation allows us to account for *wanton actions*, while my interpretation fails to do so. To support her view, she might point out that, for Frankfurt, wantons are those creatures that do not form second-order volitions, or those creatures that do not identify with their first-order desires.[83] For example, children and non-human animals are wantons since they do not care about which motive drives them to act. Adult humans can also be wantons, since not everyone always cares about which motive drives her to act. The difficulty with the case of the wanton, then, is that a wanton can act even though she does not form second-order volitions. As Frankfurt says, "humans may be more or less wanton; they may *act wantonly*, in response to first-order desires."[84] The problem for my interpretation is that, since a wanton does not form second-order volitions, it seems like I have to conclude that a wanton cannot act, since I argued that acting requires a second-order volition. Our objector might conclude that my interpretation is implausible for that reason.

Moreover, our objector might argue that the Locke-Thalberg view does not run into the same difficulty. On the Locke-Thalberg view, a wanton action is just a normal action: it consists

---

[82] Thalberg 1978, p. 216.
[83] Frankfurt 1971, p. 16.
[84] Frankfurt 1971, p. 17 (my emphasis).

in doing what one wants to do. What the wanton cannot do, on that view, is act *freely* because the freedom in question comes from our higher-order identification with the lower level. So, our objector might argue that the Locke-Thalberg interpretation has one strong advantage over my interpretation, which is that it can account for wanton actions.

That objection, however, is not decisive: I can provide a response which shows that the case of the wanton is not a real issue for my interpretation. It is true that wantons can act on Frankfurt's view: a wanton can act "in response to first-order desires."[85] So a wanton action is just a "doing that one wants to do." That being said, what I should have said earlier is that, *when a person acts*, a decision (and thus a second-order volition) is required.[86] The distinction, then, is a distinction between the action *of a wanton* and the action *of a person*. Notice that this is a distinction between two grades of agency *qua agency*, and not a distinction between "acting" and "acting freely." This distinction between two grades of agency (*qua* agency) will be further discussed in Chapter 5 (Section 5.3.2). For now, the important point is simply that we can account for wanton actions when we interpret Frankfurt's hierarchical approach as a theory of agency (*tout court*).

But then, our objector might reply that Frankfurt's account of agency is not a genuine volitional account of agency, since agency does not always require a decision. That might be so, but two points are in order here. The first is that Frankfurt was interested in the kind of actions for which we are morally responsible: what interested him is what it is *for a person* to act and not what it is for a wanton to act (wantons such as non-human animals and young children are not

---

[85] Frankfurt 1971, p. 17 (my emphasis).
[86] This is particularly clear when Frankfurt says that "*a person* is active with respect to what he does when what he does is the outcome of his identification of himself with the desire that moves him in doing it" (Frankfurt 1975, p. 54, my emphasis). Here, it is quite clear that he is concerned by what it is *for a person* to act. And he claims that when a person acts, she must perform an act of identification (a decision).

typically held to be responsible for their actions). In that context, it makes sense to say that agency requires a decision, since we are concerned by the way persons act.

The second point is that Frankfurt seems to develop, in his later work, a truly volitional account of agency. In particular, he draws a distinction between a choice and a decision, a choice being a kind of first-order act of will and a decision being a kind of second-order act of will: "making a decision is something that we do *to ourselves*. In this respect it differs fundamentally from making a choice, the immediate object of which is not the chooser but whatever it is that he chooses."[87] The point then is that wanton actions require a choice, not a decision. Thus, Frankfurt's account seems to be, at least in his later works, truly a volitional account of agency.

So far, we know that both my interpretation and the Locke-Thalberg interpretation can account for wanton actions. But is there any reason to favour my interpretation over the Locke-Thalberg view? I shall argue here that the Locke-Thalberg interpretation should be rejected because it lacks consistency.

To understand why it lacks consistency, we need to understand that there is not one mode of free agency for Frankfurt, but two. On Frankfurt's view, a person may *act freely* and *of her own free will*. In his 1971 paper, Frankfurt says: "Suppose that a person has done what he wanted to do, that he did it because he wanted to do it, and that the will by which he was moved when he did it was his will because it was the will he wanted. Then *he did it freely and of his own free will*."[88]

This passage is a bit tricky to interpret because Frankfurt discusses the two modes of free agency together. One way to interpret this passage is the following. First, one might say that

---

[87] Frankfurt 1987, p. 172.
[88] Frankfurt 1971, p. 24 (my emphasis).

"acting freely" means "doing what one wants to do" and "doing it because one wants to do it." Here, the point is simply that acting freely requires a doing, a wanting, and a link between the doing and the wanting (presumably a causal one). Second, one might say that "acting of one's own free will" means "willing that one wants to will" and "willing it because one wants to will it." The point here is that acting of one's own free will requires a willing, a second-order volition, and a link between the willing and the second-order volition (presumably a causal one). That view is, in my opinion, almost exact. There is one small problem, and it has to do with the fact that Frankfurt talks about what it is *for a person* to act freely and of her own free will. That being said, we may leave this problem aside for now because it would unnecessarily complicate the discussion.[89]

We just saw that there is a dual account of free agency in Frankfurt's work. To defend the Locke-Thalberg interpretation, our objector might point out that what Frankfurt identified as the conditions of acting freely are just, in fact, the conditions of intentional agency. When a person "does what she wants to do," she simply *acts*. Because of that, our objector might argue that it is justified to reformulate Frankfurt's view and to claim that the first level only captures the conditions of agency, while the second level captures the conditions of some kind of freedom.

I think that that response should be rejected, because the reasoning on which it is based lacks consistency. Our objector claims that a "doing that one wants to do" is simply an action, and not a free action. But then, why should we not say that a "willing that one wants to will" is simply an *active willing* (an *act of will*), and that it has nothing to do with freedom? In other

---

[89] I take Frankfurt's notion of acting freely to be equivalent to performing a bodily action, and his notion of acting of one's own free will to be equivalent to performing a decision (or an act of will). That being said, I just argued that when *a person* performs a bodily action, she necessarily also *decides* to do it. The consequence of that position, then, is that acting freely (performing a bodily action) requires acting of one's own free will (performing a decision), and thus that the conditions of acting freely are not simply first-order conditions.

words, if our objector was consistent in her reasoning, she would see that a reformulation of Frankfurt's account of "acting of one's own free will" is required as much as a reformulation of his account of "acting freely." A consistent reasoning, then, leads to the view that the second level no more captures freedom than the first level. In other words, it leads us, roughly, to my interpretation of Frankfurt: the view that his hierarchical framework is an account of agency.

2.2. Hierarchy as a dual account of free agency

In the previous section, we discussed a first interpretation of Frankfurt's hierarchical approach, the view that it is a single account of free agency. We saw that this interpretation should be rejected because it lacks consistency: a consistent reasoner would reformulate not only Frankfurt's account of "acting freely" but also his account of "acting of one's own free will." In the present section, I defend the idea that we should reformulate Frankfurt's dual account of free agency, but in order to justify this reformulation we need to gain a better understanding of his dual account. Thus, I will start with a discussion of that view.

In the previous section, we saw that Frankfurt draws a distinction between "acting freely" and "acting of one's own free will." We also saw that "acting freely" roughly corresponds to "doing what one wants to do," and that "acting of one's own free will" roughly corresponds to "willing what one wants to will." One might wonder, at this point, where this account comes from. To understand where it comes from we need to understand the relation between freedom and free agency.

On Frankfurt's view, a person may act freely even though she does not have freedom of action, and she may act of her own free will even though she does not have freedom of the will.

As he says, "[i]t is a mistake… to believe that someone *acts freely* only when he is *free to do whatever he wants* or that he *acts of his own free will* only if his *will is free*."[90] Remember that freedom of action is for Frankfurt the freedom to do otherwise[91] and freedom of the will is the freedom to decide otherwise or the freedom to constitute one's will otherwise. Thus, Frankfurt argues here that a person may act freely even if she does not have the freedom to do otherwise, and she may act of her own free will even if she does not have the freedom to constitute her will otherwise.

Frankfurt does not really develop an argument to support his position, but we can quite easily generate one using some "Frankfurt-style examples." Let us start with acting freely. Consider the case of two runners, Luke and Jones$_4$. By hypothesis, Luke has the freedom to do otherwise. So, he has the possibility to run or to do something else instead. But he wants to run and so he ends up running. Here, we may call Luke's running a "free action" because Luke had freedom of action and he exercised his freedom when acting.

Jones$_4$, on the other hand, is controlled by Black in the way described before: Black can force Jones$_4$ to do whatever he wants him to do. Thus, Jones$_4$ does not have the freedom to do otherwise: he only has one possibility for action, which is to do what Black wants him to do. But Jones$_4$ is unaware that Black controls him in this way, and so he believes that he has many

---

[90] Frankfurt 1971, p. 24 (my emphasis).

[91] There is one small difficulty with my interpretation, and it emerges in the previous quote. In that quote, Frankfurt conceives of freedom of action as the freedom to do *whatever* one wants. Thus, Frankfurt seems to conceive freedom of action as a *multi-way* power and not simply as a *two-way* power as I have argued. In other words, freedom of action seems to require, in that quote, *unlimited* possibilities and not simply *alternate* possibilities. In a later text, Frankfurt discusses more extensively the idea that freedom is an unlimited power, or the Cartesian doctrine that the will is absolutely and perfectly free (see Frankfurt 1989 and Descartes 1641). Notice that he has proposed an amendment to that doctrine.
That being said, whether this other interpretation of Frankfurt is correct or not does not really affect the argument that I put forward in this chapter. Indeed, the point is that free agency does not require freedom as a two-way power. If that is true, then free agency certainly does not require freedom as a multi-way power. Thus, the argument is valid whether we conceive of freedom as a multi-way or a two-way power.

possibilities for action. He ends up choosing to go running. Fortunately for him, this is exactly what Black wants him to do, so $Jones_4$ ends up running.

What does this tell us about acting freely? It seems like Luke's and $Jones_4$'s actions of running *are not* significantly different. Both agents perform the action in question for their own personal reasons. In the case of $Jones_4$, what constrained him played no role in the performance of his action, since Black did not have to intervene and $Jones_4$ was not even aware that Black constrained him in the way he did. Thus, if we call Luke's action a "free action," we also have to call $Jones_4$'s action a "free action," even though he does not have freedom of action. Therefore, we can conclude that a person may act freely in spite of lacking freedom of action.

The same kind of argument can be developed to support the view that a person may act of her own free will even though she does not have the freedom to constitute her will otherwise. Consider the case of two drug-takers: the experimenter and the willing addict. The experimenter has, by hypothesis, freedom of the will. Thus, she can make her first-order desire to take drugs her will or she can make any of her other first-order desire her will. She picks out her desire to take drugs and makes it her will. Here, we may say that our experimenter is taking drugs of her own free will, because she had freedom of the will and she exercised it.

The willing addict, on the other hand, has an irresistible desire to take drugs and, for that reason, she does not have freedom of the will: she can make just one of her first-order desires her will, namely, her desire to take drugs. However, she does not know that she is an addict and she believes that she has many possibilities. Then, just like the experimenter, she picks out her desire to take drug and makes it her will.

Here again, it seems like the mental acts of the two persons are not significantly different. Both constituted their will the way they did for their own reasons. In particular, the addiction of

the willing addict played no role in her mental act since she was not aware of it. Thus, if we say that the experimenter acted of her own free will, we should also say that the willing addict acted of her own free will. Therefore, we can conclude that a person may act of her own free will even though she lacks freedom of the will.

We have here two Frankfurt-style arguments that support the view that free agency does not require freedom. But, from that discussion, we do not know much about free agency as such. We simply know that it does not require freedom (alternate-possibilities freedom). We may ask now: what does free agency require? My suggestion here is that Frankfurt's account of free agency is based on an implicit argument, which we may call "the argument from conceptual elimination." This argument is actually quite simple and goes as follow: if we remove the freedom conditions from "free agency," then the only thing left are the conditions of agency. On that view, the conditions of free agency are just the conditions of agency, which means that there is no distinction between free agency and agency (*tout court*).

Notice that, if I am right, agency (*tout court*) is more fundamental than free agency, since Frankfurt reduces the conditions of free agency to the conditions of agency, and not the other way around. In that sense, Frankfurt's approach is first and foremost an account of agency (*tout court*). Notice, as well, that we may abandon Frankfurt's view of free agency (and his argument from conceptual elimination) without abandoning his theory entirely: one might think, as I do,[92] that Frankfurt's hierarchical approach is a defensible account of agency while rejecting the view that the freedom conditions of free agency are eliminable. I will defend this position in Chapter 3.

---

[92] That does not mean, however, that I think Frankfurt's view is flawless. In Chapter 6 (see 6.3.1), I will identify two main weaknesses of his approach.

An objection might be raised at this point. One might argue that my reading is implausible simply because the view that free agency is equivalent to agency is implausible. While this might be so, we need to notice that it is not an uncommon position in the literature on agency. The view in question is often attributed to Chisholm,[93] among others.[94] Remember that, on Chisholm's view, an action is defined as an event that is ultimately caused by the agent. But the agent is, for Chisholm, a "prime mover unmoved."[95] In that sense, the agent is necessarily free, that is, free from causal determination. The result is that all actions are free actions, because they necessarily involve the agent. Thus, on Chisholm's view, there is no distinction between acting and acting freely.[96]

We just saw that, for Chisholm, free agency and agency are equivalent notions and thus that Frankfurt's view is not unprecedented. Before we discuss another reason why Frankfurt might equate free agency and agency, I would like to mention a compelling piece of evidence that supports my interpretation: Frankfurt employs his hierarchical framework to discuss *both* agency and free agency. As we saw, in "Freedom of the Will and the Concept of a Person," hierarchy is used to analyze the notions of "acting freely" and "acting of one's own free will." But, in "Identification and Externality,"[97] he also employs the hierarchical framework to discuss the notions of activity and passivity: in this paper, the notions of freedom, free agency, or autonomy are not mentioned. Rather, Frankfurt uses his concept of identification and his

---

[93] Zimmerman, for example, claims that "Chisholm appears not to have sufficiently appreciated the difference between issues (1) and (3)," that is, between "the nature of action" and "the nature of free action" (Zimmerman 2010, p. 586). See also Chisholm (1964).
[94] We often attribute that view to R. Taylor (1966) and Reid (1788) as well.
[95] Chisholm 1964, p. 32.
[96] Notice that this position only makes sense as a claim about *metaphysical freedom*, and not about *political freedom*. Chisholm's view is that all actions are ultimately undetermined, because they originate in the agent. This is a point about the absence of causal determination and thus about metaphysics. We should not understand this claim to mean that all actions are free *in the political sense*.
[97] Frankfurt 1976.

hierarchical framework to analyze what it is to *desire actively*, or to have an *active passion*. The main argument of that paper is that a desire becomes active when a person identifies with it by means of a higher-order decision. Since this argument will be discussed in more detail in Chapters 5 and 6, we may leave it aside for now.[98]

There is further evidence to support the view that Frankfurt uses his hierarchical framework to analyze *both* agency and free agency, but it would be fastidious to go through all of this evidence.[99] Rather than doing that, I want to put forward a hypothesis to further justify the view in question. The hypothesis is the following: Frankfurt's account of free agency is based on the idea that *passive happenings* are constraints. This view becomes particularly clear in Frankfurt's later work.[100]

As Frankfurt argues, there are two possible kinds of passive happenings: passive *bodily* events and passive *mental* events. Let us start with the former. Passive bodily events include things like bodily spasms and tics. One way to account for passive bodily movements is in terms of agential control: passive bodily movements are those over which the agent has no control. This will be discussed at length in Chapters 4, 5 and 6, so we need not spend more time on this idea here. The hypothesis that I want to put forward is that, for Frankfurt, passive bodily events, like bodily spasms, are a kind of constraint. Think about the epileptic who is subject to frequent

---

[98] We could simply mention here that, to reach that conclusion, Frankfurt makes a few points. First, he argues that an act of identification is in fact a *decision to include* the desire among the candidates for satisfaction. And through that decision, the desire becomes internal to the person's self. Finally, a desire that is internal to the self is just an *active* desire, because, as Aristotle claims, "a thing is active with respect to events whose moving principle is inside of it." (Frankfurt 1976, p. 59)

[99] One such piece of evidence appears in a paper devoted to the notion of free agency, "Three Concepts of Free Action." In that paper, Frankfurt claims that "a person is active with respect to his own desires when he identifies himself with them, and he is active with respect to what he does when what he does is the outcome of his identification of himself with the desire that moves him in doing it" (Frankfurt 1975, p. 54). It is striking that, in the previous quote, the concepts of freedom, autonomy, or free agency are never mentioned. Rather, what is in question is the notion of *doing something actively* and *desiring actively*. This suggests, again, that Frankfurt uses his hierarchical framework to analyze *both* agency and free agency.

[100] See, in particular, Frankfurt 1976.

bodily seizures. It seems correct to say that her bodily seizures constrain her: they might prevent her from doing certain things like driving a car. Accordingly, it seems correct to say of a person who is undergoing a passive bodily happening that she is constrained. And, conversely, a person who is performing an active bodily motion is experiencing a certain kind of freedom—at the very least, we might say that she is "freer" than the person who is seized by a bodily spasm.

The same considerations can be made concerning mental happenings. As Frankfurt points out, some mental events are passive, such as "obsessional thoughts… that run willy-nilly through our heads"[101] or compulsive desires that overcome us. These passive mental happenings are analogous to bodily spasms: they are a kind of "mental spasm." Again, one way to account for this form of passivity is in terms of agential control: one might say that passive mental events are those events over which the agent has no control. This will be discussed in Chapters 5 and 6. Now, the hypothesis that I want to put forward is that, for Frankfurt, these passive mental happenings are a kind of constraint. Think about the addict who is overcome by her desire to take drugs or the jealous wife who is overcome by the obsessional thought that her husband is cheating. It seems correct to say that they are both constrained by the mental event they are experiencing. Accordingly, it seems correct to say that a person who is undergoing a passive mental happening is constrained, while a person who is experiencing an active mental happening is freer.

Notice that this account of free agency is not based on a very intuitive conception of constraints. Thus, we should make clear here what counts as constrained and what does not count as constrained on Frankfurt's view. Consider the bodily motion of the inmate who walks back and forth in her cell. The walls of the prison limit her bodily motion, and thus, we may say that

---

[101] Frankfurt 1976, p. 59.

her bodily motion is constrained or unfree. That is our intuitive view on the matter. But, on Frankfurt's understanding, the bodily motion appears to be active, that is, it appears to be an action, and thus it would not count as unfree. Similarly, a person might be the victim of a manipulator who implanted desires into her without her consent. Here again, an intuitive way to interpret this case is to say that the person is constrained in what she wants. But, on Frankfurt's view, it is far from obvious that the victim would count as unfree because nothing indicates that her implanted desires passively happens to her or that they overcome her.

We just saw that Frankfurt equates free agency and agency, and that he does so because he conceives of constraints as passive happenings. We also saw that this is not an intuitive understanding of constraints, nor an intuitive understanding of free agency. For that reason, it appears to be justified to reformulate Frankfurt's hierarchical approach as an account of agency (*tout court*). This reformulation allows us to avoid ambiguities and misunderstandings, and to gain a more precise understanding of his hierarchical approach. Moreover, nothing is lost in this reformulation, since free agency and agency are equivalent notions for him.

2.3. Hierarchy as autonomy

In the previous section, we saw that it is textually accurate to see Frankfurt's hierarchical approach as a dual account of free agency. But we also saw that this view is based on the idea that passive happenings are constraints, and that this is not an intuitive understanding of constraints. To avoid confusion and to gain a more perspicuous grasp of Frankfurt's view, I argued that we should reformulate his approach in agential terms. In this section, I will examine our last interpretation, the view that Frankfurt's two-level approach is an account of autonomy. I

will argue that this interpretation is textually correct, but that a reformulation is required for the same reasons that a reformulation of his account of free agency is required.

Let me start with a summary of the interpretation in question, the view that Frankfurt's hierarchical approach is an account of autonomy. On that reading, the second order of desiring is seen as representing a person's real or authentic self, or as being the seat of her self-conception. For that reason, when a person has the will that she wants to have, we can say that her will is *authentic* or that it represents who she *really* is. A person who is determined by this authentic will is an autonomous person where this account of autonomy is known as the "autonomy as authenticity view." Conversely, when a person does not have the will that she wants to have, we may say that she is alienated from her will or that she is heteronomous.

This interpretation is supported by textual evidence. Frankfurt indeed uses the concept of autonomy to describe his hierarchical approach. The first time he uses the concept of autonomy is in a 1973 paper, "Coercion and Moral Responsibility." Moreover, Frankfurt says that when a person is moved by the desire that he wants to have, he "does what he *really wants*."[102] Thus, it seems like his account provides a distinction between "real" or "authentic" desires and inauthentic ones. Frankfurt also claims that some passions or desires may be "discontinuous" with a person's "conception of himself"[103] or that a person may be "helplessly violated by his own desires."[104] Thus, Frankfurt suggests that a person may be alienated from her desire or that a desire may be inauthentic.

That being said, we need to better understand Frankfurt's account of authenticity and alienation and ask: what exactly is the notion of authenticity on which Frankfurt's account is

---

[102] Frankfurt 1987, p. 166.
[103] Frankfurt 1976, p. 62.
[104] Frankfurt 1971, p. 17.

based? What is it to "*really* want to do something" or to be "alienated" from our desires? My suggestion is that Frankfurt's distinction amounts to a distinction between active and passive desires. When a person is overcome by a passive desire to take drugs, for example, we may say that she is alienated from her desire. Here, the notion of alienation refers to what we might call "agential alienation." And this is confirmed by the way Frankfurt discusses alienation: on his view, a person who is alienated from her desires is a "*passive* bystander." Conversely, when a person is not overcome by her desire in this way, we may say that she really wants to do whatever it is that she wants to do: her desire is more authentic. Thus, Frankfurt's notions of authenticity and alienation concern the distinction between activity and passivity.

Consequently, to be autonomous for Frankfurt just means to be determined by an active desire as opposed to a passive mental happening. This view is summarized by Frankfurt in a later text. In a footnote, he writes that "[a]utonomy is essentially a matter of whether we are *active* rather than *passive* in our motives and choices—whether, however we acquire them, they are the motives and choices that we *really* want and are therefore in no way *alien* to us."[105] We have here the idea that autonomy is a matter of mental activity, and the idea that this mental activity is equivalent to a form of authenticity.

Now that we know that authenticity and alienation are matters of activity and passivity, we need to make clear that the distinction is *not* an intuitive one. Consider the example a young adult who is pressured by her parents to study medicine. She may want to study medicine just to eliminate the pressure. But imagine as well that she has a deeper desire to become an artist and that she feels like this is her vocation. Now on an intuitive understanding of authenticity, we might say that her desire to become an artist is more authentic than her desire to study medicine,

---

[105] Frankfurt 2004, p. 20 (my emphasis).

which might be seen as alien to her because it is the result of parental pressure. But these are not the notions of authenticity and alienation that Frankfurt has in mind. On Frankfurt's view, there is no reason to see her desire to study medicine as alien to her because it does not appear to be passive: it is nothing like a "mental spasm." Thus, on Frankfurt's view, both desires would count as authentic because both are active.

Since Frankfurt does not propose an intuitive understanding of autonomy and authenticity, his view is prone to misunderstandings. For that reason, it is favourable to reformulate his hierarchical approach in agential terms. That reformulation eliminates ambiguities and provide a more precise understanding of his approach. On that view, the notion of autonomy refers, more precisely, to the notion of what it is for *a person* to act, which involves a decision and thus a second-order volition.

2.4. Conclusion

In this second chapter, I examined the main alternatives to my interpretation of Frankfurt. I conceded that it is textually justified to interpret Frankfurt's view as an account of free agency or autonomy. But I have also shown that Frankfurt equates free agency with agency and that he equates autonomy with a form of mental agency. I explained that this is so because Frankfurt conceives of a constraint as a passive happening. Then, I argued that this is not an intuitive understanding of constraint and thus that Frankfurt's account is prone to confusion. To eliminate confusion, I argued that we should reformulate his approach in agential terms and that we gain, thereby, a sharper understanding of his view. Thus, I have shown that my interpretation of

Frankfurt, the view that his hierarchical approach is a volitional account of agency, is better than the main alternatives because it is more perspicuous.

Chapter 3: A Defense of the Hierarchical Approach

In the first chapter, I developed an original interpretation of Frankfurt's hierarchical approach. I argued that it should be seen as a volitional account of agency rather than as an account of free agency or autonomy. I also provided a first reason why Frankfurt's approach is best seen as an account of agency: because it is *more contextually sensitive* to see it as such. In the second chapter, I provided a second reason why we should favour my interpretation. I argued that it is better to see Frankfurt's two-level approach as an account of agency, because it eliminates confusions and provides a *more perspicuous* understanding of his view.

In the present chapter, I provide a third reason to see my interpretation as a better interpretation. The third reason is that Frankfurt's hierarchical approach is *more convincing* when we see it as an account of agency. To show that my interpretation is more convincing, I examine a common critique of Frankfurt's view: the claim that his approach is over-inclusive. Then, I argue that this critique convincingly shows that Frankfurt's hierarchical approach is not a plausible account of free agency or autonomy. But I also argue that when we reformulate Frankfurt's approach in agential terms, and abandon Frankfurt's view on free agency and autonomy, the critique in question loses its strength. Thus, I provide a defense of Frankfurt's hierarchical approach as a volitional account of agency.

This third chapter is divided into three sections. In the first section, I provide an exposition of the common critique of Frankfurt's view, the claim that his approach is over-inclusive. I distinguish two forms that such a critique can take: the objection from manipulation, which I attribute to Dworkin and Christman, and the objection from oppressive socialization, which I attribute to Friedman and Benson.

In the second section, I examine a possible response that one might present to defend Frankfurt's approach. That reply is based on the distinction between willing and unwilling actions that Frankfurt draws in his 1975 paper "Three Concepts of Free Action." I show that the response is unsatisfactory.

In the third section, I discussed another strategy to respond to the critique in question. That other strategy is the one that I have defended in the previous chapters; it consists in reformulating Frankfurt's hierarchical approach in agential terms and hence abandoning his view on free agency and autonomy. I show that when we interpret Frankfurt's approach as an account of agency, the critique in question loses its strength, and so I provide a defense of Frankfurt's hierarchical approach.

3.1. The main objection

In this section, I examine a common critique of Frankfurt's account, the view that his approach is over-inclusive. Moreover, I identify two forms that this critique can take: what I call the objection from manipulation and the objection from oppressive socialization.

In an early paper, "Autonomy and Behaviour Control" (1976), Dworkin argues that autonomy requires two things: authenticity and procedural independence. In that paper, Dworkin claims that a person is *authentic* when she is moved by a desire that, after critical reflection, she wants to have. Secondly, procedural independence obtains when there is an absence of "influences which in some way prevent the individual's decisions from being his own."[106] In

[106] Dworkin 1976, p. 25.

other words, there must be an absence of influences such as "manipulation, deception, the withholding of relevant information, and so on."[107]

While Dworkin does not present his view as a critique of Frankfurt's, I believe that it is, at least implicitly, an attack on his view. Indeed, it seems obvious that Dworkin maintained a dialogue with Frankfurt. While Dworkin built his 1976 account of autonomy on his own former hierarchical account of "acting freely" (1970), he also added to his view some concepts developed by Frankfurt in 1971, such as the concept of "first-order desire," "second-order desire," and "identification."[108] Thus, it seems obvious that Dworkin knew Frankfurt's work, and it seems more than plausible to assume that he was, in fact, criticizing Frankfurt's account here.

Dworkin's implicit critique can be expressed as such: authenticity is not sufficient for autonomy; a person also needs procedural independence. Dworkin's critique is that an account of autonomy based exclusively on authenticity, or based exclusively on different levels of desires, will include too many actions such as those that are the results of manipulation and deception. Thus, such an account will be over-inclusive. This is what I call "the objection from manipulation."

To make this point clear, consider the following example of manipulation. Suppose that, in order to win a vote against a social reform, a conservative manipulator captures progressivists and indoctrinates them: he implants in them false belief about the reforms in order to change their opinion and make them vote against it. One of his victims, who was a progressive and who would normally have approved the social reform, now has a desire to vote against it, and, after critical reflection, she identifies with her desire to vote against it. Thus, Frankfurt would have to

---

[107] Dworkin 1976, p. 25.
[108] Dworkin also briefly acknowledged Frankfurt's work in his preface to *The Theory and Practice of Autonomy* (Dworkin 1988, p. x).

conclude that when she acts on her desire to vote against the reform, she acts freely and of her own free will or she acts autonomously. But, in this case, the person's higher-order preference has been the product of manipulation and indoctrination and, for that reason, it seems like she should not count as autonomous or as acting of her own free will. Thus, Frankfurt seems to count as free and autonomous too many actions that should not count as such.

Christman, in his early work, has developed a similar tacit critique of Frankfurt. In "Autonomy: A Defense of the Spilt-Level Self" (1987), he has argued that the hierarchical approach also needs a supplementary criterion such as that developed by Dworkin. The issue for Christman is that the hierarchical framework is insufficient for autonomy, because the process of identification with a desire might itself be heteronomous: it might be the product of manipulation, or in general of "illegitimate external influences." Thus, on his 1987 view, a hierarchical account of autonomy must also include an absence of illegitimate external influences (such as manipulation). [109]

Christman proposes that we understand the concept of illegitimate external influence as follows. First, "these influences must be external. That is, they must emanate or originate essentially from outside the person. They are the sort of phenomena… that essentially redirect the normal cognitive processes that exemplify individual judgment and decision making."[110] Second, these influences must also be illegitimate. To determine whether an influence is illegitimate, Christman proposes the following test: "were the agent to be made aware of their presence and influence, she would be moved to revise her desire set."[111]

---

[109] Christman, however, abandoned the hierarchical framework a few years later in "Autonomy and Personal History" (1991).
[110] Christman 1987, p. 289.
[111] Christman 1987, p. 291.

To illustrate Christman's view, consider our former example, the case of the conservative manipulator. Now to determine whether it is an illegitimate external influence, we should ask, first, whether it "redirects the normal cognitive processes" of the individual she manipulates. The answer is undoubtedly "yes." The manipulator implements in her victim some false beliefs that aim, precisely, at redirecting her victim's normal cognitive processes. Second, we have to ask: were the victim aware of the manipulator's influence, would she be moved to revise her desire set? In other words, if the victim was aware of the manipulator's action, would she be moved to reconsider her desire to vote against the social reform? Knowing that she was formerly a progressive, she would surely want to reconsider her newly implemented desires. Thus, we might conclude that the victim has been subject to illegitimate external influences.

Notice that Christman's view does not depart importantly from Dworkin's critique. In fact, he sees his own account as a development of Dworkin's. Christman criticized Dworkin's notion of procedural independence for being "just as mysterious as that of autonomy itself."[112] Thus, what Christman attempted to do in that 1987 paper is to develop the notion of procedural independence.

While Dworkin's and Christman's critiques were not explicitly directed at Frankfurt's view, Friedman, on the other hand, has explicitly criticized Frankfurt. What preoccupies Friedman is not so much the presence of direct influences such as "manipulation, deception, or any other possible constraints on the reflection process," but the phenomenon of oppressive socialization. This is what I call "the objection from oppressive socialization."

Friedman points out that a person's critical reflection, and thus a person's second-order desires or volitions, might be influenced by oppressive norms or principles in such a way that

---

[112] Christman 1987, p. 288.

precludes autonomy. As an example, she gives the case of a housewife who has been taught that her place is in the home. When she reflects on her desire to flee from the responsibilities of her domestic life, she condemns it on the basis of the misogynist norms that she blindly follows. But, as Friedman points out, it is implausible to suppose that her critical reflection is itself autonomous. Thus, when she identifies with her desire to be a good housewife, she cannot count as autonomous because her critical reflection is not itself autonomous. The issue, then, is that Frankfurt's would have to claim that she is autonomous, because she identifies with the desire that leads her into action. Once again, Frankfurt's account appears to be over-inclusive.

Friedman argues then that we must account for the autonomy of these second-order desires, or for the autonomy of the process of critical reflection that leads to their formation. One way to account for the autonomy of these phenomena is to adopt the same strategy that we adopted for the first level: to look one level higher in the hierarchy. But this solution generates a regress. One could also claim that the process of critical reflection and the second-order desires to which it gives rise are autonomous in virtue of a different kind of relation. But then, we would need to provide an account of this other way, or else our account of autonomy will be incomplete.

Friedman believes that the second solution is the right one, and thus she attempts to explain how higher-order preferences become autonomous. Her strategy consists in combining the "top-down" approach of Frankfurt with a "bottom-up" approach. On this bottom-up view, one's lower-order desires constitute "touchstones of a sort for the assessment of the adequacy of one's" higher-order principles.[113] On this view, when a person's lower-order desires do not correspond to her higher-order principles or to her critical reflection, they might in fact be more

[113] Friedman 1986, p 31.

truly autonomous than her higher-order principles. For example, the woman who has been taught that her place is in the home might question her higher-order principles on the basis of her "persistent dissatisfactions and repeated urges to flee from the responsibilities and limitations which structure her domestic life."[114] It appears that these lower-order desires are in fact more autonomous than her higher-order principles. Thus, on her view, lower-order motivations have to be the object of critical reflection, but the higher-order principles that inform critical reflection must also be assessed according to their "fit" with lower-order motivations. This can be formulated as a critique of Frankfurt's hierarchical approach of autonomy: on Friedman's view, the top-down approach is insufficient and we must supplement it with a bottom-up approach.

Benson is another philosopher who has explicitly criticized Frankfurt's approach. Just like Friedman, Benson is concerned by certain forms of oppressive relationships, but his focus is on how these relationships affect a person's sense of self-worth, and derivatively, her free agency. For Benson, this sense of self-worth is a necessary condition for free agency, a condition that is overlooked in hierarchical accounts. Thus, his critique can be expressed as such: we cannot account for free agency in terms of different levels of desires; we must also include (at least) a person's sense of self-worth.

Now, a few words about Benson's concept of self-worth. On Benson's view, to have self-worth amounts to having a "sense of one's own competence to act." This sense of self-worth, moreover, is "sensitive to others' attitudes toward the agent."[115] In other words, a person's sense of self-worth depends on the relationships in which she stands to others. Thirdly, when a person lacks self-worth, she feels that she does not have the competence to participate in certain

---

[114] Friedman 1986, p 31.
[115] Benson 1994, p. 659.

relationships and, *a fortiori*, in those relationships that are at the origin of her lack of self-worth. Finally, this lack of self-worth creates "difficulties with the agent's participating in [these] relationships or interactions with others."[116] Thus, a lack of self-worth tends to lead to social exclusion.

To illustrate this notion of self-worth, Benson presents a few examples. I shall merely focus on the case of the "medically gaslighted woman." The medically gaslighted woman lives during the end of the nineteenth century. Her husband, a physician, is influenced by the medical science of his day and thinks that "excitable" women suffer from psychological illness. The woman who presents these characteristics is thus diagnosed with hysteria and ends up isolated and feeling crazy. The woman who feels crazy believes that she is not competent enough to relate to other "sane" persons, and consequently she ends up isolating herself.

For Benson, Frankfurt's account of free agency is unable to account for this form of constraint. Indeed, the medically gaslighted woman might have a desire to exclude herself from the relationships in question, act on such desire, and identify with it. In such a case, she would fulfill all the conditions stipulated by Frankfurt's account of free agency and thus she would be considered a free agent.

To remedy this issue, Benson suggests that we add a content-restriction to our account of free agency. For him, all desires and beliefs that are incompatible with a person's sense of self-worth, such as the belief that one is not honorable enough to take part in a relationship, are incompatible with free agency. On his view, Frankfurt's hierarchical approach should be supplemented with a content-restriction.

---

[116] Benson 1994, p. 659.

From this discussion it seems clear that Frankfurt's hierarchical framework is not sufficient to account for autonomy or free agency, and that we need a supplementary criterion. Dworkin, Christman, Friedman, and Benson all suggested such a criterion, whether it is "procedural independence," an absence of "illegitimate external influences," a "bottom-up" assessment of one's higher-order principles, or a content-restriction concerning one's worthiness to act. In my opinion, all these views successfully make a case against an account of autonomy or free agency that is purely based on different levels of desires. That being said, one might wonder if these criteria are successful as criteria for autonomy. My opinion is that they all fail, but I will leave that aside because it would bring us too far from our topic.

3.2. An unsatisfactory response: the notion of unwilling action

In the previous section we examined a common critique of Frankfurt's view: the claim that his hierarchical approach is over-inclusive. In the present section I will examine a possible response that one might provide to defend Frankfurt's view. This response is based on the notion of "unwilling action." Although no one (to my knowledge) has attempted to defend Frankfurt's view using this notion, a charitable and thorough reading of Frankfurt requires that we give some thought to it. That being said, I will conclude that a response based on the concept of "unwilling action" is not satisfactory.

In "Three Concepts of Free Action" (1975), Frankfurt considers a certain category of actions, those "unwilling actions" of "Type A."[117] In that kind of situation, the person does what

[117] Frankfurt 1975, p. 47.

he wants to do considering the circumstances, and wills what he wants to will, but he resents being in those circumstances. In Frankfurt's words, there is a "discordance between reality and desire": the person "regrets or resents the state of affairs with which he must in fact contend."[118] Thus, the issue here is not that the person does not desire to perform the *action* he performs, or that he does not desire to have a certain *first-order desire*; what he does not want is the *state of affairs* in which he finds himself. On Frankfurt's view, a situation of this type falls into the category of free action but it is nevertheless an "unwilling action."

To defend Frankfurt's view, one might argue that the cases presented in the previous section are all cases of "unwilling actions" in which the agent resents being in the situation in which she finds herself. Consider the case of the indoctrinated progressive who was subject to manipulation. Maybe she resents being in the circumstances she occupies. Maybe she would prefer not to be indoctrinated or not to be the subject of manipulation. On that reading, her action would count as an unwilling action.

Consider, as well, Friedman's the case of the rebel housewife. The rebel housewife has a desire to be a good housewife because she endorses the misogynist norms of her society. Here we cannot say that she rejects the oppressive relationships she has, like the indoctrinated progressive resented being manipulated. Indeed, the rebel housewife does not reject the oppression of women as such, since she endorses the misogynist norms of her society. But maybe she resents being born a woman. What she rejects, then, is her fate. On that reading, we could also say that the rebel housewife does not desire the state of affairs in which she finds herself, and thus that her action is an unwilling action.

---

[118] Frankfurt 1975, p. 47.

The same considerations could be raised concerning Benson's medically gaslighted woman. The medically gaslighted woman believes that she is a hysteric and does not have a sense of her own competence to act. She ends up wanting to isolate herself and endorsing her desire to do so. We might suppose that she endorses her first-order desire because she trusts the medical establishment of her time and believes that this is how hysterical women should be treated. Thus, she does not seem to reject the oppressive relationship she occupies. But she might resent being hysterical and reject her fate. On that reading, we could say that the medically gaslighted woman does not desire her state of affairs and thus that her action is an unwilling action.

While Frankfurt's notion of unwilling action can account for the cases discussed in the first section, I see two main problems with that response. First, it creates an obscure category of unwilling albeit free actions. The result is that the notion of free action loses its intuitive meaning: what falls into the category of free action is not always what we would normally consider a free action.

Second, it is not hard to adapt the previous cases to develop some counterexamples to Frankfurt's account. Imagine that the conservative manipulator not only implanted in her victim some desires to vote against the reform, but that she also erased her victim's memory: the victim no longer remembered being the subject of manipulation. In a case like that, it is unlikely that she would resent being in these circumstances, simply because she is not aware of them. For this reason, her action would count as both a free and a willing action on Frankfurt's view, which is simply implausible.

Consider, as well, the case of the rebel housewife and the case of the medically gaslighted woman. We supposed that both resent being who they are: the rebel housewife resents

being born a woman and the medically gaslighted woman resents being crazy. But we can easily imagine two cases in which this resentment is absent. The rebel housewife might be a religious person who believes that her fate was given by God. For that reason, she might not resent being born a woman. And the medically gaslighted woman might also have similar belief about her fate. Thus, on that reading, the actions of the rebel housewife and the medically gaslighted woman would both count as free and willing actions for Frankfurt, which is implausible.

From the previous discussion, I conclude that Frankfurt's notion of unwilling action cannot provide a satisfactory response to the previous critique, because it creates an obscure category of "free albeit unwilling action" and because that notion is still over-inclusive.

3.3. A defense of hierarchy

In the first section, we examined a common critique of Frankfurt's view, the claim that his hierarchical approach is over-inclusive because it counts as free or autonomous too many actions that should not count as such. In the second section, we saw that Frankfurt's notion of "unwilling action" cannot provide a satisfactory response to that critique. In the present section, I argue that another strategy is more successful to respond to the critique in question. The strategy consists in reformulating Frankfurt's approach as a volitional account of agency, as I have done in Chapters 1 and 2, and in abandoning his view on free agency and autonomy. I will show that my interpretation of Frankfurt is better than the main alternatives, because his approach, when reinterpreted in this way, is more convincing.

It would be useful to start our discussion with a reminder of my interpretation of Frankfurt and spell out one more time what I take to be his volitional account of agency. As we

saw, Frankfurt endorses volitionism or the trying view of action. Volitionism is the view that all actions involve an act of will, that is, a choice or a decision. For the sake of simplicity, I shall focus here on actions involving decisions. On that view, when I run, I necessarily decide to run: the decision to run is a necessary component of the act of running. Moreover, we saw that for Frankfurt a decision is a kind of mental act that involves a second-order volition. When I make a decision, I "make up my mind" or "constitute my will" by means of a second-order volition: the second-order volition makes the first-order desire stronger and allows it to effectively lead to action. Finally, we should note that Frankfurt also talks about a decision as an act of identification.

Let us now examine the previous critiques in light of this interpretation of Frankfurt. I shall start with Friedman's and Benson's counterexamples. Both philosophers develop a version of the objection from oppressive socialization: they are concerned about the effect of oppressive socialization on free agency or autonomy. Friedman introduces the case of the rebel housewife. The rebel housewife is not autonomous with respect to her second-order volitions because they are the product of oppressive norms, the view that the place of a woman is in the home. Benson, on the other hand, introduces the case of the medically gaslighted woman who is the product of the medical establishment of her time. The medically gaslighted woman believes that she is crazy, or "hysterical," and unworthy of relating with "sane" persons. As a result, she experiences some "difficulties… participating in certain relationships or interactions with others"[119] and ends up isolating herself.

The intuition that both Friedman and Benson trigger is the view that oppressive socialization is incompatible with free agency or autonomy, and Frankfurt's account of free

[119] Benson 1994, p. 660.

agency seems to be ill-equipped to interpret these cases. This seems to be on point. But what if we interpret Frankfurt's approach simply as an account of agency and abandon his views on free agency and autonomy? As we saw, this is possible because the concept of agency is more fundamental than that of free agency or autonomy. If we do this, it seems like our intuition in these cases completely shifts. The question now is not whether the rebel housewife is autonomous, or whether the medically gaslighted woman is a free agent. The question is whether they are agents, or whether they are both fully active in their behaviour. Is the rebel housewife *acting* when she cleans the dishes and endorse her desire to do so? Is the medically gaslighted woman *acting* when she isolates herself and watches television, instead of taking part in a social gathering with "sane" persons? Are they both driven by active desires? Or are they driven by passive mental episodes? Here, I see no reason why we should deny that they are both fully active in their behaviour.

The reason why our intuitions concerning these cases radically shift is simply because oppressive socialization, while incompatible (at least at first sight) with autonomy or free agency, seems perfectly compatible with agency. A person who blindly endorses social norms could still be said to be acting. In this respect, oppressive socialization is similar to coercion. While coercion is incompatible (at least at first sight) with autonomy and free agency, it seems perfectly compatible with agency. A slave who submissively follows orders could be said to be active. There is no incoherence in that claim.

Before moving on to the objection from manipulation, I should say here that my response is only a partial response to Friedman's objection. Indeed, in her critique of Frankfurt's view Friedman develops a version of the regress objection, which might be called "the oppressive socialization version of the regress objection." I showed earlier in this section that her intuitions

about oppressive socialization are not problematic for Frankfurt's approach when we interpret it as an account of agency. But one might point out that Frankfurt's view is still subject to a certain form of the regress objection. I agree with that claim. That being said, the objection should be formulated in agential terms rather than as a matter of autonomy and heteronomy. When formulated this way, the objection is simply the well-known objection to the volitional approach to agency: an objection that was forcefully formulated by Ryle, and before him by Hobbes. I will discuss that objection in more detail in Chapter 6, so we may leave it aside for now.

We just examined the objection from oppressive socialization and saw that it is easily solved when we interpret Frankfurt's approach as an account of agency. Let us now move on to the objection from manipulation. That objection is a bit more complex. As we saw, Dworkin and Christman are particularly concerned about illegitimate external influences, such as manipulation, which seem to preclude autonomy, but we may ask again whether these influences pose a problem for an account of agency (*tout court*). Here, it seems like the answer is "yes": manipulation, unlike oppressive socialization, can foreclose agency. Consider the case of a puppeteer who manipulates someone's body by means of a complex set of ropes. Her victim's body is moving, but this bodily motion cannot be said to be the victim's action. Since manipulation can foreclose agency, we need to take a closer look at that problem.

First, let us mention that there is an easy way to solve the previous issue for a philosopher of action: she could claim that a person's movement is an action only if it is caused by some mental element. The problem in the previous case of manipulation is that the bodily motion is not the outcome of the victim's mental episodes; the bodily motion is rather the outcome of the manipulator's own action. Thus, this case can be easily solved. But we might also imagine the case of a mental puppeteer: by some ingenious mechanism analogous to a set of rope, this

manipulator can control the mental episodes of his victims. Maybe the mental puppeteer is some evil demon or a very skilled neurosurgeon or a hypnotist who controls the mind of her victim while her victim is undergoing some mental happenings. Now if such a form of manipulation exists, it seems like it would foreclose agency, for it seems like the victim is passive with respect to her mental episodes.

This case of the mental puppeteer should be distinguished from the much-discussed case of the hypnotist who implants desires in her victim – let us call her "the hypnotist-implanter." The difference between the mental puppeteer and the hypnotist-implanter is that the former controls her victim *synchronously* while the latter control her victim *diachronously*. The mental puppeteer produces the mental episodes of her victim *while they occur*. The hypnotist-implanter, on the other hand, implants in her victim something that will manifest itself *later on*: a mental state or disposition, a "character or program," that will eventually lead to the production of a mental episode.

The case of the hypnotist-implanter does not raise a problem for an account of agency. Suppose that a hypnotist implants in me—who cares a lot about health—a desire to eat junk food and a false belief that junk food is in fact healthy. This belief might eventually lead me to identify with my desire to eat junk food. The question then is whether I can be said to act (*tout court*)? Of course, when I eat junk food, we cannot say that I am acting freely or autonomously. But the issue here is about agency. Can we say that I act when I am moved by a desire that was previously implanted in me by a hypnotist? It seems the answer is "yes." Even if I am manipulated in this way, we may say that I am performing an action. And we can also suppose that I am nevertheless active with respect to my desire to eat junk food. Such desire need not be a

passive mental episode that overcomes me. Thus, this form of manipulation does not raise a problem for agency.

We need, therefore, to distinguish two kinds of manipulators: the agency-inhibiting kind, or "the mental puppeteer," and "the hypnotist-implanter" who does not inhibit agency. Now this is in fact a distinction that Frankfurt draws in his early works. First, Frankfurt asks us to imagine the case of a Devil neurologist (D/n) who "manipulates his subject on a continuous basis, like a marionette, so that each of the subject's mental and physical states is the outcome of specific intervention on the part of the D/n."[120] Second, Frankfurt considers the case of a Devil neurologist who "provides his subject with a stable character or program, which he does not thereafter alter too frequently or at all."[121]

Frankfurt claims that, in the second case, the victim *acts freely*: "In that case there is no reason for denying that instances of the subject's behaviour may be members of *W* [the class of free actions]. Nor, in my opinion, are there compelling reasons either against allowing that the subject may act freely or against regarding him as capable of being morally responsible for what he does."[122] Now, as a statement about *free* agency, this is not very convincing. But when we rephrase Frankfurt's account of free agency as an account of agency *tout court*, we might agree with Frankfurt and claim that this kind of manipulation is not problematic.

The first case, the case of the mental puppeteer, is more problematic since it seems to preclude agency. But Frankfurt has a way to exclude these cases from his account of agency. Here, Frankfurt could simply argue that a victim who is manipulated by a mental puppeteer is not acting, because she is not performing an act of identification. As Frankfurt recognizes, the

---

[120] Frankfurt 1975, p. 53.
[121] Frankfurt 1975, p. 53.
[122] Frankfurt 1975, p. 53.

victims might have second-order desires and second-order volitions: her mental life might be as complex and rich as that of an agent. But the mental puppeteer cannot take the place of her victim and identify with a desire. The mental puppeteer cannot perform this act of identification, because she does not have access to the mental history of her victim, and an act of identification requires that kind of access. An act of identification is an act whereby the agent creates her identity, or connects together her mental episodes, and only the agent herself can do so.[123] Thus, we can conclude that the victim of the mental puppeteer does not act, because her behaviour does not include an act of identification.

The difference between the mental puppeteer and the hypnotist-implanter is thus that the mental puppeteer does not allow her victim to perform an act of identification, while the hypnotist-implanter does allow it. The result is that the mental-puppeteer forecloses agency, while the hypnotist-implanter does not.

To conclude, it seems clear that oppressive socialization and manipulation raise some issues for a theory of free agency or autonomy, and that any such theory must find a way to account for these forms of influence. However, oppressive socialization and manipulation do not raise the same issues for an account of agency. Only one form of manipulation appears to be problematic (the case of the mental puppeteer). But Frankfurt's notion of identification can exclude this form of influence. So, Frankfurt's hierarchical approach, when formulated *as an account of agency*, is not as flawed as it is often claimed to be.

---

[123] When discussing the case of the person manipulated by a puppeteer, Frankfurt claims that "the subject is not a person at all. His history is utterly episodic and without inherent connectedness. Whatever identifiable themes it may reveal are not inherently rooted; they cannot be understood as constituting or belonging to the subject's own nature" (Frankfurt 1975, p. 53).

3.4. Conclusion

In this chapter, I provided a third reason why we should favour my interpretation of Frankfurt's hierarchical approach over the main alternatives. I showed that when we see his approach as a volitional account of agency, as I propose to interpret it, it is *more convincing*. In order to show this, I examined the claim that his view is over-inclusive, and I distinguished two forms that such a critique may take: the objection from manipulation and the objection from oppressive socialization. After discussing and rejecting a first solution to that critique, I showed that a better strategy consists in reformulating Frankfurt's approach as an account of agency and abandoning his views on free agency and autonomy, and that this strategy allows us to solve the problem in question.

Part II: The Endorsement View of Agency

In the first part of my thesis, I argued that Frankfurt's hierarchical approach is best seen as an account of agency. In the second part, I use my interpretation of Frankfurt to develop an account of agency, what I call the "endorsement view of agency." The aim of this second part is to show that the endorsement view, which is a non-causal account of agency, is a serious contender within the metaphysical debate.

While the first part of my thesis was primarily interpretative, the second part is not although it also engages with interpretative issues. To avoid confusion, I will highlight some of these interpretive questions. In Part I, I argued that Frankfurt developed an event-causal account of agency in his 1971 paper. The first element to mention is that Frankfurt abandoned the causal approach in his 1978 paper "The Problem of Action" and developed a non-causal one, which in my opinion is more compelling and more innovative. For that reason, I will focus on this view in Part II. The second element that the reader should keep in mind is that I will interpret Frankfurt's work more freely in this second part of the thesis, since the aim is not to provide an interpretation of Frankfurt but to develop my own account of agency.

Part II is comprised of four chapters. In Chapter 4, I discuss the notion of control specifically. I argue that the endorsement view, which is based on the non-causal concept of standby control, provides the most compelling account of control because its solution to the problem of internal deviance is the strongest.

In Chapters 5 and 6, I focus on what it is to be an agent, and thus on the agential aspect of agential control. In Chapter 5, I examine the problem of the disappearing agent. I argue, *contra* Schlosser, that the problem of the disappearing agent is a serious issue for the standard account.

In Chapter 6, I examine three different solutions to the problem of the disappearing agent or different accounts of the agent: the event-causal, the agent-causal, and the non-causal accounts. I argue that the event-causal and the agent-causal views are not satisfactory. Then, I turn to Frankfurt's non-causal account of the agent. I argue that it is unsatisfactory as well, but I also show that we can modify his account and make it more compelling. Two modifications are particularly significant: I defend, unlike Frankfurt, a *dispositional* and *dialogical* non-causal account of the agent. Following these modifications, the non-causal account will emerge as the strongest.

In Chapter 7, I explain how my non-causal account of agential control can provide resources to non-causal accounts of action explanation. In particular, I examine the debate between Sehon and Mele, and I argue that my non-causal view can satisfactorily respond to Mele's critique.

Chapter 4: The Endorsement View and Standby Control

As we saw in the introduction, one common assumption in the metaphysics of agency is that agency requires agential control. In the present chapter, we will focus on the notion of control specifically. I will argue that the endorsement view is based on the most compelling account of control, which I call *standby control*. This account of control is the most compelling because it provides the most satisfactory solution to the problem of internal deviance.

The standard account or the causal theory of action is based on the idea that we exercise control over our behaviour when our behaviour is caused by an intention and a desire-belief pair. However, this notion of control has been criticized: one of the main critiques is the problem of internal deviance.[124] What this problem shows is that the notion of control on which the standard account is based is flawed, because it allows cases of deviance in which control is lost.

There have been many attempts at solving the problem of internal deviance, and thus, at developing a more satisfactory account of control. As we saw in the introduction, we can identify three main approaches in the metaphysics of agency: the event-causal, the agent-causal, and the non-causal approaches. The proponents of each of these approaches have developed their own accounts of control, and thus their own solutions to the problem of internal deviance. In this chapter, I will examine the most influential attempts made by the proponents of each of these approaches and argue that the endorsement view, which is a non-causal approach, provides the strongest solution.

---

[124] For an early discussion of the problem, see Chisholm 1964, Taylor 1966, Davidson 1973, Frankfurt 1978, Peacocke 1979, and Bishop 1989.

This fourth chapter is divided into five sections. In Section 4.1, I present the three main approaches to causation in the metaphysics of agency (the event-causal, the agent-causal and the non-causal approaches) and explain how they differ. In Section 4.2, I provide an exposition of the problem of internal deviance. In Section 4.3, I examine one of the main event-causal solutions and conclude that it is unsatisfactory. In Section 4.4, I examine two of the main agent-causal solutions and conclude, likewise, that they are unsatisfactory. In Section 4.5, I explain that the endorsement view is based on the notion of standby control, and argue that it is the most satisfactory solution to the problem of internal deviance.

4.1. Three approaches to causation

Beside elucidating the nature of agency, there is in the philosophy of agency a broader source of philosophical puzzlement, what has been called "the problem of natural agency." In the first part of this section, I examine how the problem of natural agency can be conceived as a reformulation of the free will debate. After presenting the problem of natural agency, I discuss the three main positions that one might take within the debate: the reductive or event-causal view, the non-reductive or agent-causal view, and finally the non-causal view.

The free will problem is a debate over whether the freedom required for moral responsibility is compatible with a deterministic conception of the world. On the one hand, the world is often seen as fully determined. One way to express this position is to say that "every event… is causally necessitated by antecedent events."[125] On the other side of the debate, it is

[125] McKenna and Coates 2018.

often claimed that the freedom necessary for moral responsibility is a two-way power, that is, a power to do or not do something. On this reading, an action for which we are morally responsible cannot be an action that is necessitated by prior events; hence there is a dilemma: either the world is determined and we are not free and morally responsible agents or we are free and morally responsible agents but the world is not determined.

Some philosophers have advocated a reformulation of the free will debate and suggested that we should talk about "the problem of natural agency" instead.[126] Firstly, it is far from clear that determinism is true. As Bishop and Dennett argue, natural science no longer adheres to determinism.[127] Some events have been shown to be causally undetermined; other events have been shown to be "caused according to probabilistic laws rather than deterministic laws."[128] What seems to pose a real threat to our conception of ourselves as morally responsible agents is our current natural scientific worldview. As Dennett argues, "the threat is not determinism—if it were, we could all relax since physicists now seem to agree that our world is fundamentally indeterministic—but science itself, or the 'naturalism' that is its enabling world views."[129]

Now, what is naturalism exactly? Naturalism is an ontological view according to which all the entities that exist are, or can be reduced to, natural entities – the entities that natural science postulates.[130] In the philosophy of agency, naturalism is often seen as a conception of the world that involves events, states of affairs and dispositions. According to this worldview, everything that happens can be cashed out in terms of events, states of affairs and dispositions[131];

---

[126] Dennett 1984; Nagel 1986, pp. 110-111; Bishop 1989; Velleman 1992; Lowe 2008, pp. 160-161.
[127] Bishop 1989; Dennett 1984, p. 31.
[128] Bishop 1989, p. 18.
[129] Dennett 1984, p. 31.
[130] On this, see Papineau 2020.
[131] It is often argued that dispositions are part of the natural event-causal order, because they can be reduced to events and relations among them, or to categorical properties, events, and relations among them. For example, on the simple conditional analysis (Ryle 1949, Goodman 1954, and Quine 1960), when we say that something is

an event is said to be caused by another event, state of affairs, or disposition. For example, the collapse of a bridge (an event) might be said to be caused by another event (the explosion of a bomb)[132] or a state of affairs (a structural defect)[133] or a disposition (the fragility of the bridge). But when we explain an event by means of a state of affairs or disposition, it is always on the assumption that another preceding event took place.[134] The collapse of the bridge might be caused by a structural defect *and* a particularly strong shock (an event). Thus, naturalism tends to focus on the causation of events by other events. For that reason, I shall talk about naturalism as the "natural event-causal order."

Now, naturalism is compatible with both determinism and indeterminism, depending on whether an event is causally necessitated or not – that is, depending on the laws governing this causal relation. If an event is caused according to deterministic laws, then it is determined; if it is caused according to probabilistic laws, then it is not. Naturalism is thus a worldview composed principally of events or happenings, and natural laws connecting them.

Second, it has been argued that we should not only modify the deterministic side of the debate but also the free will side. For one thing, it is far from obvious that moral responsibility requires freedom conceived of as a power to do or not do something; that is, it is far from obvious that moral responsibility requires alternate possibilities. As we saw above, Frankfurt developed a very influential critique of this view: he has argued, using what is now known as

---

"disposed to *x*," what we mean is that "if it undergoes some triggering conditions *y*, then it would produce response *x*," *x* being its characteristic manifestation (on this see Mayr 2011, p. 170 and Choi and Fara 2018.) On that view, a disposition is reducible to two events, the occurrence of the triggering conditions *y* and the production of the characteristic manifestation *x*, and a relation among these events. And since dispositions are reducible in this way, then they can be said to be part of the natural event-causal order. While this view is a controversial one (see Mayr 2011 for a discussion), it is beyond the scope of the thesis to defend it. Thus, I shall simply assume that dispositions are part of the natural event-causal order.

[132] The example is from Lowe 2008, p. 3.

[133] The example is from Davidson 1963, p. 12.

[134] Davidson 1963, p. 12.

"Frankfurt-style examples," that moral responsibility *does not* require alternate possibilities.[135] A far less controversial position is the view that we are responsible for what we *do*, that is, for the *actions* that we perform.[136] On this less controversial view, moral responsibility requires agency (among other things). So instead of opposing naturalism to freedom of the will, it has been suggested that we should oppose naturalism to agency.

The question then is whether agency is compatible with a naturalistic conception of the world, that is, with the event-causal order. According to the naturalistic event-causal order, every event is caused by a previous event like a chain of dominos.[137] On this view, every event is a mere effect of the events that precede it. The difficulty is to explain how agency is possible in such a worldview. In other words, we might wonder if there is a place for an agent within such a chain of causal reactions.

Three main positions have been defended within the debate. The first position, which is by far the most influential, is the reductive or event-causal view. According to the reductive approach, agency is a natural phenomenon and thus it can and should be described in event-causal terms. Consider an agential statement such as "John killed the Queen." Statements like that usually include an agent and an event. In our example, there is an agent (John) and an event (the killing of the Queen). Now, the question is whether this statement can be redescribed in event-causal terms. More specifically, what is salient here is whether "John" can and should be reduced to some events and states of affairs, and in particular to mental events and states of affairs. For reductionists, the answer is "yes": for them, when we say something like "John killed

---

[135] Frankfurt 1969.
[136] See for example, Pink 2010, p. 97: "Moral responsibility is for action. It is for how we act that we are responsible—not for what happens to us independently of our own doing. This is our natural intuition." But, as Pink mentions, this view has been criticized as well (see, in particular, Scanlon 1998).
[137] I borrow the analogy from Dennett 1984, p. 83.

the Queen," what we really mean is that something happening to John (an event) caused the killing of the Queen (another event).

The causal theory of action or the standard account (as it is often attributed to Davidson for example) is one influential way to develop the reductive position.[138] According to this view, in an agential statement, the agent can be reduced to mental events or, more specifically, to a belief-desire pair. When I say that "John killed the Queen," what I mean, on that view, is that the killing of the Queen was caused by a desire-belief pair such as the desire to create a political crisis and the belief that killing the Queen is conducive to it. Thus, on that view, an action is defined as an event (a bodily movement) that has been caused by a mental event or desire-belief pair. We might add here that, on the standard account, the belief-desire pair is often said to lead to the formation of an intention, which itself causes the bodily movement in question.

This reductive view is at first sight quite plausible. Consider another causal statement like "the bomb caused the collapse of the bridge." Here, we have a substance (a bomb) and an event (the collapse of the bridge); the substance is said to be the cause of the event. In a case like this, however, it seems like the statement can and should be reformulated in event-causal terms. After all, the cause of the collapse of the bridge seems to be something happening to the bomb, namely, the explosion of the bomb (an event), and not the bomb itself (*qua* substance). Thus, it seems like the reductive strategy is quite plausible.

But the reductive view, and the causal theory of action in particular, is also subject to serious difficulties. Two of them are the problem of internal deviance and the problem of the disappearing agent. The problem of internal deviance is a problem about the account of control on which the standard account is based. This issue will be discussed in the next section (Section

---

[138] See Davidson 1963.

4.2). The problem of the disappearing agent, on the other hand, is a problem about *agential* control specifically. According to that objection, the causal theory of action has failed in its reduction of the agent: the agent has disappeared. This problem will be discussed in Chapter 5.

Because of these difficulties, some philosophers have defended different approaches to causation. The main alternative to the reductive or event-causal view within the metaphysical literature is the non-reductive or agent-causal view. According to the agent-causal approach, the reductive strategy cannot be used for agential statements. Thus, for agent-causalists, an agential statement like "John killed the Queen" literally means that John (an agent) caused the killing of the Queen (an event).

There are two main agent-causal views: the *ontologically* non-reductive view and the *conceptually* non-reductive view.[139] On the ontologically non-reductive view, the agent is *a substance* and cannot be ontologically reduced to events and states of affairs. Thus, on this first view, agent-causation involves the phenomenon of substance-causation. Second, on the conceptually non-reductive view, the agent can be ontologically reduced to mental events and states, but we cannot provide a conceptual analysis of agency. For example, one might argue that the agent is reducible to active events such as the mental acts of *inferring* and *carrying out*,[140] and also argue that the inferring and carrying out in question are conceptually basic and cannot be analyzed further. On that view, what is required is not so much the involvement of the agent as the involvement of an *active* event.

While the agent-causal view can, arguably, avoid some of the main difficulties that the event-causal view faces, it does have some weaknesses as well. The main weakness of the agent-

---

[139] Clarke 2010, p. 218
[140] For an example of this view, see Bishop 1983.

causal view is that agency is, on that view, hardly intelligible. As Clarke mentions: "critics often purport to find the thesis that agents are causes unintelligible, when it is denied that agent causation is conceptually or ontologically reducible to event-causation."[141] The issue is that agent-causal views seem to be inevitably circular, since they define agency either by means of an active substance (the agent) or by means of an active event. Consequently, agency remains unanalyzed and obscure. Along with Nagel, we might say that the agent-causal view merely gives a name to a mystery.[142]

The third approach is the non-causal view. There are two main non-causal solutions to the problem of natural agency. The first is what we might call "uncaused volitionism," which is the view that agency involves an uncaused or spontaneous act of will. This non-causal view is similar to the conceptually non-reductive agent-causal view since it assumes that the agent is ontologically reducible to mental states and events but conceptually irreducible. What is distinctive of uncaused volitionism, however, is the claim that an act of will is a kind of spontaneous or uncaused mental act. That being said, this non-causal view is subject to the same difficulty as the agent-causal view, since it is based on the notion of an irreducibly active event.

The second non-causal solution to the problem of natural agency is the *structural* approach, which can be either a reductive or a non-reductive view. On the structural view, what matters is the *structural* relation between an agent and an event and not the *causal* relation. On that view, an agential statement like "John killed the Queen" entails the existence of both an event (the killing of the Queen) and an agent (John) but the relation between the two is not a causal one: the statement *does not* mean that John caused the killing of the Queen. On the

---

[141] Clarke 2010, p. 218.
[142] Nagel 1986, p. 115.

structural view, the killing of the Queen should be attributed to John, but this is so in virtue of another kind of relation.

The structural view is, in my opinion, the most promising. It is the most promising because it provides the strongest solution to the problem of internal deviance (see Section 4.5) and the problem of the disappearing agent (see Section 6.3). That being said, it does have some weaknesses as well. The main difficulty of the structural view is what some have called the "mysterious connection" issue[143]: if the agent is not causally responsible for her actions, then what is the connection between the two? One of the most plausible way to articulate this non-causal relation has been developed by Frankfurt and is based on the concept of *endorsement*. In the last section of the chapter (Section 4.5), I shall put forward an endorsement view inspired by Frankfurt's work.

4.2. The problem of internal deviance

We saw that there are three main approaches to causation in the philosophy of agency: the event-causal, the agent-causal, and the non-causal approaches. The main aim of this chapter is to show that the non-causal view provides the strongest account of control, because it provides the strongest solution to the problem of internal deviance. In order to do so, I first explain the problem of internal deviance. Second, I examine one of the main event-causal solution and argue that it fails to solve the problem of internal deviance. Third, I examine the agent-causal responses and I conclude that they are unsatisfactory. And finally, I argue that the non-causal response is

[143] The problem is originally attributed to Davidson. See Davidson 1963, p. 11.

the most satisfactory. In the present section, I start with an exposition of the problem of internal deviance.

The problem of internal deviance is one of the most enduring difficulties that the causal theory of action, or the event-causal view, has to face. Chisholm, Taylor, and Davidson himself were among the first to identify the problem.[144] The critique consists essentially in developing a counterexample, a case in which there is no action even though all the conditions stipulated by the causal theory are fulfilled. To adequately grasp this critique, let us first recall the conditions stipulated by the causal theory of action.

On the causal theory of action, a bodily motion counts as a bodily action when it is caused by the relevant mental items. These mental items include a desire-belief pair and an intention. On this view, a person who is walking to the corner store is performing an action if her bodily motion is caused by an intention to walk to the corner store, which is itself caused by a desire-believe pair—for example, the desire to buy milk and a belief that walking to the corner store is conducive to buying milk.

Many examples of internal deviance have been raised against this account of action. In the counterexamples, an intention and the relevant desire-belief pair cause a bodily movement but the causal process deviates, usually due to nervousness, anxiety or agitation. In the examples, nervousness, anxiety or agitation is part of the causal chain from mental event to bodily motion, but it seems to be an interruption that undermines our willingness to treat the resulting movement as an action. Consider the case presented by Frankfurt:

> A man at a party intends to spill what is in his glass because he wants to signal his confederates to begin a robbery and he believes, in virtue of their prearrangements, that

---

[144] Chisholm 1964, Taylor 1966, and Davidson 1973.

> spilling what is in his glass will accomplish that; but all this leads the man to be anxious, his anxiety makes his hand tremble, and so his glass spills.[145]

The problem here is that the spilling is caused by an intention (and by the relevant desire-belief pair), but it does not seem to be an action because the causal process has been interrupted by anxiety.

Cases of internal deviance suggest that the conception of control on which the causal theory of action is based is flawed. On the standard view, desires, beliefs and intentions bear agential control: we exercise control through their causal efficiency. This account of control is probably best described as a form of *initial* control. The idea is that the mental items in question bear agential control because they initiate the bodily motion in question. But this focus on initial control is precisely why the causal theory of action is subject to the problem of deviant causal chains: the theory does not look at what is going on after the initiating phase. In cases of internal deviance, control is lost somewhere in the causal chain and the resulting bodily motion is not under the agent's control. This is the reason why we tend to treat cases of internal deviance as involving no action.

4.3. The event-causal solution

We just identified the problem of internal deviance. In the next few sections, I examine how the three different approaches can respond to the problem. To begin with, I consider the event-causal solution. Many causalists have attempted to solve the problem of internal deviance

[145] Frankfurt 1978, p. 70.

with the resources of the event-causal framework.[146] It is beyond the scope of this chapter to examine all of these solutions. Instead, I shall focus on the most popular strategy, which consists in adding a requirement to the standard account: according to that view, agency involves not only initial control but also a kind of *sustained* control.[147] I argue that this event-causal solution is unsatisfactory.

To solve the problem of internal deviance it has often been argued that we should add a "causal normalcy" requirement to the causal theory of action. This requirement has to do with the *way* an intention causes the relevant bodily motion and consists in specifying that the causal process must take a "normal course" or that a bodily motion must be caused "in the right sort of way." The challenge, then, is to explain the nature of this normal causal process. One prominent way to explain the nature of this normal causal process appeals to the notion of *sustained causation*. On this view, the mental items in question must not only initiate the causal chain, they must also cause it in a *sustained* way. Thus, on this view, actions require two forms of causation or two forms of control: initial and sustained causation.

The notion of sustained causation is often explained using the negative feedback model of action.[148] Negative feedback mechanisms are those mechanisms that function to keep certain features relatively stable in spite of interference, such as a thermostat. A thermostat keeps the temperature of the room relatively stable in spite of external variations of temperature. Now, it is assumed that actions contain negative feedback mechanisms as well. The idea is that there is a

---

[146] For some of these solutions, see Goldman 1970, Morton 1975, Bach 1978, Peacocke 1979, Davies 1983, Brand 1984, Thalberg 1984, Alston 1986, Bratman 1987, Dretske 1988, Bishop 1989, Adams and Mele 1989, Mele 1992 and 2003, Enç 2003, Schlosser 2007 and 2011, Aguilar 2010 and 2012, Wu 2016.
[147] Those who developed a version of this solution include Bach 1978, Thalberg 1984, Alston 1986, Bishop 1989, Adams/Mele 1989, Mele 2000 and 2003.
[148] See Mayr 2011, p. 122.

state, usually the content of an intention, that the mechanism is trying to keep stable or to accomplish all the way to the end in spite of possible interferences. This mechanism involves a "checking for match and mismatch" and the production of the relevant bodily motion (the subsequent motion in the case of a match, or an adjustment in the case of a mismatch).[149]

Now, it is argued that in cases of internal deviance, the agent's intention initially caused the bodily motion, but it did not cause it in *a sustained way*. So, adding to the causal theory the requirement that it must involve "sustained causation," we can therefore exclude these problematic cases of internal deviance. Consider, again, Frankfurt's example of the person spilling the content of his glass. A desire-belief pair is said to cause an intention to spill the content of his glass, and this intention makes the person nervous, his hand trembles, and the person spills the content of his glass. In a case like this, it is said that the intention initiates a causal chain ending with the spilling, but the intention does not cause the bodily motion in a sustained way. Sustained causation is not involved because during the "check for match and mismatch," the intention is not responsible for the causation of the subsequent bodily motion. An incontrollable bodily process takes over.

While this solution can solve the existing cases of internal deviance, it nevertheless faces certain objections.[150] The first is that it introduces more fine-grained temporal gaps and thus that it preserves the possibility of deviance. The problem, at the origin, is that the event-causal approach is essentially historical. What distinguishes a cause from its effect is its temporality; a cause occurs prior to its effect.[151] Thus, any appeal to event-causal notions, including the notion

---

[149] See Mayr 2011, pp. 122-124.
[150] For other issues, see Mayr 2011, pp. 124-126.
[151] As Taylor mentions concerning our common conception of causation, the event-causal approach, "[t]he cause of an event, it is now almost universally supposed, is some condition or set of conditions that *precedes* some other, its effect, in time." (Taylor 1966, p. 32)

of sustained causation, will preserve the possibility of deviance. Mayr makes this point clear: "when causal processes are understood as event-causal chains, however minutely we describe the structure of such chain, it is always possible to insert between its steps a series of further steps, which can turn the original chain into deviant chain."[152]

The point is that when a "checking for match and mismatch" occurs, and the production of the subsequent bodily motion or the production of an adjustment follows, it is always possible that the process deviates. Since the production of the subsequent bodily motion or the production of an adjustment is a causal process, a deviant element such as nervousness, anxiety or agitation might be introduced at this level and interrupt the process. The result, then, would not be an action even though the behaviour involves initial and sustained control.

A second problem is that the requirement of sustained control is too strong.[153] Consider Bishop's example of "the finger movements of an accomplished violinist,"[154] which is a kind of highly skilled habitual action. An action such as this one is too fast to allow feedback and the causal production of subsequent motions or interventions. Thus, it seems implausible to assume that all actions involve sustained control, or that sustained control is a necessary component of action. For these reasons, I conclude that an appeal to the concept of sustained control is not a satisfactory solution to the problem of internal deviance, and thus that the event-causal approach remains flawed.

From the present discussion, we can identify two *desiderata* that our solution to the problem of internal deviance, and thus our account of agency, should satisfy. First, it should

---

[152] Mayr 2011, p. 134.
[153] See Mayr 2011, p. 128: "the sustaining causation strategy places too high a requirement on the performance of an action." Sehon makes a similar critique of Mele's solution to the problem of internal deviance. See Sehon 2016, p. 96.
[154] Bishop 1989, p. 171.

avoid introducing more fine-grained temporal gaps. Second, it should avoid being too strong in order to account for highly skilled habitual actions.

4.4. The agent-causal solution

We just examined the most popular event-causal solution to the problem of internal deviance and concluded that it is unsatisfactory. Another possible route consists in abandoning the event-causal framework and adopting the agent-causal approach. As we saw, there are two main agent-causal approaches: the ontologically non-reductive and the conceptually non-reductive approaches. Below, I examine an ontologically non-reductive solution, which is based on a kind of "*active power* control," along with a conceptually non-reductive solution, which is based on a kind of "*active translation* control." I argue that neither is satisfactory.

Let us start with the ontologically non-reductive solution. One way to develop the ontologically non-reductive view is based on the idea that powers are ontologically basic, or ontologically irreducible to properties, events and relations among them. Mayr has developed such an account. On Mayr's view, to account for agency we need to appeal to the ontologically irreducible notion of an active causal power. On this view, causation should be seen not as a chain of events and states connected by lawful regularities but as the actualization of a potentiality. This notion of causation involves a series of events and "the continuing presence of an underlying nature or potentiality,"[155] which is realized in that series of events.

[155] Mayr 2011, p. 133

This account can avoid the problem of internal deviance. In cases of causal deviance, a problematic element is inserted between two steps. But, on Mayr's view, this problematic element will no longer be seen as the actualization of the causal power in question. Take again Frankfurt's example of the thief. The thief has a desire and belief, which cause the intention to spill his glass. Having this intention makes him nervous, his hand trembles, and he spills the contents of his glass. Now, a case like this can be seen as the actualization of a potentiality, such as a power to move one's limb. But it is not the actualization of a power to *act*, or a power to *actively* move one's limb. Thus, with Mayr's account, we can easily exclude cases of internal deviance.

Mayr's solution to the problem of internal deviance may be seen as the development of a certain concept of control, what we may call "*active power* control." The idea here is that our behaviours count as being under our control when they are the result of the actualization of an active power. Thus, according to that view, we control our behaviour through the exercise of our active powers.

One might wonder now whether this solution is satisfactory. The first thing to say is that this approach seems to be incompatible with the natural event-causal order, since it makes use of the notion of an ontologically irreducible power. But traditional defenders of the ontologically irreducible approach, such as Chisholm, would not be too worried about that. On Chisholm's view, for example, agent-causation is affirmed as quite simply a supernatural phenomenon. In Chisholm's words, the agent has a "prerogative which some would attribute only to God: each of us, when we act, is a prime mover unmoved."[156] Thus, on that view, agent-causation is

[156] Chisholm 1964, p. 23.

incompatible with the natural event-causal order, but this should not be seen as a real issue since it makes agency a God-like phenomenon.

Among contemporary defenders of the ontologically non-reductive view, such as Mayr, this move has not been very popular. For someone like Mayr, it is a mistake to see agency as a supernatural phenomenon; rather, agency should be seen as a natural phenomenon. But that position also entails that we re-conceptualize naturalism. Thus, on a view such as Mayr's, the natural order is not an event-causal order but a substance-causal one.

While this position might appear as more promising, there are some important objections to it. One objection is that it involves a move from a Humean account of causation, in which causation is seen as a chain of events connected by lawful regularities, to an Aristotelian account of causation, in which causation is seen as the actualization of a potentiality. And, as Stout mentions, this move "involves a huge wrench from the standard modern way of thinking about causation in philosophy."[157] As Stout also adds, "many philosophers would consider this move back from Hume to Aristotle to be a move in quite the wrong direction."[158] This might not be a problem as such, but it suggests that the natural science is mistaken and that a considerable reformed is required.

A more serious objection might be that Mayr's view, like other agent-causal views, is hardly intelligible. This is the problem that we mentioned in Section 4.1.[159] The reason why Mayr's view might be seen as unintelligible is because it proposes no analysis of agency or, more precisely, because it proposes a circular analysis. To define agency, Mayr appeals to the notion of an irreducibly *active* causal power and, for that reason, agency remains unintelligible

---

[157] Stout 2010, p. 163.
[158] Stout 2010, p. 163.
[159] See Clarke 2010, p. 218.

and obscure. Along with Nagel, we might say that Mayr's agent-causal view merely gives a name to a mystery.[160]

We have just examined one version of the ontologically non-reductive view and its solution to the problem of internal deviance, which is based on a kind of *active power* control, and concluded that it is implausible for two reasons: it involves a violent rupture with our modern scientific way of conceiving causation, and agency remains, on that view, unintelligible. But there is another kind of agent-causal approach: the conceptually non-reductive approach. In "Agent-causation," Bishop has developed such an account. Let us examine his view.

Bishop agrees with the picture of action provided by the causal theory of action. He claims that an action is indeed an event that is caused by other mental events, such as a belief-desire pair and an intention, but he adds that agents have a role to play in these causal processes. Agents play the role of an intervener: they *infer* (proximate) intentions from other mental items and they *carry out* their (proximate) intentions.[161] Agents infer proximate intentions when they form such intentions from other intentions, beliefs, and desires. For example, an agent who has a desire to buy milk and believe that she can buy milk at the corner store can *infer* the intention to go to the corner store. In other words, the point is that the intention is not magically caused to happen. In addition to these mental elements, there is the agent's participation—the agent's inferring. Similarly, in an action, an intention does not simply cause the relevant bodily motion. The agent *carries out* her intention or *translates* it into the relevant bodily motion. Here again, the agent must participate; specifically, in the causal production of the bodily motion.

---

[160] Nagel 1986, p. 115.
[161] Bishop 1983, p. 74.

In light of this, the problem in the cases of internal deviance can be analyzed as follows. Remember the case in question: a man has a desire-belief pair and forms the intention to spill his glass; but forming that intention makes him nervous, his hand trembles, and he spills his glass. The problem is that spilling is not an action even though it was caused by the relevant mental items. On Bishop's agent-causal view, the reason why the spilling is not an action is simple: the intention was not translated into the relevant bodily motion *by the agent*. In this case of internal deviance, the agent's intervention or the agent's *carrying out* is missing.

Now, as we saw, Bishop's view is a conceptually non-reductive one. While it does not make use of the notion of an agent as an enduring substance, it entails that the notion of agent-causation is irreducible. As Bishop mentions, "no *definition* is here offered of the agent-causal relation. The account rests on an unanalyzed notion of an agent's himself carrying out his proximate intention."[162] On Bishop's view, the notion of agent-causation is a basic one that cannot be reduced to simpler terms: "the agency theory takes as primitive inferrings and carryings out of proximate intentions. These kinds of events are relational happenings whose subject is essentially an agent—they are intrinsically doings rather than just happenings."[163]

Thus, to solve the problem of internal deviance, Bishop appeals to the notion of an irreducibly active inferring and carrying out. This solution may be seen as the development of a certain account of control, which we may call "*active translation* control." The idea is that agential control involves, in addition to initial control, the *active translation* of our desires, beliefs and intentions into actions.

---

[162] Bishop 1983, p. 76.
[163] Bishop 1983, p. 76.

This position allows Bishop to avoid a simple objection to his solution to the problem of internal deviance. Indeed, suppose that inferrings and carrying out were reducible or analyzable. Then, they could be reduced to mental events with certain mental antecedents. But if they were analyzed as such, we would introduce a temporal gap between the causal antecedent and the mental event, and with this temporal gap we would preserve the possibility of causal deviance. Thus, this solution would not be a satisfactory one. But when Bishop claims that inferrings and carryings out are "intrinsically doings" or intrinsically active events, he claims that they are active in virtue of their own nature and not in virtue of a certain causal antecedent. Thus, this view avoids the reintroduction of causal antecedents and eliminates the problem of internal deviance.

Moreover, Bishop's appeal to the concept of "active translation control" appears to have one strong advantage over the concept of "active power control": it is compatible with the natural event-causal order. Indeed, on Bishop's view, agency is ontologically reducible to events and states; more specifically, it is reducible to mental events and states (desires, beliefs, intentions, inferrings and carryings out) and to bodily motions. As we saw, this was not the case with the concept of active power control: it appeals to the notion of an ontologically irreducible power.

That being said, Bishop's view is also subject to some major difficulties. I shall mention three difficulties here. The first is that his view seems to be *ad hoc*[164]: the only reason to believe that inferrings and carryings out are unanalyzable is that we would otherwise reintroduce the possibility of causal deviance. Bishop fails to provide any good reason to support his position. Relatedly, Bishop's view remains mysterious. As we saw in the case of the substance-causal

[164] For a similar point about "mental action theory," see Brand 1984, p. 13.

view, agent-causal views provide a circular analysis of agency: they appeal to some active element to define agency. The result is that agency remains unanalyzed and obscure.

The third objection is that Bishop's view fails to provide a unified account of agency. On his view, bodily actions can be reduced to simpler terms. For example, the action of walking can be reduced to a desire-belief pair, an intention, a bodily motion, and the inferring and carrying out of the intention. But the inferring and carrying out, which are themselves actions, cannot be analyzed in the same way: they are intrinsically active events. The issue is that if we are talking about the same thing here, namely, *actions*, then we should be able to provide the same analysis for both. In other words, the problem is that Bishop's account of agency is not a unified one: bodily actions are analyzable while the mental acts of inferring and carrying out are unanalyzable.

From this discussion, we can identify four other *desiderata* that we are expecting from our solution to the problem of internal deviance. In the previous section, we saw that our solution should avoid the reintroduction of temporal gaps and that it should not be too strong. The third *desideratum* is that our solution should be compatible with the natural event-causal order. The fourth *desideratum* is that it should avoid *ad hoc* arguments. The fifth is that it should avoid circularity. And the sixth it that it should provide a unified account of agency.

4.5. The non-causal solution

We just examined how the event-causal and the agent-causal approaches can solve the problem of internal deviance and concluded that the two solutions are unsatisfactory. In the present section, I discuss the non-causal view. I shall leave aside the first version of the non-

causal view, uncaused volitionism, because it is subject to similar difficulties as the conceptually non-reductive view.[165] Instead, I will focus on the structural view and argue that it provides the most compelling solution to the problem of internal deviance. Before making that argument, however, I will develop a structural account that I call "the endorsement view of agency," which is based on Frankfurt's works.[166]

4.5.1. The endorsement view

In this first sub-section, I explain in more details the endorsement view and argue, in particular, that it is based on an account of control that I call *standby control*. Before we dig into that question, however, I will discuss the main assumptions on which the endorsement view is based.

The endorsement view takes a structural approach to agency, which means that instead of looking at the way a behaviour is caused or at the nature of reasons for action, it looks at the structure of a behaviour. On the endorsement view, an action is composed of a complex structure that involves a bodily or mental motion and the agent's endorsement. Moreover, the endorsement view, just like causal views, is based on the idea that an action is a behaviour that is attributable to the agent.[167] But unlike causal approaches, the endorsement view does not assume that attributability is a form of causal responsibility. Rather, a behaviour is attributable to the agent, and thus is an action, when the agent endorses it or makes it her own. The endorsement view is

---

[165] As we saw, uncaused volitionism is the view that all actions involve an uncaused and irreducible act of will. Since this view is a conceptually non-reductive one, it is circular and thus hardly intelligible.
[166] See in particular Frankfurt 1976, 1978, and 1987.
[167] Here, I have in mind a minimal notion of attributability that should not, or need not necessarily, be associated with moral responsibility.

also based on a certain account of agential control. On that view, we exercise control through our endorsements. This is a form of control that I call *standby control*.[168] Moreover, our endorsements "play the role of the agent," as one might say. For that reason we exercise, through our endorsements, a kind of *agential* control. This is a question that we will examine in more detail in Chapter 6.

In this sub-section I shall explore the concept of standby control specifically. As we saw, one of the main assumptions in the metaphysics of agency is that agency requires control over our body and/or mind. Control, however, can take many forms. It is useful to begin our discussion by distinguishing three forms of control. To illustrate these forms of control I will use Frankfurt's analogy of the car coasting downhill[169] and explain how the movement of a car may be under a person's control. I say that this is an analogy because the example suggests that we control the movement of our body and mind in the same way that we may control the movement of the car. But the example may be taken more literally if we assume, for example, that the car is an extension of the person's body.

First, suppose that a car is situated on top of a hill and that a person is standing behind it. Imagine, moreover, that the person pushes the car down the hill and gives it its initial impulse. We might say that the movement of the car is under the person's control: she can control, to some extent, its speed and direction. This is what we previously called "initial control." Now imagine that a second person is inside the car. The second person adjusts the movement of the car once it is moving: she gently hits the brakes to slow it down and turns the steering wheel to make sure that the car stays on the road. In other words, she adjusts the movement when there is

[168] I borrow the expression from Pettit 2012 and Tappolet 2016.
[169] See Frankfurt 1978.

interference. We can say that the movement of the car is also under the control of the second person, but this form of control is different. This is what we previously called "sustained control."

Now imagine a further scenario. The person inside the car is entirely satisfied with the movement of the car: she does not need to intervene and adjust it. Even if the person does not intervene in this scenario, we can still say that the movement of the car is under her control. What matters is that she is ready to intervene and adjust the movement of the car in case there is interference, and that she has the capacity to do so. Since she does not actually intervene, we may call this form of control "*standby* control." This notion of control is equivalent to what Frankfurt calls "guidance."[170] This is the notion of control on which the endorsement view is based.

These three concepts of control involve, as the concept of control generally does, counterfactual conditionals.[171] Initial control and sustained control both involve a counterfactual conditional that expresses a dependence between this form of control and the relevant movement. The counterfactual conditional takes the following form: if the movement in question had not been under the person's control, then it would not have occurred the way it did. This suggests that these two concepts of control are *causal* notions. Indeed, the counterfactual conditional perfectly matches counterfactual analyses of causation.[172]

But things are a bit different in the case of standby control. Indeed, this notion of control does not involve the previous counterfactual conditional: we can say that even if the movement had not been under the person's standby control, it would have occurred exactly as it did. This

---

[170] See Frankfurt 1978.
[171] See Mayr 2011, p. 32.
[172] See Menzies 2017.

shows that this notion of control is not a causal one. This point will be discussed at length in Sections 7.4 and 7.5. While standby control is not a causal notion, it does involve another kind of counterfactual conditional. This counterfactual conditional expresses a dependence between *the presence of interferences* and an attempt to intervene. It can be expressed as follows: if the movement had been subject to interference, the person would have intervened.

Let us now clarify this notion of standby control since this is the concept of control on which the endorsement view is based. A first thing that we might say is that standby control can be cashed out as a *disposition to intervene*. We saw that a person has standby control when the following statement is true: "if there were interferences, the person would intervene." This clause mentions a triggering condition, the presence of interferences, and a characteristic manifestation, an actual intervention. And, as is well known, dispositions typically involve these two elements.[173] Thus, exercising standby control seems to amount to being disposed to intervene.

Intuitively speaking, this seems to be an adequate way of speaking of standby control. We saw that a person has standby control when she is ready to intervene and when she has the capacity to do so. First, imagine that the person in the car is sleeping: we would not say that she is ready to intervene; likewise, we would not say that she was disposed to intervene. Second, imagine that the car was broken and that the person did not have the capacity to intervene. Then, we would not say either that she was disposed to intervene. So, it seems like the language of disposition adequately captures was is going on when a person exercises standby control.

---

[173] To say that dispositions typically involve a triggering condition and a characteristic manifestation does not mean that dispositions can be reduced to these two elements. As we saw, this last point is a controversial view in recent debates over dispositions. In particular, it seems to be subject to counterexamples such as finking cases (see, for example, Martin 1994) and masking cases (e.g., Johnston 1992 and Bird 1998). It is beyond the scope of the present thesis to settle that issue.

One issue arises at this point. One might argue that bodily processes can also be disposed to intervene. The dilatation of the pupils would be one example.[174] My eyes are disposed to intervene in case of interference (e.g., when the light shifts). When the light dims, my eyes intervene by adjusting the opening of my pupils. Thus, it seems like nothing prevents us from saying that my eyes are also disposed to intervene. The problem is obviously that we do not want to say that my eyes are acting. Thus, we need to further specify this notion of standby control in order to exclude those cases.

What we need at this point is a way to distinguish what it is *for an agent* to be disposed to intervene and what it is for a bodily process to be disposed to intervene. My suggestion is to stipulate that a mental element is involved when an agent is disposed to intervene, which is obviously not the case when bodily processes are disposed to intervene. In other words, to be disposed to intervene is just, for an agent, to be in a certain mental state. But what mental state? Here, like Frankfurt suggests, I believe that we should appeal to the concept of *endorsement*. But I shall use the term in a slightly different way than Frankfurt does.

Frankfurt focuses on the idea that we can endorse our *desires*, a process that he calls a process of identification.[175] The reason why Frankfurt focuses on the endorsement of our desire is because he is concerned, as we saw in Chapter 2, with what it is *for a person* to act: a kind of action that involves self-reflexivity. But I see no reason to deny that we can endorse our bodily motion as well as our desires. Maybe this is what a wanton action involves for Frankfurt: the endorsement of a bodily motion that does not include the endorsement of a desire. Whether this

---

[174] This example is also from Frankfurt. See Frankfurt 1978.
[175] The idea here is that when we endorse our desires, we create our identity. For that reason, to endorse a desire is a kind of identification. On this, see Frankfurt 1987.

interpretation is correct or not does not really matter for the present purpose. What matters is that the idea is a coherent one, which it is.

To make this point sharper, consider what is happening when the person is in the car coasting downhill and does not intervene. A way to describe the scenario is to say that the person *endorses* the movement of the car or that she accepts it as it is. This endorsement means that no intervention is required. On the other hand, if there was interference she would not endorse the movement of the car: she would reject it. The same could be said about the endorsement of our bodily motions. Consider a case developed by Mele, the case of Peter who "awakes to find himself rolling down a snow-covered hill" because he was pushed by "his prospective fraternity brothers."[176] Peter is satisfied with the movement of his body and does not intervene even though he could have done so. Another way to describe this scenario is to say that he *endorses* the movement of his body or that he accepts it as it is. And this endorsement, again, shows that he was disposed to intervene but did not do so. Thus, it seems fully coherent to employ the notion of endorsement in regard to our bodily motions. This is the way I will use the concept in the rest of the chapter.

Now there are a few other elements that we need to clarify about this notion of endorsement. First, we need to specify that when we act, our endorsement must be *causally effective*. Suppose, for example, that the person who is coasting downhill endorses the movement of her car but the car is broken such that she cannot intervene and modify its course. Clearly, we would not say that she has control over the car. The reason for that is simply because her endorsement is not causally effective: if she were to reject the movement of the car, she would not be able to intervene. Likewise, imagine that Peter who is rolling downhill endorses the

---

[176] Mele 1997, p. 139.

movement of his body but, unbeknown to him, he is also suddenly paralyzed. Clearly, we would not say that Peter is in control of the movement of his body. The reason, similarly, is simply that his endorsement is not causally effective.

But that is not yet sufficient. We need also to specify that an endorsement must have *direct* causal efficacy or that the agent has *direct* standby control when she endorses her bodily motion. Imagine that I were to endorse my heartbeat. Maybe I am satisfied with its speed and approve it as it is. In this case, we could also say that I have the capacity to intervene with my heartbeat. If I were to be unsatisfied with it, I could run to increase its speed or mediate to diminish its speed, but we obviously do not want to say that my heartbeat is an action that I am performing. The issue is that the kind of control in question is a kind of *indirect* control, not a kind of *direct* control. One way to cash out the distinction between direct and indirect control is the following: indirect control involves the performance of an action while direct control does not.[177] In our example, I have to *run* or to *meditate* in order to intervene with the movement of my heart. This is a kind of indirect control. What is required for agency is *direct* standby control, a capacity to intervene that does not require the performance of a further action.

Let us now wrap up the elements discussed so far. On the endorsement view, an action involves a complex structure composed of a bodily or mental motion and the agent's endorsement, an endorsement that should be directly causally effective. Peter wakes to find himself on top of a high, snow-covered hill and is pushed downhill by his fraternity brothers. His body starts sliding downhill in virtue of gravitational force and the initial impulse given by his fraternity brothers. But Peter also endorses his bodily motion and does so in a directly causally effective way: he could, at any time, directly intervene with the motion and stop it or change its

[177] See for example, Wilson and Shpall 2016 and Mele 1992.

trajectory. Imagine further that, mid-way down the hill, Peter realizes that a kid is in his way and that unless he intervenes he will hit the child. Peter nevertheless endorses the motion of his body and does not intervene. In that case, he *hits* the child. On the endorsement view, Peter performed an action even though his bodily motion was not caused by some mental antecedent. And this seems to be the most natural thing to say.[178] The reason why his bodily motion is an action is that, through his endorsement, Peter took ownership of the motion and the motion became attributable to him, and that is just what an action is: a motion that is attributable to the agent.

4.5.2. The endorsement view and the problem of internal deviance

In the previous sub-section, I developed a version of the endorsement view inspired by Frankfurt's work. In this sub-section, I explain that the endorsement view, which is based on the non-causal concept of standby control, avoids the problem of internal deviance and that it fulfills three of the six *desiderata* that we identified so far. In Chapter 6, I will show that the endorsement view fulfills the other three *desiderata*. Thus, standby control will emerge as the most compelling account of control.

Let us first recall the problem of internal deviance. In a case of internal deviance, a desire-belief pair causes an intention which itself causes the relevant bodily motion, but the process deviates due to nervousness and the result is not an action. In Frankfurt's example, the person has an intention to spill the contents of his glass, this makes him nervous, and his hand

[178] Causalists, on the other hand, would have to say that Peter did not do anything or that he merely let his body hit the child. But imagine that Peter says, "I did not hit the child, I just let my body do it." This would sound like a disingenuous way to avoid blame and thus like an improper description of the scenario. I shall further discuss that question in Chapter 7.

trembles. But the trembling is simply not an action since the person has no control over it. The problem, for the proponents of the causal theory of action, is that they have to conclude that the hand trembling is an action.

The endorsement view avoids this kind of problem because one of its conditions is not fulfilled: the criterion of causal efficacy. Remember that what matters on the endorsement view is that a person endorses her bodily or mental motion in a causally effective way. Now, in cases of causal deviance, the agent's endorsement or rejection is causally ineffective. In Frankfurt's example the person's hand trembles, and whether the person rejects the motion or endorses it he can only do so in a causally *ineffective* way: the person cannot, for example, stop the motion or intervene with it. Thus, cases of internal deviance simply do not count as actions on the endorsement view because causal efficacy is missing.

In Sections 4.3 and 4.4 we discussed six *desiderata* that our solution to the problem of internal deviance should have. First, it should not reintroduce temporal gaps. Second, it should not be too strong. Third, it should be compatible with the natural event-causal order. Fourth, it should avoid *ad hoc* arguments. Fifth, it should not be circular. Sixth, it should not create a disunity. In the next paragraphs, I argue that the endorsement view fulfills the first three *desiderata*.

The first *desideratum* is that our solution should not reintroduce temporal gaps. The endorsement view fulfills this first *desideratum* because the concept of a causally effective endorsement is a *synchronous* concept. First, it is a synchronous concept because an endorsement cannot be a causal antecedent: we cannot endorse something that did not happen yet. When we endorse something (e.g., a bodily or mental motion), it must already exist. Second, it is a synchronous notion because a *causally effective* endorsement cannot be anterior to the

thing that is endorsed. If I were to endorse a bodily or mental motion that happened in the past, my endorsement would not be causally effective since I would not be able to intervene and adjust it. Thus, the notion of a causally effective endorsement is a synchronous notion. For that reason, it does not reintroduce temporal gaps.

The second *desideratum* is that our solution to the problem of internal deviance should not be too strong. As we saw, this is an issue for causalists who appeal to the notion of sustained causation, since this condition seems to be too strong to account for fast actions such as the finger motion of a violinist. That kind of actions seems to be a kind of highly skilled habitual action. One might argue that the endorsement view is also too strong because it is implausible to suppose that habitual actions involve an actual endorsement. Does the violinist endorse his finger motion? Or, when I absent-mindedly swipe my feet on the mat, do I endorse my bodily motion?

I concede that the endorsement view seems to be too strong to account for some actions, especially habitual actions: habitual actions are typically performed absent-mindedly[179] and, for that reason, it seems implausible to suppose that they involve an endorsement. But we can easily solve that issue, if we stipulate that agency requires either an *actual* endorsement or a *hypothetical* endorsement.[180] In other words, what is required is either that I actually endorse my behaviour or that, *if I were to become consciously aware of it*, I would endorse my behaviour.

This hypothetical endorsement can be cashed out as a disposition to endorse the behaviour in question. As we saw, what is essential to dispositions is that they have a

---

[179] On this, see Pollard 2010.

[180] Doris and Gorman have made a similar point in their discussion on moral responsibility. On Doris's view, "Identification [or endorsement] may be said to obtain if a person *would have* identified with the determinative motive of her behaviour at the time of performance had she subjected it to reflective scrutiny. Accordingly, unreflective persons—as all of us are sometimes—may be quite legitimately responsible for their unreflective behaviours" (Doris 2002, p. 141). Similarly, for Gorman, "what's important… are the psychological dispositions and not the endorsement itself, the actual act of endorsement can be merely hypothetical" (Gorman 2019, p. 148).

characteristic manifestation and some triggering conditions. Here, the characteristic manifestation is the actual endorsement. The triggering condition, on the other hand, is the conscious awareness of the behaviour in question. Thus, to say that "I am disposed to endorse my behaviour" means that "if I were to become consciously aware of my behaviour, I would endorse it."

Consider again my habitual action of wiping my feet when entering the house. On the hypothetical version of the endorsement view, the behaviour in question counts as an action if I am disposed to endorse my behaviour. And this is actually the case: if I were to become consciously aware of my behaviour, I would endorse it. Moreover, we may add that my endorsement would be causally effective because I would be able to directly stop my behaviour if there was interference. Imagine that I have gum under my feet. In that case, I would be able to directly stop my motion, which suggests that my endorsement would be causally effective.

The third *desideratum* is that our solution to the problem of internal deviance should be compatible with the natural event-causal order. The question then is whether the endorsement view is compatible with this. As we saw, the endorsement view is a non-causal structural approach. That means that instead of looking at the way a behaviour is caused, the endorsement view looks at the way it is structured. That being said, the endorsement view does not deny that actions are caused. In particular, the endorsement view does not deny that actions are caused by prior events. The endorsement view is, in fact, compatible with different causal frameworks, such as the event-causal or the substance-causal frameworks.

While the endorsement view does not deny that actions are caused, it does deny that the causal history of an action is essential to it. In other words, it denies that we should look at the way an action is caused in order to define it. In that sense, actions are similar to events such as

the collapse of a bridge. The collapse of a bridge may be caused by an earthquake (another event), but nobody would want to say that the earthquake is essential to the collapse of the bridge or that, to define what the collapse of a bridge is, we need to identify its causal history. And the reason is simply that the collapse of a bridge may be caused in many different ways (it may be caused, for example, by a hurricane). Actions are similar with this respect. While they may be caused by prior events (such as prior mental events), this is not an essential feature of actions. What is essential is their structure.

In Chapter 6, we will see how the endorsement view fulfills the fourth, the fifth, and the sixth *desiderata*. Assuming that it can fulfill these *desiderata*, we may conclude that the endorsement view provides the most satisfactory solution to the problem of internal deviance, and thus that the account of control on which it is based (standby control) is the most compelling.

4.6. Conclusion

In this chapter we discussed the notion of control and saw that the problem of internal deviance suggests that the standard account is based on a flawed conception of control. We also examined the most popular solutions to the problem of internal deviance and concluded that the endorsement view, which is based on the non-causal concept of standby control, provides the most compelling solution and thus the most compelling account of control.

Chapter 5: The Disappearing Agent Objection

One of the most fundamental assumptions in the metaphysics of agency is that agency requires agential control. In the previous chapter, we examined different notions of control and we saw that the endorsement view, which is based on the non-causal notion of standby control, provides the strongest account of control. In the present chapter, we shall start a more in-depth discussion of what it is for *an agent* to control her behaviour – a discussion that will continue in the next chapter. Here, I shall argue that the standard account fails to provide an account of *agential* control.

On the standard account, to exercise agential control just consists in controlling one's behaviour by means of mental attitudes such as desires, beliefs and intentions. The idea here is that the agent, on the standard view, can be reduced to these mental attitudes. But according to some critiques, something went wrong in the reduction: the agent disappeared.[181] This critique is known as the "disappearing agent objection." What the critique suggests is that the standard account fails to provide an account of *agential* control.

The problem of the disappearing agent has not always been very well understood in the literature. For that reason, I will spend the entire chapter just to explain the issue. One of the most extensive discussions of the problem is developed by Schlosser in his "Agency, Ownership, and the Standard Theory" (2011). In my explanation of the objection, I will build on Schlosser's view and distinguish, as he does, two versions of the problem.[182] That being said, Schlosser also

---

[181] For a discussion of the objection, see Melden 1961, Chisholm 1964, Taylor 1966, Nagel 1986, Velleman 1992, Bratman 2001, Enç 2003, Mele 2003, Hornsby 2004, Schroeter 2004, Lowe 2008, Aguilar and Buckareff 2010, Schlosser 2011, and Steward 2013.

[182] Aguilar and Buckareff also make a distinction between two versions of the problem. See Aguilar and Buckareff 2010.

argues that the disappearing agent objection is not a real issue for the causal theory of action (or the event-causal approach). I think that Schlosser is mistaken, so I will explain why Schlosser's rejection of the objection has failed. I will therefore argue, *contra* Schlosser, that the standard account has failed to provide an account of *agential* control.

This chapter is divided into three sections. In Section 5.1, I provide an exposition of the disappearing agent objection as it is understood by Schlosser and explain why he dismisses the objection. In Section 5.2, I develop my own interpretation of the problem and explain how my reading differs from Schlosser's. In Section 5.3, I argue that Schlosser does not provide a satisfactory response to the problem of the disappearing agent when the objection is properly understood. Thus, I conclude that Schlosser's dismissal of the objection has failed and that the disappearing agent objection is a real issue for the standard account.

5.1. Schlosser's interpretation and responses

In this first section, I provide an exposition of Schlosser's interpretation of the disappearing agent objection as he developed it in "Agency, Ownership, and the Standard Theory" (2011). In that text, Schlosser identifies a distinction between two versions of the disappearing agent objection, which he calls "the challenge of disappearing agency" and "the challenge of disappearing agents."[183] After explicating these two problems, I will explain why Schlosser dismisses them.

[183] Schlosser 2011, p. 22.

To begin with, let us recall the metaphysical position on which the standard account is based. As we saw, on the standard account, an action can be defined as a behaviour that is caused by an intention and a desire-belief pair. Moreover, the standard account is a reductive view: it reduces the agent to mental attitudes such as intentions, desires and beliefs.

Schlosser identifies a first version of the disappearing agent objection, which he calls "the challenge of disappearing agency," and which he attributes to Melden and Nagel.[184] This objection is, on his view, a fundamental challenge: the critique is that "the event-causal theory altogether fails to capture the phenomenon of agency, as it reduces activity to mere happenings."[185] According to that objection, the agent is, on the event-causal approach, reduced to mere happening and for that reason an agent is a "*victim* of causal pushes and pulls" or a "*mere locus* in which events take place."[186]

Schlosser rejects the challenge of disappearing agency for a few reasons. In the following paragraphs, I shall focus on the most important points he has made. The first reason why he rejects the objection is that "its proponents have not produced a single *argument* to support their case, and they have certainly not identified a philosophical *problem*. Their case is entirely based on intuition, and in some cases on mere metaphor and rhetoric." As he mentions, Melden and Nagel sometimes uses the metaphor of a victim or bystander to make their point, but their argument is not based on anything other than intuitions.

The second reason why he rejects the challenge of disappearing agency is that the objection seems to deny that we can do things with our minds. Schlosser's example is that of remembering. He says that

---

[184] Melden 1961, Nagel 1986.
[185] Schlosser 2011, p. 22.
[186] Schlosser 2011, p. 22.

> Events may be called *happenings* in virtue of the fact that they *occur* in time. But the fact that events are occurrences does not entail or show that an agent's mental events and movements are things that *happen to* the agent, in the sense that they assail or befall the agent, or in the sense in which we say that a bad or unjust thing happened to us. When I remember something, for instance, I am a constitutive part of an event, but I am no victim or helpless bystander.

Thus, it seems like Melden and Nagel are committed to the view that mental events are passive and that we cannot do things with our mind (such as remembering, calculating, deciding, imagining, etc.), which is implausible.

Another reason why he rejects the objection is that it seems to suggest that the event-causal approach fails to capture agential control: "We can interpret this [critique] as saying that [the event-causal approach] fails to capture the fact that agents can exercise control over their behaviour."[187] But Schlosser argues that this view is false, because the causal theory of action can account for the distinction between behaviours that are controlled by the agent and those that are not controlled by the agent. For Schlosser, "an agent exercises control only if the behaviour in question is caused by mental states and events that rationalize its performance," and when deviant causal chains are excluded.[188] This account of agential control allows us to distinguish active and passive behaviours. For example, passive bodily motions (e.g., bodily spasms) are those motions that are not caused by mental states and events (or not caused in the right way by mental states and events). Active bodily motions, on the other hand, are those motions that are caused by mental states and events in the right way.

The second version of the objection is what Schlosser calls "the challenge of disappearing agents," an objection that he attributes to Velleman, Bratman, Enç and Schroeter.[189]

---

[187] Schlosser 2011, p. 13.
[188] Schlosser developed a solution to the problem of deviant causal chains. See Schlosser 2007.
[189] Velleman 1992, Bratman 2001, Enç 2003, Schroeter 2004.

As Schlosser construes it, "the challenge of disappearing agents" is a challenge concerning higher forms of agency such as full-blooded agency and autonomous agency. According to that objection, the causal theory of action can account for lower forms of agency but not higher forms. As Schlosser puts it, the objection says that the causal theory of action "fails to account for the agent's participation or proper role in the performance of higher kinds of agency."[190]

To response to that critique, Schlosser makes two main points. First, he accepts the claim that there are different grades of agency:

> It is very plausible to think that the aspects or kinds of human agency form a spectrum, or a hierarchy, from lower and basic to higher and more refined kinds of agency. At the bottom of this hierarchy one finds behaviour that is purposeful but to a high degree driven by environmental stimuli (such as instinctive, automatic or highly habitual reactions). Moving up the hierarchy we get intentional, rational, deliberative, reflective and self-controlled agency, and towards the top we find autonomous and free agency.[191]

But he adds that it is unhelpful to talk about grades of agency in terms of the participation or lack of participation of the agent: that "creates a bipartition that does not match up with the varieties of human behaviour."[192] In other words, there seems to be more than two grades of agency and the opposition between "the participation of the agent" and "the missing agent" only allows for two grades.

Second, Schlosser denies that the agent is missing in the lower kinds of agency: "The most natural thing to say, and the most natural assumption to begin with, is that all instances of agency involve an agent. Wherever there is agency, there is an agent participating, playing a role as the agent." Schlosser proposes another way to distinguish the higher forms of agency from the lower forms. Instead of being a distinction between the participation of the agent and the lack of

---

[190] Schlosser 2011, p. 27.
[191] Schlosser 2011, p. 27.
[192] Schlosser 2011, p. 27.

participation, we should say, according to Schlosser, that "the agent instantiates certain properties or exercises certain abilities [in higher kinds of agency], which are not instantiated or exercised in lower kinds of agency."[193]

5.2. An alternative interpretation

In the previous section, we examined Schlosser's interpretation of the disappearing agent objection. In the present section, I will develop an alternative interpretation. Just like Schlosser, I will distinguish two versions of the objection, but I will also argue that we can find two sub-versions of the first objection: a strong one and a weak one. Moreover, I will propose a different interpretation of the second objection: instead of focusing on the failure of the standard story to account for higher forms of agency, I shall focus on its failure to account for what is typically human about agency.

As we saw in the previous section, Schlosser identifies two versions of the objection, what he calls "the challenge of disappearing agency" and "the challenge of disappearing agents." In this section, I shall identify a similar distinction. What Schlosser calls "the challenge of disappearing agency" is what I will call "the missing source of activity objection." And what he calls "the challenge of disappearing agents" is what I will call "the missing humanity objection."

To introduce these two points, it is instructive to look at one particularly influential account of the agent, namely, Chisholm's account. Chisholm is often considered to be the modern father of the agent-causal view. On Chisholm's view, we have, as agents, "a prerogative

[193] Schlosser 2011, p. 28.

which some would attribute only to God: each of us, when we act, is a prime mover unmoved."[194] There are two important ideas here. The first is the idea that the agent is a prime mover unmoved. The agent is, for Chisholm, necessarily active and cannot be passive: she is a *source of activity*. The second idea is that the agent shares something with God. The agent has, for Chisholm, a God-like capacity: the ability to do and decide otherwise. This is what distinguishes human beings from non-human animals and explains why human beings are responsible for their actions while non-human animals are not. Accordingly, the agent is, for Chisholm, a distinctively *human* creature.

Thus, the agent is for Chisholm both a source of activity and a distinctively human creature. When critics say that the agent disappeared in the causal theory of action, they mean that there is either no source of activity or nothing distinctively (or typically) human. Let us take a look at each of these critiques, starting with the missing source of activity objection.

5.2.1. The missing source of activity objection

In the present section, I develop an interpretation of the first version of the problem of the disappearing agent, namely, the missing source of activity objection (what Schlosser calls "the challenge of disappearing agency.") I show that we can find two sub-versions of this objection: a strong and a weak version.[195] This is where my interpretation differs from Schlosser's.

A first critique of the causal theory of action is the strong version of the missing source of activity objection, as formulated by Melden, for example. On Melden's view, mental events such

---

[194] Chisholm 1964, p. 23.
[195] I attribute the strong version to Melden (1961) and Nagel (1986), and the weak version to Taylor (1966) and Frankfurt (1976).

as desires, beliefs, and intentions are mere happenings, not things that we do: "these causal factors [volitions, desires, interests, etc.] are happenings in, or to, me, rather than things that I do."[196] It is convenient here to use the notions of passivity and activity to distinguish a mere happening from an action. Melden's claim is that all mental events (such as beliefs, desires and intentions) are *passive*.

Now what issue arises from this exactly? Melden later argues that since desires, beliefs, and intentions are mere happenings, or passive, they cannot give rise to bodily actions; all that they can produce is other passive events: "A happening can only produce other happenings…It is futile to attempt to explain conduct through the causal efficacy of desire—all that can explain is further happenings, not actions performed by agents."[197] Thus, the causal theory of action seems implausible, because it entails that something passive gives rise to something active.

Melden's critique has been called the "problem of the disappearing agent" because the agent, which is for him *a source of activity*, has been reduced by causal theorists to something passive, to mere happenings. So, it seems that the agent has disappeared: in the causal theory of action, there are only passive happenings and no source of activity.

Melden's critique is what I call the *strong* version of the missing source of activity objection. It is a strong version because it says that *all* mental events are mere happenings or passive. Another critique of the causal theory of action is the *weak* version of the missing source of activity objection, a critique that we may attribute to R. Taylor and Frankfurt.[198] According to that view, only *some* mental events are passive. This objection says that the distinction between passive bodily movements and active bodily movements extends to the mind as well. Thus, there

---

[196] Melden 1961, p. 9.
[197] Melden 1961, p. 128.
[198] See Taylor 1966, p. 73 and Frankfurt 1976.

are passive mental happenings and there are mental actions. The problem is that the causal theory of action reduces the agent, a source of activity, to mental events that are *sometimes* passive. In those cases, the agent has disappeared after all. Thus, the *weak* version of the disappearing agent objection stipulates that the agent *sometimes* disappears in the causal theory of action.

Notice that Frankfurt makes this distinction. Indeed, he claims that "[i]n our intellectual processes, we may be active or passive,"[199] and that the distinction between "those movements of a person's body that are mere happenings in his history, and those that are his own activities" has "its analogues in the psychological domain."[200] Frankfurt has also provided some compelling examples of active and passive thoughts. Active thoughts include processes such as "[t]urning one's mind in a certain direction, or deliberating systematically about a problem."[201] Passive thoughts, on the other hand, comprise those "obsessional thoughts… that run willy-nilly through our heads."[202] Moreover, Frankfurt has argued that there are active and passive passions (or desires): "there is a useful distinction to be made, however awkward its expression, between passions with respect to which we are active and those with respect to which we are passive."[203]

Now Frankfurt does not really explain the consequences that his distinction has for the causal theory of action, but we may extrapolate on his view. I see two possible ways to articulate the problem that Frankfurt's distinction raises for the causal theory of action.

First, we may argue that Frankfurt's distinction can be used to provide a counterexample to the causal theory of action. Consider this case: I sometimes watch horror movies in which murderers are hiding in dark places to attack their victims. Every time I watch that kind of

---

[199] Frankfurt 1976, p. 59.
[200] Frankfurt 1976, p. 59.
[201] Frankfurt 1976, p. 59.
[202] Frankfurt 1976, p. 59.
[203] Frankfurt 1976, p. 60.

movie, I am overcome by the obsessive thought that someone is hiding in a closet and by the urge to make sure that nobody is hiding there. These passive mental elements may have some behavioural consequences: I might walk in direction of the closet and check what is in there. The problem here is that the behavioural consequences cannot be actions, because something passive cannot give rise to something active. To say that I "walked" in direction of the closet and "checked" what was in the closet suggests that I acted, but this is erroneous: we should rather say that I behaved passively. The issue is that all the conditions stipulated by the causal theory of action are fulfilled, and thus if the causal theory of action were correct, the result should be an action.[204] On this first option, what is required to fix the issue is a criterion that allows us to rule out cases that involve passive mental episodes.

The first option does not seem very plausible. Intuitively speaking, it seems more plausible to say that I am acting when I walk in direction of the closet and check what is in there, than to say that I am behaving passively. And that is so even if I am overcome by a passive thought and a passive desire. My behaviour is nothing like a bodily spasm or a corporeal tic. If I am acting, the example is not a counterexample to the causal theory of action. But then what is the problem exactly? This leads us to our second option.

According to the second option, passive mental events, such as passive thoughts and desires, may be involved in the causal production of an action. The problem is that a bodily motion cannot become active in virtue of these passive mental events: there must be something

---

[204] Bratman has made an argument along these lines. In "Two Problems About Human Agency" (2001), he draws a distinction between "(merely) motivated behaviour" and "action determined or governed by the agent" or "full-blown agency" (2001, pp. 311-312). Merely motivated behaviours include those cases in which the agent is "moved by desires … from whose role in action he is, as we sometimes say, estranged" (2001, p. 312). Merely motivated behaviours are caused by passive desires and for that reason, the agent who performs them "seems himself less the source of the activity than a locus of forces" (2001, p. 312). Notice that Bratman does not consider these behaviours to be actions.

else that confers activity on the bodily motion. What is required then is the addition of another element, an element that is necessarily active (such as the agent). What we need to identify is a reliable source of activity, something that will necessarily confer activity on bodily motions. Notice that what is required here is not an element that will rule out cases that involve passive mental episodes, because we conceded that passive mental episodes can be involved in the causal production of an action.

Thus, the problem with the causal theory of action is that it fails to identify a reliable source of activity that will necessarily confer activity to bodily motion. In response to this objection, a causal theorist might claim that it is the role of an intention to make a bodily event active. But this is not convincing. If it were the role of intentions to make bodily events active, then we would have to explain why intentions are active. And the only plausible response to this is that intentions are active in virtue of the belief and desire that cause them. But this is the problematic view with which we started. Beliefs and desires, because they can be passive, are not reliable sources of activity or things that necessarily confer activity.

We have here what I take to be the weak version of the missing source of activity objection. The problem, again, is that the causal theory of action assumes that beliefs and desires make bodily movements active (through an intermediate intention), and this is implausible considering that beliefs and desires are sometimes passive. Causal theorists have failed to identify a reliable source of activity. So, in cases such as my walking to the closet, the standard account fails to identify a source of activity. The agent – a source of activity – has disappeared.

5.2.2. The missing humanity objection

We just examined the first version of the disappearing agent objection, namely, the *missing source of activity* objection. Now we will examine the other version of the disappearing agent objection, which Schlosser calls "the challenge of disappearing agents" and which I call the "missing humanity objection." My interpretation differs importantly from Schlosser's since we focus on two different aspects of the problem: while Schlosser focuses on the failure of the standard story to account for higher forms of agency, I focus on its failure to account for what is typically human about agency.

The missing humanity objection was developed by Velleman in his influential paper "What Happens When Someone Acts?"[205] but it can also be attributed to Frankfurt before him. The critique in question is based on a set of distinctions. In "Freedom of the Will and the Concept of a Person" (1971), Frankfurt argues that what is distinctive about human beings is that they can form second-order desires: "it seems to be peculiarly characteristic of humans … that they are able to form what I shall call 'second-order desires' or 'desires of the second-order'."[206] Non-human animals, on the other hand, are for Frankfurt incapable of forming second-order desires because they lack what we might call "self-reflexivity," namely, the capacity to reflect on their mental states (e.g., their first-order desires). As Frankfurt puts it, "No animal other than man, however, appears to have the capacity for reflective self-evaluation that is manifested in the formation of second-order desires."[207]

Moreover, Frankfurt draws another distinction, that between second-order desires and second-order volitions: a second-order volition is, as we saw in Chapter 1, a desire to have a certain first-order desire as one's will. Then, he claims that second-order volitions are distinctive

---

[205] Velleman 1992.
[206] Frankfurt 1971, p. 12.
[207] Frankfurt 1971, p. 12.

of *persons*: "it is having second-order volitions, and not having second-order desires generally, that I regard as essential to being a person."[208] To the concept of a person, Frankfurt opposes the concept of a wanton. As he says, "[t]he class of wantons includes all nonhuman animals that have desires and very young children. Perhaps it also includes some adult human beings as well."[209] He also concedes that human beings are sometimes wantons: "humans may be more or less wanton; they may act wantonly, in response to first-order desires."[210]

Before I explain how Frankfurt's distinctions can be seen as a critique of the standard account, I would like to comment on them. The first distinction is that between human beings and non-human animals. On Frankfurt's view, the distinctive feature of human beings is self-reflexivity (the capacity to reflect on their mental attitudes); non-human animals lack self-reflexivity. However, whether or not non-human animals lack this characteristic is an empirical question: it cannot be settled by arm-chair philosophers. Secondly, Frankfurt's position might not be true. In recent studies on animal cognition, it has been argued that some non-human animals and human infants are capable of metacognition.[211] For example, it has been argued that dolphins,[212] rhesus monkeys,[213] great apes[214] and human infants[215] are aware of their own epistemic states: they know whether they know or do not know something. If that is correct, then we might think that some non-human animals are capable of forming second-order desires since they seem to possess self-awareness. A more careful distinction would be, then,

---

[208] Frankfurt 1971, p. 16.
[209] Frankfurt 1971, p. 16.
[210] Frankfurt 1971, p. 17.
[211] See Beran et al. 2012, Crystal and Foote 2009, Shettleworth and Sutton 2006, Proust 2013, Andrews 2016.
[212] Smithe *et al*. 1995
[213] Hampton 2001
[214] Call & Carpenter 2001
[215] Call & Carpenter 2001

the distinction between those creatures that possess a *typically* human feature (self-reflexivity) and those creatures that do not. This leaves room for the possibility that some non-human animals possess the feature in question.

The second distinction is the distinction between a wanton and a person. This second distinction has almost the same extension as the previous distinction, with the exception of one case: those human beings that form second-order desires but not second-order volitions. As an example, Frankfurt mentions the "physician engaged in psychotherapy with narcotics addicts" who wants to know what it is to have a compulsive desire for drugs, but who does not really want to be moved by the desire to take drugs.[216] When he is moved by his second-order desire, the physician is a human being but not a person. The difference between the physician and a person is that, while a person's second-order desires have an impact on what she does, the physician's second-order desire has no such effect.

The reason why Frankfurt introduces this nuance is because, as we saw in Chapters 1, 2, and 3, Frankfurt is interested in *agency*, and second-order desires that are not second-order volitions have no effect on what someone does. That being said, the distinction does not seem to capture something that is very prevalent in our human experience. For that reason, and for the sake of simplicity, we may leave it aside. In the rest of the thesis, I shall use the terms "human being" and "person" interchangeably when discussing Frankfurt's view.

Now that we have a better understanding of Frankfurt's distinction between a wanton and a person, we may wonder what implication it has for the causal theory of action. Frankfurt does not present his view as an objection to the causal theory, but it is not hard to see

---

[216] Frankfurt 1971, pp. 14-15.

how someone could read it as such. One might argue, following Frankfurt, that the causal theory of action only captures *wanton actions*. On the causal theory, an action involves first-order mental attitudes (beliefs, desires, intentions), but many non-human animals have these mental attitudes as well.[217] Thus, the causal theory of action seems to capture a form of agency that is not typically human. The problem is that we are expecting a theory of *human* action, a theory about the way human beings typically act. In the causal theory of action, there is nothing typically human: a *human* agent is missing.

There are two points to notice here. The first is that the missing humanity objection arises from a concern about moral responsibility. One of the main reasons why we tend to believe that a human action is distinct from a non-human action is because the former is typically open to moral assessment whereas the latter is not. The problem with the causal theory of action is that nothing in the theory can explain why human actions are typically open to moral assessment while non-human actions are not. Thus, the missing humanity objection could just as well be called the missing *moral being* objection.[218] As we saw in Section 4.1, the problem of natural agency is a problem concerning the compatibility between our moral perspective on the world, according to which we are morally responsible agents, and our scientific perspective on the world, according to which every event is caused by a previous event. The causal theory of action, in its attempt to develop an event-causal or scientific account of agency, has neglected the moral dimension (or the typically human dimension) of agency.

---

[217] On this, see for example Nagel 1977. Needless to say, this is a controversial issue within the philosophy of mind. Davidson might reply that the formation of mental attitudes requires linguistic competence, and thus that non-human animals do not have mental attitudes (See Davidson 1982; cf. Allen and Bekoff 1997).

[218] When formulated as such, the critique also leaves open the possibility that some non-human animals are moral creatures. In recent studies on animal cognition, it has been argued that some non-human animals are "full-blown moral agents." On this, see Bekoff and Pierce 2009, and Andrews 2016.

The second point to note concerns Frankfurt's account of wanton action. Frankfurt's notion of wanton action was developed in 1971, before he developed his "missing source of activity objection" in 1976. That is important because Frankfurt would have to conclude, from 1976 onward, that the causal theory of action fails to provide *both* an account of what is typically human about agency and an account of wanton actions. That is so because a wanton action should include a reliable source of activity, and the causal theory of action fails to include this reliable source of activity: first-order desires, beliefs, and intentions can be passive.

We just saw that Frankfurt's distinction between wantons and persons can be used to develop an objection against the causal theory of action. While Frankfurt did not explicitly present it as a critique of the standard account, Velleman, who was strongly influenced by Frankfurt's work, did. In "What Happens When Someone Acts?" Velleman explicitly argues that the causal theory of action fails to capture what is *distinctively* human about agency: "I shall argue that the standard story describes an action from which the distinctively human feature is missing."[219] As we just saw, this is not a very careful statement and it rests on empirical considerations. A more careful formulation would be that "the standard story describes an action from which the [*typically*] human feature is missing."

The problem, again, is that if we follow the causal theory of action we could attribute agency to many species of non-human animals. Indeed, it seems far from absurd to say that some animals have beliefs and desires, or even intentions. And thus, it seems far from absurd to claim that many animals behave in the way described by the causal theory. But what we are expecting from a theory of human agency is that it accounts for what is typically human about agency—

---

[219] Velleman 1992, p. 462.

and the causal theory fails to do this. Then we might say, alongside Velleman, that a *human agent* is missing in the causal theory of action.

In the elaboration of his view, Velleman has drawn a distinction between two kinds of actions: defective and full-blooded actions. Defective actions are, for Velleman, those actions that do not involve anything distinctively (or anything typically) human. Full-blooded actions, on the other hand, do involve something distinctively (or typically) human. Velleman's critique then is that the standard account only captures defective agency; it does not capture full-blooded agency. This is why we may interpret Velleman's critique, as Schlosser does, as saying that the standard story fails to account for higher forms of agency.

5.3. A critique of Schlosser's response

In the previous section, I developed an interpretation of the disappearing agent objection that differs from Schlosser's interpretation. In this section, I will explain why Schlosser's response to the objection has failed. In brief, the issue is that he has failed to take into considerations the aspects of the problems that I highlighted.

5.3.1. Schlosser's response to the missing source of activity objection

In this section, I will argue that Schlosser does not present an adequate response to the missing source of activity objection because he fails to consider the *weak* version of the objection. Moreover, I will examine a possible reply to my critique on behalf of Schlosser and conclude that the reply is unsatisfactory.

As we saw, there are two versions of the missing source of activity objection. On the strong version, *all* mental attitudes count as passive happenings. On the weak version, *some* mental attitudes count as passive. The issue, on both views, is that these passive attitudes cannot confer activity to behaviours (bodily motions for example) and thus that a source of activity seems to be missing in the standard account.

In response to this objection Schlosser develops three points. First, Schlosser argues that the objection is not based on a philosophical argument. This seems correct as a response to the strong version: the objection seems to be merely based on intuitions. But this is certainly not the case for the weak version. The weak version of the objection *is* based on a philosophical argument. More specifically, it is based on an argument from *consistency*. The argument goes as follow. First, the standard account assumes that events can be active or passive: on the standard view, a bodily action is an active bodily event and a corporeal spasm is a passive bodily event. If that is so, then we must conclude, *for the sake of consistency*, that mental events can also be active or passive since they are also events. The problem, then, is that passive mental events (such as thoughts and desires) may be involved in the causal production of an action, but they cannot confer activity to behaviours. Thus, the action must be active in virtue of something else. The problem with the standard account is that it fails to include this something else (a reliable source of activity).

Schlosser's second point is that the objection seems to deny that we do things with our minds. Again, this is an accurate critique of the *strong* version of the objection; indeed, on this version, mental events are mere happenings. Thus, the objection seems to deny the entire realm of mental actions, that is, everything that we can do with our minds. For example, it seems to deny the existence of mental actions such as deliberating, performing a mental calculation,

focusing on a problem, and so on. That being said, the weak version of the objection is not committed to this implausible position. On the weak version, there are *both* active and passive mental events. Thus, Schlosser's second point is not an adequate response to the weak version of the objection.

Finally, Schlosser's third point is that the causal theory of action can account for agential control in spite of what the objection suggests. For Schlosser, those mental states and events mentioned by the causal theory of action are bearers of agential control when they rationalize the behaviour that they cause and when they cause it in a non-deviant way: in other words, "our agency springs from our mental states and events."[220] This response, however, would not satisfy the proponents of the weak version of the objection. The reason is that mental states and events can be passive, and these passive mental states and events may rationalize a behaviour and cause it in a non-deviant way. But agency cannot spring from these passive mental attitudes. Something else must be involved, namely, a reliable source of activity.

Therefore, it seems like Schlosser's three objections fail to respond to the weak version of the missing source of activity objection. That being said, Schlosser might want to reply to my critique and claim that he has considered Frankfurt's objection. Frankfurt's objection is what he calls the "challenge from ownership." Here, I will explain why Schlosser misconstrued Frankfurt's objection and argue that he failed to provide a convincing response to it.

The objection that Schlosser attributes to Frankfurt goes as follow. It says that mental states and events cannot bear agential control or that they do "not guarantee agency, because the agent may not identify with being moved by those states and events."[221] In that sense, the

---

[220] Schlosser 2011, p. 26.
[221] Schlosser 2011, p. 24.

resulting behaviour "may still not be a true and proper expression of the *agent's own* agency."[222] As an example, Schlosser mentions Frankfurt's unwilling addict, who is moved by a desire to take drugs even though he does not identify with his desire. On Schlosser's reading of Frankfurt, the behaviour of the unwilling addict does not count as a true expression of his own agency.

In response to this "challenge from ownership," Schlosser presents three points. Below, I will examine each of these points and explain why I reject them. First, Schlosser claims that Frankfurt's unwilling addict (the addict who does not want his desire to take drugs to be effective) is not behaving passively but acting. He says that the "unwilling addict is not a mere bystander or locus in the flow of events. He is capable of a good degree of control and agency, and he exercises this ability in the pursuit of drugs."[223] In that sense, the behaviour of the unwilling addict must be an expression of his own agency.

I agree with this point. When discussing Frankfurt's objection in Section 5.2.1., I conceded that passive mental events (such as the addict's compulsive desire to take drugs) may be involved in the causal production of an action. But the point is not that the willing addict is behaving passively. The point is that his desire to take drugs is passive, and thus that the addict is acting *in virtue of something other than his desire to take drugs*. The problem is that the causal theory of action has failed to identify that something else. Thus, Schlosser's first point does not provide an adequate response to Frankfurt.

Schlosser's second point is similar to the first. He claims that the addict's desire and the resulting behaviour must be "his *own* in some basic or minimal sense." To illustrate his view, he contrasts the behaviour of the willing addict to some behaviour that we would definitely *not*

---

[222] Schlosser 2011, p. 24.
[223] Schlosser 2011, p. 25.

attribute to the agent: "serious cases of schizophrenia where patients report that their actions are under the control of some external agent" and cases of anarchic hand syndrome "where patients report that one of their hands moves on its own."[224] Schlosser's point, then, is that the behaviour of the unwilling addict appears to be his own in a minimal sense, compared to the behaviour in these other cases.

I agree with the claim that the behaviour of the unwilling addict is his own. This is equivalent to saying, for Frankfurt, that the behaviour is active or that it is an action,[225] and I already conceded that point. But Frankfurt's point is not that the addict's behaviour is not his own. His point is that the addict's behaviour cannot be his own in virtue of the desire that causes it, because the desire in question is not itself his own. There must be another "source of ownership" and the causal theory has failed to identify that other source of ownership. Thus, Schlosser's second point is also unsatisfactory.

Schlosser's third point is that the challenge from ownership is not a fundamental challenge because it is a challenge for a higher form of agency or a challenge for *autonomous agency*. It is not a challenge for agency as such.

On that point, I very much disagree with Schlosser. In Chapter 2, I showed that autonomy is, for Frankfurt, a matter of mental activity. Thus, to say that the challenge from ownership concerns autonomy and not agency as such does not solve the issue since autonomy is cashed out, for Frankfurt, in agential terms. Second, the challenge from ownership is, in fact, a fundamental challenge for agency: it is the challenge of the prime mover unmoved or the challenge of the origin of agency. To understand why, we need to understand the relation

---

[224] Schlosser p. 25.
[225] I will say more on the equivalence between ownership and activity below.

between ownership and activity. In "Identification and Externality," Frankfurt invokes the Aristotelian principle according to which "a thing is active with respect to events whose moving principle is inside of it."[226] Thus, activity is linked to internality. Frankfurt goes on to argue that a desire is internal when the person *identifies with it* or *makes it her own*: when a person identifies with a desire, she makes the desire internal to her self. The result is that the desire becomes active, in virtue of the Aristotelian principle. Thus, Frankfurt's notion of identification or ownership is a solution to the problem of the origin of agency: it explains how our desires become active and able to confer activity on our bodily motion.

That being said, we should make clear that actions do not necessarily involve a decision or an act of identification, like the action of the unwilling addict or the action of the wanton. To account for these cases, as I have argued in Chapters 2 and 4 (see Section 4.5.1 in particular), Frankfurt seems to introduce another kind of act of will: a choice, which is a kind of first-order endorsement through which the agent confers activity to her bodily motion. Thus, to account for the action of the unwilling addict, we might suppose that there is another act of endorsement (a first-level act of endorsement).

5.3.2. Schlosser's response to the missing humanity objection

In the previous section, we saw that Schlosser's response to the missing source of activity objection failed. In the present section, I shall examine his response to the disappearing humanity objection. I shall argue that it also fails and that this is so for two main reasons. First, Schlosser

[226] See Frankfurt 1976, p. 59.

did not correctly identify the main issue. Second, the distinction between two grades of agency is not as problematic as Schlosser suggests.

According to the missing humanity objection, the standard account has failed to identify a typically human form of agency. On Frankfurt's view, the standard account has identified what it is for a wanton to act, and not what it is for a human being or a person to act. On Velleman's view, the standard account has identified a form of defective action, not a form of human action *par excellence* or full-blooded action. In response to that critique, Schlosser presents two points. First, he argues that the bipartite distinction is problematic because there is a wide spectrum of grades of agency. Second, he argues that we should not draw a distinction between the participation and the lack of participation of the agent, because an agent must be involved in all actions.

This response is problematic for two reasons. First, it is problematic because it focuses on a secondary aspect of the objection: the distinction between wanton/defective action and human/full-blooded action. But the objection does not depend on this. The objection is first and foremost that the standard account has failed to include something typically human. Thus, one may argue, as I will do in the next chapter,[227] that we should abandon the distinction between wanton and human action, and that even when we do one may still believe that the standard account is flawed because it fails to include something typically human. On that view, all actions performed by (full-grown) humans involve something typically human.

Schlosser's response is also problematic for another reason. Schlosser argues that there are more than two grades of agency and that the distinction between wanton action and human action does not adequately capture it. But Schlosser includes in these grades of agency things

[227] See sections 6.1. and 6.3.2.

like instinctive, automatic, intentional, rational, self-controlled, free, and autonomous actions. For that matter, we may also include virtuous actions such as courageous and generous actions. The issue is that Schlosser does not identify grades of agency *qua* agency, while the distinction that Frankfurt and Velleman are drawing is a distinction between two grades of agency *qua* agency. Let me explain that point.

Agency is often cashed out in terms of agential control, and both Frankfurt and Velleman seem to endorse that view. Control is obviously a quantitative concept: we may have more or less control over something or we may have control over one thing or over many things. Since control is quantitative, we may identify different grades of agency that correspond to the level of agential control that we have. Frankfurt's and Velleman's accounts of agency are based on the idea that there are two broad categories of control: we may have control over our bodily motion and control over our mental happenings. When we have control over our bodily motion only, we are not very different from other animals. This is the first grade of agency. When we have control over our bodily motion *and* over our mind, we exhibit something more, something typically human: the capacity to reflect on and control our mental processes. This is the higher grade of agency. Thus, it seems like the notion of two grades of agency (*qua* agency) is fully coherent, because there are two sorts of thing over which we may have agential control: our body and our mind.

Schlosser's last point is that we should not talk about grades of agency in terms of the participation or lack of participation of the agent because "the most natural assumption to begin with, is that all instances of agency involve an agent."[228] But this is not the kind of claim that should worry Velleman, since it rests on a disagreement over terminology. If we construe the

---

[228] Schlosser 2011, pp. 27-28.

agent as a source of activity, like Chisholm does, then of course all instance of agency involve an agent. But if we construe the agent as a typically human creature, a creature that can control her mind or a creature that can "intervene between reasons and intention, and between intention and bodily movements,"[229] then it could be the case that some actions involve an agent and some do not. For that reason, Velleman's critique of the standard account remains in force: his point is simply that the standard account fails to include a *typically human* being.

5.4. Conclusion

In the present chapter, we examined the disappearing agent objection, which is the claim that the standard story fails to account for the "agential" dimension of "agential control." We saw that there are two ways to formulate the objection, which correspond to two different ways of conceptualizing the agent. On the first objection, the standard account fails to include a *source of activity*. On the second objection, the standard story fails to include a *typically human being* (or a *moral being*). We also saw that Schlosser's attempt to respond to the objection fails, and thus that the objection remains a real issue for the standard account.

[229] Velleman 1992, p. 463.

Chapter 6: The Endorsement View and the Agent

As we saw, one of the main assumptions in the metaphysics of agency is that agency requires agential control. In Chapter 4, we focused on the concept of control and we saw that the endorsement view, which is based on the concept of standby control, provides the most compelling account of control. In Chapter 5, we started a discussion of what it is for *an agent* to control her behaviour and we saw that the causal theory of action has failed to account for it. In the present chapter, we shall pursue that discussion. More specifically, I will defend a variation of Frankfurt's endorsement view, which is a non-causal view, and argue that it provides the most compelling account of what it is for *an agent* to control her behaviour.

According to the standard account, the agent can be reduced to mental attitudes such as desires, beliefs and intentions. We saw two issues with that view. First, desires and beliefs can be passive, so the agent, which is a source of activity, cannot be reduced to them. This is what I call the "missing source of activity objection." Second, desires and beliefs are not characteristic of human beings, so a *human* agent cannot be reduced to them. This is what I call the "missing humanity objection." These two objections constitute two versions of the disappearing agent objection.

There have been many attempts to solve the disappearing agent objection, and thereby to develop a more satisfactory account of *agential* control. As we saw, we can identify three main approaches in the metaphysics of agency, namely, the event-causal, the agent-causal, and the non-causal approaches. The proponents of each of these approaches have developed their own account of the agent, and thus their own solution to the problem of the disappearing agent. In this chapter, I will examine the most influential attempts made by the proponents of each of these

approaches and I will argue that my version of the endorsement view, which is a non-causal approach, provides the strongest solution.

This chapter is divided into three sections. In Section 6.1, I examine one prominent event-causal solution to the disappearing agent objection, that of Velleman, and conclude that it is not satisfactory. In Section 6.2, I examine two types of agent-causal solutions and conclude, similarly, that they are not satisfactory. Finally, in Section 6.3, I examine the non-causal solution. I start with an exposition of Frankfurt's non-causal solution (Section 6.3.1), which is a volitional and hierarchical view, and show some of its weaknesses. Then (Section 6.3.2), I develop my own solution and argue that it avoids the difficulties to which Frankfurt's account is subject. My own version of the endorsement view is a *dispositional* and *dialogical* view.

6.1. The event-causal solution

In the previous chapter, we discussed the disappearing agent objection, which suggests that the standard account is not based on a satisfactory account of *agential* control. One might argue that in order to account for agential control, or for the proper role of the agent, we simply need to refine the standard account or the event-causal approach. In this section, I examine one prominent attempt at refining the event-causal approach, Velleman's attempt, and argue that it is unsatisfactory.

In his paper "What Happens When Someone Acts?" Velleman attempts to solve the disappearing agent objection using the event-causal framework. The first thing to mention is that Velleman agrees with the picture of action proposed by the causal theory of action. For him, an action is a bodily event that is caused by an intention and a relevant belief-desire pair. But

Velleman argues that something is missing from this picture, namely, whatever plays the role of the agent. For Velleman, the role of the agent is the role of an intervener. In a full-blooded action, an agent *forms* an intention and *translates* that intention into action: "The agent thus has at least two roles to play: he forms an intention under the influence of reasons for acting, and he produces behaviour pursuant to that intention."[230] Thus, the element that plays the role of the agent is an intermediary between reasons (desires and beliefs) and intentions, and between intentions and bodily motions.

This story of human action is in fact Bishop's story, the one that we examined in Section 4.4. On Bishop's view, an action involves a desire-belief pair, an intention, a bodily motion and the *inferring* and *carrying out* of the intention. But unlike Bishop, Velleman believes that we can provide a reductive account of the role of the agent.

As we saw, the role of the agent is for Velleman that of an intervener. Moreover, an intervention is for him a process of practical reflection. First, when an agent engages in practical reflection, she considers her reasons for action (desires and beliefs) and forms an intention as a result. Second, through practical reflection, an agent calculates the strength of her reasons in order to ensure that a certain bodily motion is produced. That being said, the agent is not reducible to this process of practical reflection for Velleman. Practical reflection is itself an action, and to reduce the agent to it would mean that he has failed to provide a conceptual reduction.[231] Therefore, Velleman argues that we should look at the motive behind practical reflection, the motive that drives practical thought. This motive is a "desire to act in accordance

---

[230] Velleman 1992, p. 462.

[231] Mele claims that reducing the agent to the act of practical reflection would generate a regress: "If all actions are caused by their agents, an action that is supposed to play the role of the agent should itself have an agent as its cause, and Velleman therefore would have another instance of agent causation to reduce." (Mele 2003, p. 225)

with reasons." On Velleman's view, the agent's participation amounts to this desire being present in action.

Let us now examine whether this account of the agent can solve the disappearing agent objection, starting with the first version of the objection: the missing source of activity objection. We saw that according to one plausible version of the objection, desires and beliefs are sometimes passive and thus we cannot reduce the agent, a source of activity, to these passive mental elements. It seems obvious that Velleman's proposal is subject to that objection. Velleman wants to reduce the agent's participation to a desire to act in accordance with reasons. But such a desire, just like any other desire, may be passive: a person may be driven by an obsessive desire to act rationally. Thus, adding this desire to the causal chain may fail to account for the active nature of action. If the desire to act in accordance with reasons is passive, it will only incorporate another passive contender. For that reason, Velleman's view does not solve the missing source of activity objection (on the weak version).

And what about the missing humanity objection? Velleman does have a solution to that objection, but it is important to see that it is not very different from Frankfurt's. This should be no surprise, since Velleman himself claims that his approach is simply a modification of Frankfurt's approach.[232] As we saw in Section 5.2.2, Frankfurt believes that self-reflexivity, which is a form of reflection on our mental attitudes, is a typically human feature of agency. Velleman's notion of practical reflection is not very different, since it is, for him, a form of reflection on our reasons (desires and beliefs) and intentions. Arguably, this form of reflection is something that human beings typically do. Most non-human animals do not seem to intervene

[232] Velleman 1992, p. 476.

with or reflect on their motives; they seem simply to be driven by their motives.[233] If that is so, then we could suppose that the motive behind practical reflection, the "desire to act in accordance with reason," is something typically human. Thus, Velleman's view seems to account for a form of agency that is typically human and to provide a solution to the disappearing humanity objection.

The question now is whether this solution is satisfactory. I shall argue that it is not, for the same reason that Frankfurt's solution is not satisfactory. The reason why both views are unsatisfactory is because they entail that when a human being does not manifest self-reflexivity in her action, she behaves like a non-human being. In other words, a human action lacking self-reflexivity (what I shall call, in a somewhat paradoxical way, a "human wanton action") counts as non-human, since it possesses nothing typically human on Frankfurt's and Velleman's view. This position is problematic. To see why, it is useful to take a look at the literature on moral responsibility and in particular at a popular critique of Frankfurt's account of moral responsibility. The reason why this literature is relevant here is because, as we saw in Section 5.2.2, the missing humanity objection is grounded in a concern for moral responsibility: the objection is based on the idea that human actions are distinct from non-human actions because, contrary to non-human actions, the former are typically open to moral assessment.

On Frankfurt's view, moral responsibility requires self-reflexivity or the formation of a second-order volition. More precisely, it requires a harmony or "mesh" between a person's second-order volitions and her first-order desire.[234] This view has been widely criticized with one

---

[233] This is not to say that all non-human animals lack that capacity. As we saw, in recent studies on animal cognition it has been argued that some animals are capable of metacognition and thus that they possess self-reflexivity.
[234] Frankfurt claims that the "assumption [that a person is morally responsible for what he has done] *does* entail that the person did what he did freely, or that he did it of his own free will." And then he claims that a person has done something freely and of his own free will when the person "has done what he wanted to do, that he did it because he

common critique being that it is under-inclusive: it leaves out too many actions for which we are morally responsible, such as akratic actions[235] and human wanton actions.[236] Here, I shall focus on the problem of human wanton actions. Consider Hagi's example of Poppy:

> Poppy finds herself with a first-order desire to take the drug, a desire which she can resist. Suppose that initially, she has no second-order volitions concerning this desire. She reflects for a while. She is aware that taking the drug isn't good for her, but decides that she should indulge anyway. Pretty clearly, none of Poppy's simple reflections over whether she should take the drug *need* be 'second-order;" they could be, but they needn't be. Assume, then, that none of her reflections are second-order, and assume, finally, that she acts on her decision. It seems that Poppy could, under these circumstances, be morally appraisable for taking the drug: there are no "appraisability-undermining factors" in the actual sequence of events that culminate in Poppy's taking the drug.[237]

Hagi's point is that moral responsibility does not seem to require self-reflexivity or a form of reflection on the person's mental attitudes: it is enough for moral responsibility that the person reflects on her *action*. This indicates that Frankfurt's account of moral responsibility is flawed.

A similar point about human wantonness is developed by Fischer. On Fischer's view, "an individual might well be criticizable precisely for failing to form any second-order volitions. That is, an individual could in principle be morally blameworthy for failing to care about which first-order desire moves him to action."[238] On Fischer's view, an individual who fails to form a second-order volition when she should have and could have done so might be considered morally responsible. As an example, consider McKenna and Van Schoelandt's example of Gluttonia. Gluttonia is capable of caring about her first-order desires and of forming second-order volitions, but there is an area of her life where she simply does not care (although she should): she does not

---

wanted to do it, and that the will by which he was moved when he did it was his will because it was the will he wanted." (See Frankfurt 1971, p. 24)

[235] For this critique, see Vihvelin 1994, Hagi 1998, 2002, Fischer 2012a, 2012b, McKenna 2011, McKenna and Van Schoelandt 2015, Strabbing 2016, and Gorman 2019.

[236] For this critique, see Hagi 1998, Doris 2002, Fischer 2012b, and McKenna and Van Schoelandt 2015.

[237] Hagi 1998, pp. 71-72.

[238] Fischer 2012b, p. 135.

care about whether or not she is moved by a desire for food and drink. Further, Gluttonia "binge eats or drinks booze recklessly, since she has no high-order preferences about whether her efficacious desires are efficacious."[239] Now, should we think that Gluttonia's wanton action is not open to moral assessment because it lacks self-reflexivity and a second-order volition? McKenna and Van Schoelandt claim that this would be an implausible conclusion, and I agree with them.

What does this critique tell us about Frankfurt's and Velleman's solution to the missing humanity objection? The previous critique suggests that human wanton actions (or Velleman's defective actions) are open to moral assessment, even though they lack self-reflexivity. And since we usually reserve moral assessments for typically human actions, it also suggests that human wanton actions *do have* characteristics that are typically human. If that is correct, then it seems that Frankfurt's and Velleman's solution to the missing humanity objection is unsatisfactory because it entails the opposite view. In locating humanity in self-reflexivity, Frankfurt and Velleman have to endorse the implausible conclusion that all action lacking self-reflexivity is a kind of non-human action.

To make this point even more concrete, consider Velleman's example of his interaction with an old friend:

> Suppose that I have a long-anticipated meeting with an old friend for the purpose of resolving some minor difference; but that as we talk, his offhand comments provoke me to raise my voice in progressively sharper replies, until we part in anger. Later reflection leads me to realize that accumulated grievances had crystallized in my mind, during the weeks before our meeting, into a resolution to sever our friendship over the matter at hand, and that this resolution is what gave the hurtful edge to my remarks.

---

[239] McKenna and Van Schoelandt 2015, pp. 50-51.

This is, according to Velleman, a case of defective or wanton action, since it does not involve the intervention of the agent or a process of self-reflection. The question, however, is whether this action is open to moral assessment. When Velleman makes some hurtful remarks, is he open to moral assessment? My intuition in this case is that he is, and thus that there must be something typically human about his behaviour.

As we saw, human wanton actions seem to be open to moral assessment and, for that reason, they seem to possess a typically human feature. But what feature can that be? One plausible answer is that the feature in question has to do with the fact, if it is a fact, that human beings are typically *answerable* for their wanton actions, unlike most non-human animals. Thus, one plausible alternative rests on the notion of answerability. The latter notion has been developed in recent works on moral responsibility,[240] and I shall discuss it in more detail in Section 6.3.2. For now, some general considerations will suffice.

To say that human beings are typically answerable for their wanton actions just means that if they were pressed to provide an explanation or to justify their behaviour, it would be reasonable to expect them to do so. And this seems to be on point. Consider Velleman's example of his hurtful remarks and Hagi's example of Poppy. To say that Velleman is answerable for his action just means that the why-question ("why did you φ?") applies or that it makes sense to ask Velleman the question. And this seems to be the case. It is clearly conceivable that his old friend might reasonably ask him why he made these remarks; the question, moreover, would seem to be rightly asked. Similarly, one may ask Poppy, the drug taker, why she did what she did. Here again, it is clearly conceivable to ask her to justify or explain her behaviour and the question

[240] For a development of this view, see Hieronymi 2008, 2009, 2014; Shoemaker 2011, 2015; Pereboom 2014; Smith 2015.

appears to be rightly asked. Thus, Velleman and Poppy seem to be answerable for their wanton actions. If that is so, then these actions seem to possess a typically human feature.

From this discussion we may conclude that Velleman's event-causal solution to the problem of the disappearing agent is unsatisfactory. We may also identify one other *desideratum* that we are expecting from our account of agential control. In Chapter 4, we identified six *desiderata*. The seventh *desideratum* is that our account should avoid treating those human wanton actions as a kind of non-human action.

6.2. The agent-causal solution

In this section, I will discuss the two types of agent-causal views, namely, the ontologically non-reductive and the conceptually non-reductive views. The question is whether they can provide a satisfactory response to the missing source of activity objection and to the missing humanity objection. I will argue that their solutions to both these problems are unsatisfactory.

Let me start with the ontologically non-reductive agent-causal view. On this view, agent-causation is ontologically fundamental, which means that the agent is an irreducible substance. Chisholm was among the first to develop an ontologically non-reductive account. On his view, all actions necessarily involve an agent, conceived of as an irreducible substance. The agent, moreover, has a "prerogative which some would attribute only to God: each of us, when we act,

is a prime mover unmoved."[241] For Chisholm, the agent is both a necessarily active substance and a god-like creature.

While Chisholm's view can solve the two problems at stake, it does have some very unappealing consequences. Let me start with the first problem, the missing source of activity objection. By requiring the presence of an agent as a necessarily active substance, Chisholm ensures that all actions involve a source of activity or an active principle. Thus, this approach quite obviously avoids the weak version of the disappearing agent objection. That being said, Chisholm's solution also entails circularity, since an action is defined as an event caused by an active substance. As we saw in Section 4.4, this is why the agent causal view is often seen as mysterious: on that view, agency remains unanalyzed.

The second objection at stake is the missing humanity objection (or the missing moral being objection). Chisholm's view can also easily solve the issue, but his solution generates some unappealing consequences. On Chisholm's view, the agent is a god-like or supernatural entity. This view draws a sharp distinction between human agents and other non-human beings: only human beings (and god) enjoy the prerogative of being a prime mover unmoved, that is, the power to do or not do something which makes us morally responsible for what we do. Animals, on the other hand, are neither responsible for what they do nor do they possess this two-way power. For this reason, Chisholm's view can be said to capture something that is typically human about agency. That being said, Chisholm's view entail that human beings have a supernatural power, which is a view that most philosophers are not willing to endorse nowadays.

Contemporary proponents of the substance-causal view have generally abandoned Chisholm's supernatural view and attempted to naturalize substance-causation. In other words,

---

[241] Chisholm 1964, p. 23.

proponents of the ontologically non-reductive view have attempted to develop an account of agency that it compatible with the natural order, but the natural order, on their view, is conceived differently: it is not the natural *event*-causal order endorsed by the natural science, but a natural order based on the notion of substance-causation (or a kind of power-causal view). As we saw, one such view is Mayr's. On Mayr's account, an action always involves an irreducibly active causal power which stands for the agent, but this causal power belongs to the natural realm.

Just like Chisholm's view, Mayr's account can easily solve the weak version of the missing source of activity objection, but at the cost of circularity. By requiring the presence of an active power, Mayr makes sure that all actions involve a source of activity. But this entails that his view is circular: it defines agency by means of an *active* power. For that reason, agency remains unanalyzed and mysterious.

As to the missing humanity (or the missing moral being) objection, it is not clear that Mayr's account can solve it. For Chisholm, the missing humanity objection can be solved by endorsing the view that agent-causation is a supernatural phenomenon. When rejecting this view, Mayr seems to be vulnerable to the objection that Chisholm's supernatural account was meant to solve. Indeed, Mayr defines the active causal power necessary for action as the "abilities for physical action" or the "abilities to act *at will*."[242] But that kind of ability, such as the ability to raise one's arm at will, seem to be far from unique to human beings. Thus, it seems like Mayr fails to provide an account of agency that captures what is typically human about agency.

The second type of agent-causal view is the conceptually non-reductive view. On that view, agent-causation is conceptually irreducible but not ontologically fundamental. According

[242] Mayr 2011, p. 222.

to this second view, agent-causation is "realized in causation by events"[243] but no reductive analysis is possible. As we saw, one example of this second type of agent-causal view is Bishop's. Bishop in fact endorses the standard account of agency (or the causal theory of action), but he adds that agents *infer* (proximate) intentions from other mental items and that they *carry out* their (proximate) intentions.[244] This inferring and carrying out are ontologically reducible to events, but they are not conceptually reducible: they are intrinsically active events.

Bishop's view can avoid the missing source of activity objection, but his solution is not satisfactory. First, Bishop correctly assumes that the first objection is not that the standard account includes no agent as an ontologically irreducible substance. The issue is rather that there may be no active principle, no source of activity, in the causal theory of action. Thus, Bishop rightly assumes that if we identify an active event that occurs in all cases of action, we can solve the problem. Thus, when Bishop claims that all actions must include an inferring and a carrying out, and that these are intrinsically active events, he can successfully respond to the missing source of activity objection.

That being said, Bishop's solution is subject to some difficulties, which we discussed in Section 4.4. The first is that Bishop's view is *ad hoc*: Bishop does not provide any good reason to believe that inferrings and carryings out are unanalyzable (other than to avoid the problem it is meant to solve). Second, his view is circular: it appeals to the notion of an intrinsically active doing to define agency. For that reason, it remains obscure. Third, Bishop's view is not unified: bodily actions can be reduced to simpler elements while the acts of inferring and carrying out cannot.

---

[243] Clarke 2010, p. 218.
[244] Bishop 1983, p. 74.

As for the missing humanity objection, Bishop's view seems to be able to solve it for the same reason that Velleman's account does, but it is subject to the same difficulty as Velleman's view. As we saw, Bishop and Velleman endorse a similar view, since they both claim that an element is missing from the causal theory of action, and this element is whatever connects a belief-desire pair to an intention, and an intention to a bodily motion (on Bishop's view, this is an *inferring* and a *carrying out*). For Bishop, human beings are not simply moved by desires and beliefs; they typically take an external perspective on these motives and translate them. And this self-reflective perspective is one that most non-human animals seem to lack. This is why Bishop's account of agency can be said to be an account of *human* agency as such, and thus to avoid the missing humanity objection.

As we saw in the previous section, the problem with this solution is that human wanton actions, those actions that are performed by human beings and that lack self-reflexivity (which are mere behaviours for Bishop), count as non-human and this seems implausible considering that they are typically open to moral assessment.

Before we move on to the non-causal solution, some remarks about Bishop's view are in order. One might wonder why Bishop does not claim that inferring and carrying out can be analyzed in the same way that he analyzes bodily actions. We saw one reason for that in Chapter 4: to avoid the introduction of temporal gaps. Indeed, if inferrings and carryings out were analyzable in terms of events and their causes, it would reintroduce the possibility of causal deviance. But there is another reason why Bishop claims that inferrings and carryings out cannot be analyzed in the same way that he analyzes bodily actions: he wants to avoid the regress objection. Indeed, if inferrings and carryings out could be analyzed in the same way, then we could reduce them to a belief-desire pair, an intention, a mental event, and to an inferring and a

carrying out. But the second inferring and carrying out, being themselves actions, could be reduced to the same elements and this would generate an infinite regress.

There is one important lesson to remember from that discussion. When we appeal to the concept of an active mental doing to define agency (a mental act such as an act of inferring, an act of carrying out or an act of will), we have to face a dilemma. On the one hand, we can say that the mental act in question is analyzable and reducible, but that position generates a regress. On the other hand, we can say that the mental act in question is unanalyzable, but that view is circular and creates a disunity. Understanding the nature of this dilemma is crucial to grasp Frankfurt's solution and my own solution to the problem of the disappearing agent, to which I shall now turn.

6.3. The non-causal solution

We just examined some event-causal and agent-causal solutions to the disappearing agent objection. As we saw, these solutions are not very satisfactory. In the present section, I examine the non-causal solution, or more specifically, the structural view. In the first sub-section, I examine Frankfurt's non-causal solution to the problem of the disappearing agent and argue that it is unsatisfactory. Next, I develop an alternative non-causal solution and argue that it is the most compelling solution to the problem of the disappearing agent.

6.3.1. Frankfurt's endorsement view: a volitional and hierarchical account

In this first sub-section, I discuss Frankfurt's endorsement view and explain how it solves the disappearing agent objection. I argue that Frankfurt's solution is a volitional and hierarchical one. More specifically, I show that Frankfurt develops a *reductive volitional* account of agency to solve the missing source of activity objection, and I argue that he develops a *hierarchical* account to solve the missing humanity objection. Finally, I explain why Frankfurt's account is unsatisfactory: because his solution generates a problematic regress and because it treats, just like Velleman's account does, human wanton actions as non-human actions.

As we saw, Frankfurt's non-causal account of agency can be characterized as a structural view. The main idea of the structural view is that we should look not at the way an action is caused or at the nature of reasons for action, but at the way an action is structured. As we saw in Chapter 2, Frankfurt draws a distinction between two kinds of action: the actions of a person and the actions of a wanton. On Frankfurt's view, the action *of a person* is composed of a complex structure involving a bodily motion, a first-order desire, and the agent's endorsement of the first-order desire. The action *of a wanton*, on the other hand, simply involves a bodily motion and the agent's endorsement of the bodily motion. This last claim is admittedly much more controversial, but we need not be bothered with interpretive issues here.[245]

One question arises at this point: what is it exactly to endorse a bodily or mental motion? One of Frankfurt answers to that question is that an endorsement is an act of will, either a decision or a choice. To endorse *a desire*, on his view, is to decide to include it among the candidates for satisfaction.[246] To endorse *a bodily motion*, on the other hand, amounts to something like choosing to make it your own. That second point is, again, much more

---

[245] As we saw, the main aim of Part II is to develop my own account of agency. For that reason, I take much more freedom in my interpretation of Frankfurt.
[246] Frankfurt 1976.

controversial. But I take it to follow from Frankfurt's discussion of the distinction between a decision and a choice. As Frankfurt says, a decision is a second-order mental act while a choice is a first-order mental act: "making a decision is something that we do *to ourselves*. In this respect it differs fundamentally from making a choice, the immediate object of which is not the chooser but whatever it is that he chooses."[247]

The first question is whether Frankfurt's volitional account of endorsement can provide a solution to the missing source of activity objection. The issue is that an account of agency should include a reliable source of activity. As we saw, the standard account seems to fail to do so because it appeals to desires, beliefs and intentions, all of which could be passive. Does the same issue arise with Frankfurt's view? It seems like Frankfurt's view easily avoids the issue since the important concept is, for him, that of endorsement, and endorsement is an act of will – something necessarily active.

But this means that Frankfurt's view is facing the dilemma that we just discussed. The first option of the dilemma, which is the one Frankfurt opts for, is to say that acts of will are analyzable and reducible. As we saw, this option generates a regress, which I will discuss in more detail below. The second option is to say that acts of will are unanalyzable and irreducible. In that case, we do not generate a regress, but we end up with a circular and disunified account. As noted, Frankfurt opts for the first option, and thus develops a reductive account of acts of will. The most interesting development of this view is to be found in a paper that has been largely overlooked in the secondary literature: "Concerning the Freedom and Limits of the Will" (1989). In this text, Frankfurt explains why decisions and choices are necessarily active. It is

[247] Frankfurt 1987, p. 172.

useful to distinguish two steps in his argumentation. In the next paragraphs, I shall discuss these two points.

The first point, which I find quite compelling, is that "there can be no discrepancy… between deciding to make a certain decision and making that decision."[248] The point here is that it is impossible to decide something and, at the same time, to reject the decision or to decide not to make it. It is important to understand that this statement is based on Frankfurt's *ahistorical* view, so the point is just that we cannot make a specific decision and decide not to make it *at the same time*. Obviously, we can reject a decision that we made in the past. Similarly, we can reject a decision that we will in fact make in the future. But the real question here is: can we reject a decision that we are making at the moment we are making it? Of course, I may make a decision that I wish I did not have to make. I may, for example, decide in favour of a desire to give my money to a robber when he threatens my life, and I may wish I did not have to make that decision. But what I reject here are the conditions I am in, not my decision itself. Considering the conditions which I am in, I still endorse my decision: the decision is one that I decide to make. Thus, to the question "Can we reject a decision that we are making at the moment we are making it?", it seems that we have to answer "no," and thus that we have to conclude that no discrepancy is possible between making a specific decision and deciding to make that decision.[249]

---

[248] Frankfurt 1989, p. 78. From now on, I will focus on the concept of decision to simplify the discussion.

[249] Here I provide a defense of Frankfurt's argument according to which a decision cannot be passive. His argument, in other words, is the claim that an agent cannot be alienated from his decision. That view, however, has been criticized by Velleman and Mayr. Let me explain why I do not agree with Velleman and Mayr.

Velleman has developed a case that is supposed to show that an agent can be alienated from his decisions: the example of his "unwitting decision to break off a friendship" (Velleman 1992, p. 472). His example, however, is problematic for two reasons: it treats decisions as first-order phenomena and the alienation seems to occur at a later time (and not at the same time as the decision is made). To fix these two issues, Mayr has developed a version of Velleman's example, which goes as follow: "Imagine that John has, some time earlier, quarreled with James and has ever since had the conflicting desires to re-establish their relationship or to definitely sever the friendship. When, on an occasion, they meet in the street and begin to talk, John suddenly finds himself raising his voice and shouting abuse at James without provocation—a behaviour which is completely unintelligible to himself. Later reflection on the event shows John that he has already unconsciously decided in favour of the desire to sever the friendship.

Notice that this first point is compatible with a scenario where someone decides something and simply does not make a decision to decide. In that case, there is no discrepancy between a decision and a decision to decide, simply because there is no decision to decide. To exclude that kind of scenario, Frankfurt makes a second point. He argues that whenever someone decides something, she also necessarily decides to make the decision. This second point is expressed through his argument from logical equivalence. For Frankfurt, to decide to make a certain decision is "logically tantamount" to making the decision. As he says, "deciding not merely *to make some decision*, but *to decide to do that rather than to do this*, is logically tantamount *to deciding to do that*."[250] One way to understand that point is as follows.

According to the logical equivalence argument, whenever a decision to φ occurs, then a decision to decide to φ also occurs. And, conversely, whenever a decision to decide to φ occurs, then a decision to φ occurs. Thus, deciding to φ is logically equivalent to deciding to decide to φ. Schematically, we get:

- Decision to $\varphi \Rightarrow$ decision to decide to φ
- Decision to decide to $\varphi \Rightarrow$ decision to φ
- Thus, decision to $\varphi \Leftrightarrow$ decision to decide to φ

Though such decision was present, John did not identify with this decision, but felt 'violated' by it when it manifested itself in his insulting behaviour—in the same way as he feels 'violated' when overcome by a spasm of emotion" (Mayr 2011, pp. 60-61).

I believe that Mayr's argument is not a convincing counterexample. His point is that an agent may *feel* alienated from his decision. Velleman mentions on the other hand *a belief* that one's decision is alien. But why should we think that the *belief* or the *feeling* that one's decision is alien matters at all? In other words, what gives this belief or feeling the authority to speak for the agent? This belief or feeling of alienation might in fact be obsessional, or passive, and thus it is not clear that it has the authority to speak for the agent. To make a strong case against Frankfurt's approach, one would have to show that it is possible to reject a decision, but this rejection has to be made by means of a higher-order decision—not by means of a thought or feeling. Thus, it is not clear that Velleman and Mayr succeeded in making a case against Frankfurt.

Moreover, Mayr claims that the person "felt 'violated' by [his decision] when it manifested itself…—*in the same way as he feels 'violated' when overcome by a spasm of emotion*" (Mayr 2011, p. 61). One might wonder here if this is even possible. In other words, one might wonder if it is possible to feel violated by a decision in the same way that we feel violated by a spasm of emotion. My personal intuition is that this is not possible: decisions simply cannot be compared to spasmodic emotions in this way. Mayr has failed to give us any good reason to believe that the comparison is possible, and thus his counterexample makes, at best, a weak case against Frankfurt.

[250] Frankfurt 1989, p. 78.

If this argument is correct, then it means that whenever someone decides something, she also necessarily decides to make the decision. But this second decision, being itself a decision, also logically entails a decision to make the decision. And this goes on *ad infinitum*. This is one way to understand the "resonance effect" that Frankfurt talks about.[251]

From these two points, Frankfurt concludes that decisions are *necessarily active*: they are a reliable source of activity, just like Chisholm's concept of the agent. Notice that this account of decision is a reductive (although a non-causal) one. On Frankfurt's view, decisions confer activity to bodily motion (through an intermediate desire). But decisions are not intrinsically active or unanalyzable. Decisions are active, on his view, because other higher-order decisions confer them activity. Another way to put the point is to say that Frankfurt analyzes decisions in the same way that he analyzes bodily actions. On his view, the bodily action of a person is composed of a complex structure involving a bodily motion, a desire to perform it, and an endorsement (i.e., a decision to include the desire among the candidates for satisfaction). But decisions themselves are composed of a similar structure: a mental motion, a desire to perform it (i.e., a second-order desire) and an endorsement (i.e., another decision). Obviously, that option generates a regress since our analysis of decisions involves another decision, and our analysis of that other decision involves a third decision, and so on *ad infinitum*.

Now what should we think about Frankfurt's reductive solution and, more specifically, about the regress (or the resonance effect) that it generates? Is it really as unproblematic as Frankfurt seems to suggest? The first point to mention is that the regress is often seen as problematic, according to many action theorists, because it has some implausible temporal

---

[251] Frankfurt 1987, p. 168.

implications.[252] Imagine what would happen if Frankfurt's account of decision was an event-causal rather than a non-causal account. As we saw, the event-causal approach is essentially historical. On this view, a cause is distinct from its effect partly in virtue of its temporality: a cause precedes its effect. Now suppose that an event-causalist were to say that an act of will, being itself an action, needs to be caused by a second act of will and that this second act of will, being also an action, needs to be caused by a third act of will, and so on *ad infinitum*. Here, the regress means that all actions start at the beginning of time, which is obviously an implausible view.

However, Frankfurt's endorsement view, which is a non-causal or structural account of agency, avoids this problem quite easily. The structural view does not look at how an action is caused or at whatever mental antecedent it has. It looks at the way an action is structured: what matters is that the bodily or mental motion is endorsed by the agent at the moment it is occurring. We might say, then, that the regress occurs *vertically* instead of *horizontally*. Thus, the temporal issue is easily avoided.

While Frankfurt's account of decision can easily avoid the temporal issue, which is one of the reasons why the regress might be problematic, it is not clear that his account generates a completely unproblematic regress. There are other reasons in addition to temporal implausibility why a regress might be problematic. I will here suggest two other possible reasons, which I call the *capacity issue* and the *phenomenological issue*.

First, Frankfurt's account is subject to the capacity issue. Here, it is important to understand that Frankfurt's argument from logical equivalence, according to which the decision

---

[252] Brand has made this point. According to him, because "willings are themselves actions, there must be a willing to will, and a willing to will to will, and so on *ad infintum*. The regress is vicious, it is alleged, because each willing takes time, thus requiring an impossibly long time for the performance of a single action" (Brand 1984, p. 12).

to φ is logically equivalent to the decision to decide to φ, does not mean that the two decisions are *identical*. The argument just means that whenever one decision occurs, the other also occurs and vice versa. But these two decisions are two distinct entities or two distinct occurrences. Accordingly, the regress that Frankfurt's view generates entails an infinite number of mental occurrences. The problem then is that we do not have the mental capacity for infinite mental occurrences. As finite beings, we have limited mental capacities, and it is implausible that we could form, at the same time, a decision to φ, a decision to decide to φ, a decision to decide to decide to φ, and so on *ad infinitum*.

Finally, consider the phenomenological issue. On that view, the regress is problematic because people usually do not experience third, or fourth, or fifth-order mental attitudes. For example, people usually do not make a decision to decide to decide something. Frankfurt's account of agency is problematic because it supposes that every single action involves these third and fourth and fifth-order mental occurrences. Thus, Frankfurt's view is problematic because it entails the occurrence of mental events that are rarely, if ever, experienced. If they are never experienced, there is good reason to think that they do not occur at all.

Now that we have examined Frankfurt's solution to the first objection, we can move on to the second objection, the missing humanity or the missing moral being objection. What should we say about Frankfurt's solution to this objection? As we saw in Section 6.1, Frankfurt relies on self-reflexivity to solve the missing humanity objection. For him, an endorsement is a kind of self-reflective mental action, and as such it seems to be a typically human thing to do. But we also saw that this solution should be rejected because it has some problematic consequences. In the next paragraph, I shall simply summarize the argument presented in Section 6.1.

We saw that Frankfurt's reliance on self-reflexivity is problematic because it entails that actions lacking self-reflexivity count as non-human, even when they are performed by full-grown human beings. I call these actions "human wanton actions." The problem, more specifically, is that human wanton actions seem to be open to moral assessment in spite of what Frankfurt suggests, and thus they must possess something typically human. For that reason, I argued that we should abandon Frankfurt's solution to the missing humanity objection.

From the present discussion, we may identify a final *desideratum*. In the previous sections and chapters, we identified seven *desiderata* that our account of agency should satisfy. The eighth *desideratum* is that our account should not generate a problematic regress.

6.3.2. A dispositional and dialogical alternative

In this section, I develop an account of agency inspired by Frankfurt's endorsement view. To better solve the disappearing agent objection, I argue that we need to modify Frankfurt's account in two main ways. First, I argue that the best way to solve the first version of the objection, the missing source of activity objection, is to develop a *dispositional* account of agency. Second, I argue that the best way to solve the second version of the objection, the missing humanity objection, is to adopt a *dialogical* framework (which need not involve second-order desires and the like).

Let us start with the missing source of activity objection. My strategy to solve the objection is as follows. First, just like Frankfurt, I identify a mental act that is central to agency, a decision, and I propose a reductive account of decision. That will generate a regress, but I argue

that the regress is unproblematic if the central notion is that of a *disposition to endorse*, and not that of an actual endorsement.

Let us start with the first step: identifying a mental act that is central to agency. In Chapter 4, I defended the endorsement view of agency and argued that the crucial mental element is an endorsement. But I did not explain what kind of thing an endorsement is. Here, I shall adopt Frankfurt's strategy and claim that an endorsement is a certain kind of act of will or decision. Unlike Frankfurt, however, I will not draw a distinction between a decision and a choice, or between a first-order and a second-order act of will. If an endorsement can be construed as a decision, we need to ask: what kind of decision exactly? The decision in question is something like the decision to make a bodily or mental motion one's own. This is what is sometimes called the "ownership view." The idea behind it is that a bodily action is just a bodily motion that is attributable to the agent, and a bodily motion is attributable to the agent when the agent makes it her own or takes ownership for it.

Appealing to the concept of a decision allows us to avoid the missing source of activity objection, because it ensures that all actions necessarily involve an active element. But then we have to face the dilemma that we discussed in 6.2 and 6.3.1: either decisions are reducible and analyzable, in which case we end up with a regress, or they are unanalyzable and irreducible, in which case our account of agency is circular and disunified. I believe, as does Frankfurt, that the first option is the best and that we can generate an *unproblematic* regress. To show how the regress can be unproblematic, I propose in the following paragraph an analysis of decisions that builds on my account of habitual actions discussed in 4.5.2.

In Chapter 4 (Section 4.5.2), I argued that habitual actions count as actions because we are disposed to endorse the bodily motions that compose them. The idea was that what is

required is not an actual endorsement but a hypothetical endorsement. What is required is that *if the person were to become consciously aware of the motion in question, she would endorse it*. My suggestion here is that decisions count as actions for the same reason. In other words, a decision is an action because *if I were to become consciously aware of it, I would endorse it*. Notice that since an endorsement is itself a decision, this amounts to saying that *if I were to become consciously aware of my decision, I would decide to make it*. Thus, my account builds in part on Frankfurt's conception of decision and his view that "there can be no discrepancy between… deciding to make a certain decision and making that decision."[253]

Now it is obvious that this view generates a regress. The decision in question is active because I am disposed to endorse it. But this endorsement is also a decision, so it means that I have to be disposed to endorse it. And this goes on *ad infinitum*. But the regress, as I will show below, is unproblematic. We saw three reasons why a regress might be problematic: the temporal, the capacity and the phenomenological issues. In the next paragraph, I argue that my account of decision avoids all of these problems.

Let us start with the temporal issue. As we saw in the last section, on some views the regress means that all actions start at the beginning of time, which is implausible. But we also saw that the structural view avoids the problem quite easily, because it focuses on the structure of an action at a time instead of on causal antecedents. Thus, my structural account easily avoids the temporal issue.

The second issue that the regress might generate is the capacity issue. As we saw, Frankfurt's account requires the occurrence of an infinite number of decisions. This view is implausible because we do not have the mental capacity to perform all of these decisions *at the*

---

[253] Frankfurt 1989, p. 78.

same time. One might think that my dispositional account fares no better with respect to the capacity issue. While my account does not require the occurrence of an infinite number of decisions, one might think that it requires the disposition (and thus the capacity) to perform an infinite number of decisions at the same time, which we clearly do not have.

This possible objection, however, is mistaken. To understand why, we need to draw a distinction. My account does not require the disposition (and the capacity) to perform an infinite number of decisions *at the same time*, but it does require the disposition (and the capacity) to perform each and every one of these decisions. In other words, it requires the disposition (and the capacity) to perform each and every one of these decisions but it does not require that one be disposed (and that one has the capacity) to actualize all of these dispositions *at the same time*.

But is it plausible to suppose that a person is disposed to perform each and every one of these decisions? In other words, is it plausible to suppose that we have the capacity to perform third, fourth and fifth-order decisions, for example? I would argue that it is, because the only things that are required are the capacity to make decisions and some self-reflective capacity, of which we have both. If one can accept the idea that we can form second-order decisions, for example, then I see no reason to deny that we have the capacity for third-order decisions. All that is required is the capacity to reflect on the second-order decision and to form another higher-order decision. And we seem to have these capacities. Thus, my dispositional account can avoid the capacity issue.

The third issue is the phenomenological issue. As we saw, the regress might be problematic because it might be phenomenologically implausible: people rarely (if ever) experience third, or fourth, or fifth-order decisions. But on my account, what is required is not that we perform or experience these decisions. What is require is that we are *disposed to make*

them. And this makes all the difference. To say that we are disposed to make second and third and fourth-order decisions is phenomenologically speaking quite plausible: it just means that if we were to become consciously aware of our lower-order decisions, we would endorse them by means of a higher-order decision. Moreover, my view can quite easily explain why we do not actually make these second and third and fourth-order decisions: because the triggering condition rarely (if ever) occurs. The relevant triggering condition is the person's conscious awareness of the lower decision. The reason why we rarely (if ever) make second and third and fourth-order decision, even if we are disposed to make them, is because we are rarely (if ever) consciously aware of our lower-order decisions; we rarely take the time to reflect on them. For that reason, my account seems to be phenomenologically plausible.

We just examined my non-causal solution to the missing source of activity objection, and we saw that it satisfies the eighth *desideratum*: it does not generate a problematic regress. While it does generate a regress, the regress is unproblematic because it avoids the temporal, capacity, and phenomenological issues. Moreover, since my solution is a reductive one, it easily satisfies the fourth, the fifth, and the sixth *desiderata* that we discussed in Chapter 4. In the next paragraph, I shall explain how my account satisfies these three other *desiderata*.

The fourth *desideratum* is that an account of agency should not be *ad hoc*. My solution, which is based on the concept of a disposition to endorse, fulfills this fourth *desideratum*. My account is not *ad hoc* because we have some good positive reasons (other than to avoid the regress objection) to believe that agency requires a disposition to endorse rather than actual endorsement. In particular, the presence of this disposition explains a broad category of actions, namely, unreflective actions such as habitual and spontaneous actions.[254] The fifth *desideratum*

[254] On this, see Section 4.5.2.

is that our account of agency should not be circular. Although my solution appeals to the notion of decision, which is a mental act, on my view a decision is not intrinsically active, and consequently my account of agency is not circular. It is also more intelligible for the same reason: we know that decisions count as actions because we are disposed to endorse them. The sixth *desideratum* is that our account of agency should be unified. On my view, agency appears as a single phenomenon that can be uniformly analyzed. A decision counts as an action for the same reason that a bodily action counts as one: because the agent is disposed to endorse both of them. Thus, on my view, all actions have the same structure.

We just saw that my account of agency easily satisfies the four *desiderata* previously discussed. For that reason, it seems to propose the strongest solution to the missing source of activity objection. Before we conclude this section, we must examine the second objection, the missing humanity objection. As we saw, one issue with the standard account is that it fails to include an agent, which is a human or moral being. The question now is whether the endorsement view, as I have developed it, can solve that issue. In the next paragraphs, I shall argue that the endorsement view is best conceived as a *dialogical* view and that this allows us to solve the missing humanity objection. In arguing for a dialogical account of endorsement, I am following Westlund's suggestion,[255] although I am offering a dialogical account of *agency* rather than a dialogical account of *autonomy*. To develop this dialogical account of agency, I build on Hieronymi's account of answerability,[256] which is itself a reinterpretation of Anscombe's view.[257]

---

[255] See in particular Westlund 2009.
[256] Hieronymi develops her view in a series of articles. See Hieronymi 2008, 2009, 2014.
[257] Anscombe 1957.

On Hieronymi's view, our actions are open to moral assessment because we are answerable for them. To say that we are answerable for our actions, on her view, just means that the why-question (why did you φ?), which is a request for one's reasons, is "given application" or is rightly asked. But then, when is the why-question "given application"? On Hieronymi's view, the question "is given application just in case the assumptions naturally made in asking it are met."[258] So we shall ask: what are these assumptions?

One such assumption is that an intentional action is performed. As Hieronymi and Anscombe claim, the question "is given application by intentional actions."[259] Importantly, that includes intentional actions that are not performed for any particular reason. As Anscombe mentions, "[t]he question is not refused application because the answer to it says that there is no reason, any more than the question how much money I have in my pocket is refused application by the answer 'None.'"[260] Thus, the why-question is rightly asked even when an intentional action is done for no reason.[261]

The why-question is refused application, on the other hand, when a person does not act intentionally or when she is not aware of what she is doing. To make this clear, consider Davidson's example of the person who flips the switch, turns on the light, illuminates the room and, without being aware of it, alerts the prowler. Alerting the prowler is not something that the

---

[258] Hieronymi 2014, p. 14.
[259] Hieronymi 2014, p. 13.
[260] Anscombe 1957, section 25.
[261] The idea that we can act intentionally, or simply act, for no reason is a controversial one. Some causalists have argued that it is not possible (see, for example, Davidson 1971, Goldman 1970 and Mele 1988). Others have argued that the notion of acting for no reason is fully coherent. Anscombe is one example (1957). Frankfurt is another: "The supposition that people cannot make decisions or performs actions except for a reason strikes me as belonging to an excessively rationalistic conception of human life… I cannot see that reasons are required either by actions or by decisions" (Frankfurt 2002, p. 89).

person does intentionally. So, in this case it would not make much sense to ask her "why did you alert the prowler?" The why-question would not be rightly asked.

We just saw that the why-question is "given application" by intentional actions, but is it the only assumption that is made in asking the why-question? If that were the only assumption, then many non-human animals, assuming that they can act intentionally (which seems plausible to assume), would be rightly asked why they did what they did. It would make sense to ask a cat, for example, why it did what it did. But this view seems implausible: it seems implausible to assume that the why-question is rightly asked of a cat (for example), and that is so even if one could speak cat language. Thus, it seems that another assumption is made in asking the why-question: that a *typically human being* or an *answerable being* is acting intentionally. Obviously, we might want to leave open the possibility that animals other than human beings (e.g., great apes, dolphins) are answerable beings. That is after all an empirical question that arm-chair philosophers cannot settle.

One strategy to exclude non-answerable beings is to give the notions of intention and intentional action a thicker meaning than what is supposed in the standard story. Hieronymi opts for that solution. On her view, "to intend to φ is to settle positively the question of whether to φ."[262] So, that means that the why-question is only given application to beings who can settle or answer the question of whether to φ. Presumably, that excludes a broad category of non-human beings that are not able to answer the question. For that reason, Hieronymi's view is a plausible one. That being said, I shall argue that the endorsement view is a plausible solution as well.

---

[262] Hieronymi 2014, p. 14.

As we previously saw, an endorsement is a kind of decision, the decision to make a bodily or mental motion one's own. This is more or less the view that Frankfurt puts forward.[263] But, on Frankfurt's view, an endorsement entails something else: to endorse something is also to support or "get behind" it.[264] Thus, an endorsement might be best conceived as a complex decision involving two things: a decision to make a bodily or mental motion one's own and to stand behind it. Although Frankfurt does not say much about this notion of "getting behind" something, one natural way to understand the notion is in dialogical terms. What it could mean to "get behind" or "stand behind" a motion is that the person is ready to defend it in the face of possible challenges. In other words, a person who stands behind a motion is ready to justify it or to provide an explanation (if there is any): she is ready to answer the why-question. Thus, just like Hieronymi's solution, the endorsement view seems to exclude a broad category of non-answerable beings: all of those creatures that do not manifest this readiness to answer the why-question.

This dialogical account of endorsement can easily solve the missing humanity objection. As we saw, one issue with the standard account is that it fails to include a typically human being: it does not explain what it is for a human being to act. We also saw that this worry finds its origin in a moral concern, that is, in the need to distinguish typically human actions, which are open to moral assessment, from non-human actions, which are not open to such assessment. On the endorsement view that I have defended here, an action necessarily involves an endorsement or a disposition to endorse. Further, this endorsement entails answerability or the readiness to answer the why-question. This could be characterized as a typically human feature since most non-

[263] Frankfurt 2002, p. 87.
[264] Frankfurt 2002, p. 87,

human animals do not seem to be ready to answer the why-question. Moreover, it is also a moral feature, since it could explain why human actions are typically open to moral assessment. Thus, the endorsement view provides a solution to the missing humanity or moral being objection.

In Section 6.1, we identified a seventh *desideratum* that our solution to the missing humanity objection should fulfill: it should avoid treating "human wanton actions" as non-human actions. As we saw, this was a problem for Frankfurt's and Velleman's solutions, which locate the typically human feature of agency in self-reflexivity (a form of reflection on our mental attitudes). On the dialogical view that I have proposed here, human wanton actions count as typically human, even if they lack self-reflexivity, because they involve an endorsement and this endorsement entails a typically human feature: the readiness to answer the why-question. Thus, my solution satisfies the seventh *desideratum* and seems to be, for that reason, superior to Frankfurt's and Velleman's.

6.4. Conclusion

In this chapter, we examined different accounts of the agent which are meant to solve the disappearing agent objection. We saw that the event-causal and the agent-causal views are not satisfactory. We also saw that Frankfurt's non-causal solution, which is a volitional and hierarchical one, presents some difficulties as well. I argued that a *dispositional* and *dialogical* non-causal view can solve these objections and that it is the most satisfactory solution. For that reason, it seems to provide the strongest account of the agent.

Chapter 7: The Causalism and Anti-Causalism Debate

In Chapters 4, 5, and 6, I discussed the metaphysical question of agential control and argued that the endorsement view, as I developed it, provides the most compelling account of what it is for an agent to control her behaviour. In the present chapter, I show how this version of the endorsement view can help settle the epistemological debate in favour of the non-causal account of action explanation.

The causal theory of action is based on a certain account of action explanation. According to causalists like Davidson, to explain an action we need to identify the reason for which the action is performed, which is also the cause of the action (for example, a desire-belief pair). This view has been characterized as a form of "psychologism" and it has been widely criticized. The main issue is that reasons for actions do not appear to be mental attitudes *per se*, but their content.

To avoid that issue, many opponents have developed an alternative anti-causal view of action explanation. That view has been criticized by causalists in turn. One of the most prolific critiques of the anti-causal view, Mele, has rejected some of these anti-causal accounts of action explanation on metaphysical grounds. In this chapter, I will examine the debate between Mele and Sehon and argue that the endorsement view easily avoids Mele's critique. Thus, I will show that the endorsement view can help settle the epistemological debate in favour of the non-causal view.

This chapter is divided into five sections. In Section 7.1, I provide an exposition of the causalism and anti-causalism debate in the epistemology of agency. In Section 7.2, I discuss the debate between Sehon and Mele more specifically. In Section 7.3, I argue that the endorsement

view provides a strong solution to Mele's critique, and thus that it can strengthen the case for the non-causal approach. In Section 7.4, I examine a possible reply that Mele might raise against my view. In Section 7.5, I show that the reply is unsuccessful.

7.1. Causalism and anti-causalism in action explanation

In this first section, I provide an exposition of the causalism and anti-causalism debate in the philosophy of action. The debate, as I construe it here, concerns the nature of action explanation.[265] It is generally agreed that to explain an action we need to identify the reason for which the action is performed, a process known as "rationalization."[266] The debate concerns the nature of these reasons for action. Davidson famously argued that the reason for which a person acts is the cause of her action, namely, the desire-belief pair that causes the action.[267] Thus, Davidson's view is a *causal* account of action explanation.

However, his view, which is often characterized as a form of "psychologism," [268] has been the subject of many attacks. One problem with Davidson's "psychologism" is that reasons for action seem not to be psychological states themselves, but their content, which can be construed as facts, states of affairs, or true propositions. [269] But on this view, reasons do not seem to be causes, since facts, states of affairs, and true propositions are not typically construed as

[265] On that debate, see D'Oro and Sandis 2013 and Schumann 2019.
[266] Davidson 1963.
[267] Davidson 1963.
[268] See, for example, Alvarez 2017.
[269] See, for example, Dancy 1993, Raz 1999, and D'oro and Sandis 2013.

causes.[270] This rejection of Davidson's view can be called an *anti-causal* account of action explanation.

In response to Davidson's psychologism, many anti-causal views have emerged.[271] One of these, which is particularly influential, is the teleological anti-causal account of action explanation. Sehon, for example, has developed such a view.[272] On his view, reasons are states of affairs or goals towards which we aim. And these states of affairs, which are not yet realized, are not the causes of actions. Suppose that Claire went to the corner store. A possible reason why she went to the corner store is *to get milk*, a state of affairs towards which her action was directed. But, clearly, getting milk cannot cause the action of walking to the corner store, because it is not yet realized when the action is performed.

The anti-causal view is subject to one recurrent difficulty, which is Davidson's original worry about non-causal accounts of action explanation. The worry is that if rationalization is not a species of causal explanation, it might just be epiphenomenal: rationalization could render an action intelligible but it might not be a genuine explanation in the sense that the identified reason might not be the actual reason for which the person acted. This worry gave rise to what is known as "Davidson's challenge," a challenge that Davidson raised for non-causal accounts of action explanation.

Davidson's challenge is the following: how can we account for the distinction between "acting for a reason" and merely "having a reason to act" without an appeal to the notion of causation? Consider a case in which a person has two reasons to act. For example, Claire might

---

[270] Cf. Mellor 1995.
[271] See Wilson 1989; Tanney 1995, 2005, 2009; Hacker 1996, 2009; Rundle 1997; McCann 1998; Hutto 1999; Dancy 2000; Schroeder 2001; Schueler 2003, 2009, 2019; D'Oro 2007; Ruben 2009; Candish and Damnjanovic 2013.
[272] Sehon 1994, 2005, and 2016.

have two reasons to walk to the corner store: she might want to see her friend who happens to work there, and she might want to buy milk. Suppose as well that she actually walked to the corner store, and that she did it *for* just one of these reasons—to buy milk. How, then, can we account for the fact that she is only acting for one of these reasons and not for the other?

Davidson argues that the only plausible way to distinguish acting for a reason and having a reason for acting is to postulate that there is a causal link in the case of acting for a reason. In other words, on his view, we must postulate that when Claire acts for a reason, the reason *caused* her behaviour. In our example, the desire to buy milk *caused* her walking to the corner store. Her desire to see her friend, on the other hand, is merely a reason that she has to walk to the corner store because it remained causally ineffective: it did not cause her behaviour.

Some anti-causalist have attempted to answer the challenge, including Wilson, Sehon, Ginet and Wallace.[273] Mele, on the other hand, has been one of the most important detractors of these anti-causal attempts to solve Davidson's challenge.[274] One of the main critiques he has raised concerns the metaphysical basis of these accounts of action explanation. As Mele says, "because teleologists have not offered an acceptable account of what it is to *act*, or to 'direct' one's bodily motions, they have not offered an acceptable account of what it is to act for the sake of a particular goal."[275] Thus, it seems like anti-causal accounts of action explanation have a major weakness, which is that they have failed to provide a satisfactory account of what it is for an agent to control her behaviour. In the next section, we will examine in more detail this issue, focusing on the debate between Sehon and Mele.

---

[273] See Wilson 1989, Ginet 1990, Sehon 1994, 2005, 2016, and Wallace 1999.
[274] See Mele 2000, 2003, 2010, and 2019.
[275] Mele 2000, p. 287.

7.2. The Sehon-Mele debate

In this section, I examine a debate that has been going on for over twenty years (from 1994 to 2019) between Sehon and Mele. The debate originates in Sehon's attempt to answer Davidson's challenge, and thus it originally concerned the notion of "acting for a reason." But Mele critiques Sehon for providing an implausible metaphysics of agency and thus the debate has turned mostly to the metaphysical issue of what an action is.

In "Teleology and the Nature of Mental States" (1994), Sehon has argued that the teleological view can account for the distinction between "having a reason to act" and "acting for a reason" with no appeal to causation. Sehon argues that to account for the distinction, a teleologist needs to look at the set of counterfactual conditionals that the notion of "acting for a reason" supports but that the notion of "having a reason to act" does not. Sehon gives the example of Heidi, who has two different reasons to "lift a heavy book up to the top of a bookshelf": to put the book where it belongs and to strengthen her biceps.[276] Now he assumes that Heidi acts for only one of these reasons. To determine for which reason, we need to make a counterfactual test.

Suppose that Heidi lift a heavy book to the top of the bookshelf in order *to place it where it belongs*, while exercising her biceps is merely a reason that she has. If this is the case, some counterfactuals will be true. For example, if the book had belonged on the bottom shelf, she would not have lifted the book up to the top of the bookshelf. That counterfactual is true if she acted *in order to* place the book where it belongs. But suppose that this was just a reason that she had and that she acted for a different reason: to exercise her biceps. Then the counterfactual

[276] Sehon 1994, p. 67.

would be false: it is false to say that "if the book had belonged on the bottom shelf, she would not have lifted the book to the top of the bookshelf." Indeed, if she lifts the heavy book to the top of a bookshelf to exercise her biceps, then she would have lifted it up even if it belonged on the bottom shelf.

Thus, a teleological account can distinguish the concept of "acting for a reason" and "having a reason for acting." "Acting for a reason" comes with a set of true counterfactual conditionals that are not true in the case of "having a reason to act."

Mele has developed a first critique of Sehon's non-causal account of acting for a reason in "Goal-Directed Action" (2000/2003), in which he develops a case involving Norm and the Martians. The case was originally meant to repudiate Wilson's anti-causal teleological view,[277] but it soon became the center of the debate between Sehon and Mele. What the case is supposed to show is that Sehon's teleological view cannot account for what it is to *act for a reason*, because it does not adequately account for what it is *to act*. Thus, Mele critiques Sehon's account of action explanation, which is an epistemological position, on metaphysical grounds.

The case of Norm and the Martians involves Norm, a man that has many reasons to climb a ladder: to fetch his hat, his tool kit, and a basket of bricks. But Norm can also fetch just one of these things at a time. Moreover, Norm is controlled by Martians. As Mele puts it,

> Norm has learned that, on rare occasions, after he embarks on a relatively routinized activity (e.g., tying his shoes, climbing a ladder), Martians take control of his body and initiate and sustain the next several movements in the chain while making it seem to him that he is acting normally. He is not sure how they do this, but he has excellent reason to believe that they are even more skilled at this than he is at moving his own body, as, in fact, they are. (The Martians have given Norm numerous demonstrations involving other people.) The Martians have made a thorough study of Norm's patterns of peripheral bodily motion when he engages in various routine activities. Their aim was to make it seem to him that he is acting while preventing him from even trying to act by selectively shutting down portions of his brain. To move his body, they zap him in the belly with M-

[277] Wilson 1989.

> rays that control the relevant muscles and joints. When they intervene, they wait for Norm to begin a routine activity, read his mind to make sure that he plans to do what they think he is doing (e.g., tie his shoes, or climb to the top of a ladder), and then zap him for a while unless the mind-reading team sees him abandon or modify his plan. When the team notices something of this sort, the Martians stop interfering and control immediately reverts to Norm.
> A while ago, Norm started climbing a ladder to fetch his hat. When he reached the midway point, the Martians took over. Although they controlled Norm's next several movements while preventing him from trying to do anything, they would have relinquished control to him if his plan had changed (e.g., in light of a belief that the location of his hat had changed).

So, Norm has an intention or a goal in mind; he is "directing his behaviour" toward fetching his hat. But once the Martians notice it, they zap Norm in the belly, control his body, and produce the bodily motions required to accomplish his goal. And while they do all that, they make it seem to Norm as if he is acting normally.

In a case like this, Sehon would have to say that Norm *acted for a reason*, or that he climbed the ladder *in order to fetch his hat*. Indeed, Norm would pass Sehon's counterfactual test. Consider this counterfactual: if Norm's hat had been in the garden, he would have walked to the garden instead. This counterfactual is true in the present case. Indeed, we suppose that the Martians would have produced the appropriate bodily motion to allow Norm to fetch his hat (in this counterfactual case, walking to the garden). Now imagine that Norm had directed his behaviour at fetching his tool kit instead, and that fetching his hat was merely a reason that he had. In that case, the counterfactual would be false; indeed, it would be false to say that if his hat had been in the garden, he would have walked to the garden instead. The Martians, who are very good at reading minds, would know that Norm was planning to fetch his tool kit (and not his hat) and thus they would have produced the bodily motion that is conductive to it (climbing the ladder).

Thus, the case of Norm passes Sehon's counterfactual test. The issue, for Mele, is that it is implausible to say that Norm climbed the ladder in order to fetch his hat, because it is implausible to say that Norm *climbed the ladder* in the first place. That is, it is implausible to suppose that he is acting. The reason why it is implausible to suppose that he is acting is because he has no control over his body: the Martians are controlling his body. As Mele would say concerning Sehon's counterfactual test, "the result would be false, since [Norm] was not directing [his] behaviour—that is, acting—at all in this case. Rather, the Martians were controlling the motions of [his] body."[278]

Sehon has developed a first answer to Mele's critique in his book *Teleological Realism: Mind, Agency, and Explanation* (2005). In that book, Sehon still endorses the view that we can use a counterfactual test to distinguish "acting for a reason" and merely "having a reason for acting." But Sehon has also refined his teleological account of action explanation so it is useful to start with an examination of his account.

Specifically, Sehon develops an "account of the epistemology of teleology."[279] He argues that in order to explain an action we need to identify a goal that respects two criteria. First, the goal has to be such that the action appears to be the most appropriate means to achieve that goal. Call this "the means-end optimization principle." Second, the goal has to be the most valuable state of affairs toward which the action could be directed. Call this "the optimal value principle." We need to add that these two criteria should be based on a "viable theory of the agent's intentional states and circumstances," and not on an abstract and ideal agent.

---

[278] Mele 2000, p. 287.
[279] Sehon 2005, p. 169.

To illustrate this view, take Sehon's example of Sally, who withdraws life support from her terminally hill and comatose father. To explain her behaviour, we need to identify a goal that satisfies both the means-end optimization and the optimal value criteria, while taking into consideration Sally's intentional states and circumstances. A possible explanation of her behaviour is that she withdrew life support *to allow her father to die with dignity*. First, the action (withdrawing life support) would appear to be the most optimal way to achieve the end (allowing her father to die with dignity) given Sally's beliefs about human dignity. Second, the end (allowing her father to die with dignity) would also appear to be the most valuable state of affairs towards which the action could be directed. Sally might value another state of affairs that the action will realize. Since Sally lives in the United States, pulling the plug on her father will relieve her of the enormous hospital bills, which is something that she also values. But given Sally's intentional states, her love for her father, and her extravagant personality, allowing her father to die with dignity appears to be a more valuable goal.

To answer Davidson's challenge, Sehon still relies on a counterfactual test. Let us suppose that Sally acted for one reason only: to allow her father to die with dignity. Suppose as well that Sally had more than one reason to act. As we saw, allowing her father to die with dignity and saving some money were two reasons that she had. To show that she withdrew life support to allow her father to die with dignity, and only for that reason, we may run a counterfactual test based either on the means-end optimization criterion or on the optimal value criterion. According to the means-end optimization principle, "agents act in ways that are appropriate for achieving their goals given the agents' circumstances, epistemic situations, and

intentional states."[280] Thus, if we alter an agent' circumstances, but the appropriate way to achieve a goal has not changed, we would expect the same behaviour.

Consider this counterfactual scenario. Imagine that "the hospital charged a large fee for withdrawing life support such that this course of action was actually more expensive for Sally than allowing her father to stay on the machines."[281] In this counterfactual scenario, withdrawing life support is still the best way to achieve her goal, which is to allow her father to die with dignity. Thus, we can assume that she would have done the same thing in that counterfactual scenario: it is true to say that "if the hospital had charged a large fee, she would have withdrawn life support." But if "allowing her father to die with dignity" was not the reason for which she acted and was merely a reason that she had, then the counterfactual would be false. Indeed, suppose that she withdrew life support to save money; then she would *not* have withdrawn life support because withdrawing life support is not appropriate to achieving her end of saving money. Thus, Sehon's solution to Davidson's challenge is still the same. On his view, "acting for a reason" supports a set of true counterfactual conditionals that "having a reason to act" does not support.

After presenting his solution to Davidson's challenge, Sehon discusses Mele's critique. To repudiate Mele's objection, Sehon examines the case of Norm and the Martians and questions Mele's intuitions on that case. More specifically, he questions Mele's intuition according to which Norm is not acting. To determine whether Norm is acting or not, Sehon argues that we need to know more about the case, so he imagines two possible scenarios. In the first scenario, the Martians are completely reliable, just like Mele seems to suggest. In the second scenario,

---

[280] Sehon 2005, p. 157.
[281] Sehon 2005, p. 157.

they are not. Sehon argues that in the first case, it is mistaken to suppose that Norm is not acting, while in the second case it is correct.

In the first scenario, the Martians are making Norm's body move exactly as Norm had planned to make it move, and this is an "ironclad promise." In a case like this, we are very close to a "variety of occasionalism,"[282] in which Martians, rather than God, are producing all of our bodily movement in a perfectly reliable way. In this case, Sehon argues that Norm is acting in spite of the unusual causal chain. To make this point more acute he suggests a case in which Norm's behaviour triggers some moral intuition.[283] Imagine that Norm is shooting his philosophy professor instead of fetching his hat and that the Martians are fully reliable.

> At the moment when Norm is picking up the gun and about to shoot it, the Martians take over his body and make it carry out the dirty deed. However, they make it seem to Norm as if he is acting, and if Norm had changed his mind and decided to put the gun down, the Martians would have immediately relinquished control and Norm would not have committed the murder.[284]

In a case like that, it seems like Norm is responsible for *shooting* his professor and not just for having a plan or an intention to shoot his professor. Thus, it appears that Norm is acting in this first scenario.[285]

In the second scenario the Martians are not reliable. While this is not suggested by Mele, it seems to follow from our intuitions concerning extra-terrestrial life and the imperfect nature of these Martian creatures (Martians are not God). As Sehon claims, "[t]he Martians have chosen not to disrupt Norm's plans on this occasion, but there are no guarantees that they will always

---

[282] Sehon and Mele engage in a discussion on occasionalism (Sehon 2005/2016 and Mele 2010/2019) and what would happen to our conception of agency if that view were true. I omit it here since I do not think that it is very useful for the debate.
[283] Sehon argues that "some intuitions are more worthy of reliance than others. In particular, when it comes to what counts as an action, we should take most seriously our common sense reactions as embodied in ordinary practices of praising, blaming, and, in general, holding one another responsible" (Sehon 2016, p. 58).
[284] Sehon 2005, p. 168.
[285] Schueler has made a similar argument. See Schueler 2019, p. 64-65.

use their powerful technology in such a benign manner."[286] In this second scenario, the Martians could fulfill Norm's aim or not: this is at their discretion. Sehon argues that Norm is not acting in this second scenario because his "behaviour will be far from appropriate for achieving his goals."[287] In other words, his behaviour will fail to satisfy the means-end optimization principle that we previously discussed.

Remember that, for Sehon, a behaviour counts as teleological (that is, as an action) when it is the best way to achieve a certain aim, and this is something that can be verified with a counterfactual test. Mele claims that Norm climbed the ladder in order to fetch his hat. But when the Martians are unreliable, Norm's behaviour fulfills his goal simply because the Martians have decided to collaborate. In many other counterfactual scenarios in which the Martians do not collaborate, Norm's behaviour would not be an adequate way to achieve his aim. Thus, it seems like Norm's behaviour fails to satisfy the means-end optimization principle. For that reason, it does not count as an action.

In the end, the problem with Mele's case is that it suggests something implausible: imperfect Martians who are perfectly reliable. When we separate these elements and either talk about imperfect Martians or God-like Martians, then, for Sehon, a teleological anti-causalist account seems to provide the right verdict.

Following Sehon's response, Mele has come back to the charge in his "Teleological Explanations of Actions: Anticausalism versus Causalism" (2010). In that paper, he argues that Sehon's conclusions about the two scenarios are mistaken.

---

[286] Sehon 2005, p. 169.
[287] Sehon 2005, p. 169.

The first scenario involves reliable Martians and, instead of fetching his hat, Norm shoots his philosophy professor. Sehon argues that Norm is acting in that scenario because we tend to see him as responsible for shooting his philosophy professor. We would not simply condemn Norm for planning or intending to kill the professor. As a response to Sehon, Mele identifies a distinction that he seems to ignore, the distinction between being responsible *for shooting* the professor and being responsible *for the shooting* of the professor. Think about a person, say John, who hires a hitman. John may be responsible *for the shooting* of his victim without being responsible *for shooting* his victim. Indeed, it would be strange to say that John is responsible for shooting his victim since he did not shoot his victim. Mele argues that Sehon's case is similar: while Norm seems to be responsible for the shooting of the professor, he is not responsible for shooting him since he did not perform an action. Thus, on Mele's view, our moral intuitions in this case do not show that Norm performed an action.

In the second scenario, the Martians are unreliable. Sehon claims that in such a case, it is correct to say that Norm is not acting. But he also explains that this is not a problem for his account: according to his epistemology of teleology, Norm's behaviour does not count as goal-directed because the means-end optimization principle is not satisfied. Mele rejects Sehon's argument and claims that his "explanation of why Norm is not acting is seriously problematic." To explain why Sehon's argument is problematic, Mele supposes the opposite scenario, a scenario in which the Martians do not intervene even though they could have intervened. In this scenario, Norm acts as he usually does: he climbs the ladder or walks to the kitchen. But in this scenario, there are also "indefinitely many" counterfactual scenarios in which the Martians intervene and in which Norm's behaviour is not appropriate to achieve his goal, so Norm's

behaviour would not respect the means-end optimization principle and thus would not count as an action on Sehon's view. This seems implausible for Mele.

Mele's second critique is not the end of the story. Sehon has further defended his intuitions concerning the case of the reliable Martians in his *Free Will and Action Explanation: A Non-Causal, Compatibilist Account* (2016). It should be noted that he has not defended his intuition concerning the case of the *un*reliable Martians. The reason for that is probably because this task appears to be superfluous. Indeed, both he and Mele reach the same conclusion concerning the case of Norm and the unreliable Martians: in that scenario, Norm is not acting. Consequently, Sehon has focused on defending his interpretation of the scenario involving the reliable Martians.

As we saw, Sehon wants to show that Norm is acting when controlled by the reliable Martians. In order to do that, he develops a version of the case that appeals to our moral intuitions. In that scenario, Norm plans to kill his philosophy professor instead of fetching his hat. Based on that scenario, Sehon has argued that we would take Norm to be responsible for shooting the professor. In response to that argument, Mele has claimed that Sehon failed to distinguish between "being responsible *for shooting* the professor" and "being responsible *for the shooting* of the professor." On Mele's view, we take Norm to be responsible for the shooting, and not for shooting. Thus, the case does not show that Norm is acting.

Against Mele's critique, Sehon argues that it is not clear that we would take Norm to be responsible *for the shooting*. The analogy with the person who hires a hit man is in fact confusing. In that case, the person is performing an action that is conducive to the death of her victim, namely, she hires a hit man. So, it appears that the person is responsible for the shooting because she did something that resulted in the shooting. But, on Mele's interpretation of Norm,

Norm did not perform any action. Thus, it seems like Mele could not claim that Norm is responsible for the shooting.

Sehon also considers the possibility that Norm is responsible for the shooting because he failed to act, which would be a case of omission. But Sehon rejects that possibility because "there is… no failure or omission for which we would hold Norm responsible."[288] However, Sehon's response might be a bit too quick. Mele could claim that Norm is responsible for the shooting because he failed to intervene when he should have done so. But notice that this view would entail that Norm is responsible for *allowing the shooting to happen*, and *not for shooting*. And, although this is a very controversial issue,[289] we typically treat cases of allowings not as severely as cases of doings. The issue is that it is not clear that we would want to diminish Norm's responsibility in this case. If Norm were to say, "I did not shoot the professor, I just let it happen," we would probably treat that claim as a disingenuous way of avoiding blame and we would probably reject it. Thus, *contra* Mele, it seems like Norm is not simply responsible for the shooting: he is responsible *for shooting*.

Mele has developed a recent response to Sehon's last position. In his "Causalism: On Action Explanation and Causal Deviance" (2019), Mele further defends his intuitions concerning Norm and the reliable Martians. In that text, however, Mele does not introduce any significantly new information, except maybe concerning occasionalism,[290] and it appears as if the debate has reached a dead end.

---

[288] Sehon 2016, p. 59.
[289] On this debate, see Woollard and Howard-Snyder 2016. See also Rachel 1975 for a critique of the distinction between doing and allowing in the context of euthanasia.
[290] As mentioned previously, I omit the discussion about occasionalism.

The first thing to say is that Mele reiterates his distinction between "being responsible for shooting" and "being responsible for the shooting." In his discussion of the distinction, Mele does not respond to Sehon's point according to which "being responsible for the shooting" involves some kind of action (the person hiring a hitman performed an action and is, for that reason, responsible for the shooting). Instead of responding to that point, Mele strengthens his case about Norm and the shooting and presents a few similar cases. One of these involves a Martian with tentacles, a case that he had previously discussed in Mele 2010. In that scenario, a Martian has long invisible tentacles:

> a Martian makes himself invisible and then, with his slim but powerful tentacle, pulls Norm's paralyzed finger down on the trigger. (The Martian also makes it seem to Norm as though Norm is pulling the trigger, and "if Norm had changed his mind and decided to put the gun down, the [Martian] would have immediately relinquished control" to Norm.) Obviously, Norm did not pull the trigger. So Norm did not shoot the professor.[291]

The main difference between this case and the previous case (the Martians zapping Norm) is that the "Martian is pushing and pulling Norm from the outside" rather than "pushing and pulling Norm from the inside."[292] That being said, it is not clear what the distinction is supposed to add to the debate between Mele and Sehon.

7.3. The solution from endorsement

In the previous section, we examined the debate between Mele and Sehon and saw that, although the debate originally concerned the notion of acting for a reason, it became a debate over the nature of agency. In the present section, I will argue that the endorsement view, which is

---

[291] Mele 2010, p. 188.
[292] Mele 2019, p. 51.

based on the notion of standby control, provides a more satisfactory response to Mele's objection than Sehon's view. In doing this, I also show that Mele's objection is not a serious issue for non-causalists.

As we saw, the Sehon-Mele debate has been formulated in terms of reliability and unreliability. My suggestion here is that the real issue lies in a related distinction, that between *direct* and *indirect* control. I say that it is a related distinction because indirect control is often unreliable while direct control is typically more reliable. But that need not be so. We could easily imagine a scenario in which indirect control is very reliable and direct control is unreliable. It is in these scenarios that the difference between my solution and Sehon's solution will be the most obvious. Before we examine in more detail the distinction between direct and indirect control, I would like to go back to Mele's argument as it was formulated in "Goal-directed Action" (2000).

In "Goal-directed Action," Mele claims that Norm's behaviour is not an action because Norm has no control over his bodily motion. For example, concerning the similar case of Heidi, Mele says: "Heidi was not directing her behaviour—that is, acting—at all in this case. Rather, the Martians were controlling the motions of her body."[293] What Mele's view suggests is that because the Martians took control of Norm's or Heidi's body, Norm and Heidi lost control over their body until the Martians "relinquish control." Thus, Mele suggest that control is something that passes from hand to hand and not something that two agents can possess at the same.

That assumption is seriously problematic, for we often suppose that two persons can have control over the same thing. Two spouses might share a bank account, in which case they both have control over the same bank account. Similarly, imagine that Norm were controlled by a

[293] Mele 2000, p. 287.

puppeteer with a complex set of ropes.[294] And imagine that the ropes were so fragile that Norm could easily break them. Then surely the puppeteer controls Norm's body. But it would be implausible to claim that, because of this, Norm does not also have control over his body. Norm preserved control all along. The case of Norm and the puppeteer is a case where two agents have control over the same body.[295] The case of Norm and the Martians seems to be similar to this one. Mele thinks that Norm has no control over his body, but it appears to me that there are many ways in which Norm could be said still to have control over his body. As Mele concedes, "Norm is prepared to adjust or modulate his behaviour, and we may even suppose that he is *able* to do so."[296] He is able to do so, because, according to Mele, if he were to change his plan the Martians would relinquish control to him. But this readiness and ability to intervene is what I have called, in Chapter 4, "standby control," and surely Norm has standby control over his body. Moreover, we may also say that Norm has some kind of initial control over his body, since the Martians just accomplish whatever Norm is intending to do. Norm's bodily motion depends on what he is intending to do, and thus Norm certainly has a form of initial control.

Therefore, Mele's conclusion that Norm's behaviour is not an action because Norm does not have control over his body seems to be unjustified. What Mele could have said instead is that Norm does not have the *right kind of control* over his body.[297] But notice that this argument is rather different. That being said, I do agree with Mele's conclusion and believe that Norm is not acting, but the reason is not because he lacks control over his body. Rather, the reason is because

---

[294] A suggestion made by Mele as well. See Mele 2019 p. 51.
[295] One might point out that the case of Norm and the puppeteer is not identical to the case of the two spouses sharing a bank account, and I would agree with that claim. The case of the two spouses sharing a bank account is a case in which two persons both have full control over something, while the case of Norm and the puppeteer is a case in which a person seems to have partial control (the puppeteer) and another to have full control (Norm). For the sake of simplicity, I have left aside the difference between these two cases.
[296] Mele 2000, p. 285.
[297] For a similar point, see Sehon 2005, p. 168.

he does not have the right kind of control. In particular, the issue seems to be that, while he has *indirect* standby control, he does not have *direct* standby control. As we saw in Section 4.5.1, *direct* standby control is the notion of control on which the endorsement view is based.

The distinction between direct and indirect control is crucial to the philosophy of agency.[298] As we saw in Section 4.5.1, a typical way to draw the distinction is the following: a person has indirect control over something when *an action* needs to be performed to control that thing, whereas a person has direct control when *no action* is required. To illustrate the case, consider the case of a person's heartbeat. To control her heartbeat, a person may start to run: that will accelerate her pulse. The form of control involved in this case is what we might call "indirect control," because it requires the performance of an action (running). On the other hand, we do not have direct control over our heartbeat: we cannot control it unless we perform some kind of action. Compare this case with the movement of a person's left arm. Surely, the person can indirectly control the movement of her left arm: she may, for example, grab her left arm with her right arm and make it move. This is a case of indirect control because it requires an action (the arm grabbing). But the person can also directly move her left arm if she simply raises it. This is a case of *direct* control.

What is required for agency is *direct* control and not indirect control. There is one obvious conceptual reason for this: if our account of agency were based on a notion of control that requires the performance of an action, we would have a circular account of agency. There is another reason why agency seems to require direct control: because it corresponds to our intuitions. Consider the case of a person's heartbeat. No one would want to say that the

[298] See for example, Wilson and Shpall 2016 and Mele 1992.

movement of her heart is an action that the person is performing. The most plausible explanation for this is that, while she has indirect control over her heartbeat, she cannot directly control it.

With that in place, we can easily explain why Norm is not acting when he is controlled by the Martians: because he does not have direct control over his bodily motion. The control he has is merely indirect and requires that the Martians preform some kind of action. For example, Norm has standby control because he is ready to intervene and is able to do so. But Norm's standby control is indirect: he is only able to intervene because the Martians would stop zapping him in the belly if he were to change his plan. Thus, his control over his bodily motion depends on the performance of an action by the Martians. It is an indirect form of control. For that reason, Norm's behaviour does not count as an action on the endorsement view. Thus, the case of Norm and the Martians does not appear to be a problem.

We saw that indirect control requires the performance of an action and direct control does not. This is why indirect control is typically less reliable than direct control, especially when it requires the performance of an action by another agent: we can never be sure that the other agent will not change her mind or fail to perform the required action. But indirect control could sometimes be very reliable: I can very reliably control my heartbeat by running. Thus, indirect control is not simply a form of unreliable control. Similarly, direct control is not simply a form of reliable control. In the rest of this section, I would like to examine the two scenarios mentioned by Sehon in the light of the distinction between direct and indirect control. I shall argue that interpreting these cases in terms of direct and indirect control allows us to better explain our intuitions than the distinction between reliability and unreliability.

The first scenario involves reliable Martians. In that scenario, I tend to side with Sehon's intuition and believe that Norm is acting. But this is not so much because the Martians are

reliable; this is rather because they appear to behave like robots. And the kind of control these robotic Martians have over Norm's body mimics the kind of direct control we have over our own bodies. As we saw, direct control does not require the performance of an action. Similarly, if Norm has control over his body by means of these robotic Martians, and we assume that robots do not act, then Norm has a kind of direct control over his body. This is why we tend to see him as acting.

This point is also suggested by an analogy that Sehon draws between the case of Norm and the Martians and the case of Jane and the device. In his 2016 book, Sehon develops another scenario to strengthen his case against Mele. The case goes as follow:

> Jane's right index finger has, through some sort of unfortunate mishap, become completely paralyzed. However, doctors discover that they can make the finger move in all of the ways it used to be able to by installing a microscopic device within the finger, and then by sending appropriate signals to the device via Wi-Fi. Moreover, the doctors have a way of reading Jane's brain with a separate device implanted in her skull, such that if Jane simply moves about her life in the usual way, the device detects when she was about to move her index finger, and it sends corresponding signals, also via Wi-Fi, to a centralized computer server which then broadcasts a signal back to the device in her finger, making her finger move in just the way she had planned.[299]

This scenario is supposed to show that Jane is acting even though her intentions did not cause her bodily motion. Rather, the device and the Wi-Fi did it: "Jane controls her finger's behaviour *via* the Wi-Fi."[300] I agree that Jane is acting. But according to Sehon, this is because the device is a reliable tool. Where this gets implausible is when we imagine that Jane uses an unreliable device. Suppose that the device, which is after all a new invention, only work 25% of the time. Then, Sehon would have to say that Jane is not acting even when the device does what Jane intends to do. This seems implausible.

---

[299] Sehon 2016, p. 60.
[300] Sehon 2016, p. 61.

The reason why Jane is acting when she uses the device is because she *directly* controls her bodily motions by means of the device: no action is required to control these motions. And that is true in the case of the unreliable device as well. When it functions properly, Jane is acting because she directly controls her bodily motion.

This brings us to the second scenario and Mele's critique of it. In this scenario the Martians are unreliable. According to Sehon, Norm is not acting in that scenario because his behaviour fails to satisfy the means-end optimization criteria. His behaviour is not a suitable way to achieve the intended result or, in other words, "in a range of nearby counterfactual situations his behaviour is not appropriate to his goals."[301] Against that point, Mele has developed a case where the Martians do not intervene although they could have: "Imagine a case in which the Martians consider interfering with Norm but decide against doing that. Norm walks to the kitchen for a beer without any interference from the Martians."[302] In that case, Sehon would have to say that Norm is not acting because "in a range of nearby counterfactual situations his behaviour is not appropriate to his goals." But this seems implausible. When Norm walks to the kitchen, he *is* acting.

As far as I know, Sehon has not developed a response to this case. The reason is probably just because this second scenario is not a critique of anti-causal views. But Mele's point here is instructive: it suggests that the distinction between reliability and unreliability is not the relevant distinction, because it would lead Sehon to endorse counterintuitive results. Rather, the important distinction seems to be the distinction between direct and indirect control. For example, we could say that when the Martians do not intervene and do not zap Norm, Norm is acting because he

---

[301] Sehon 2005, p. 170.
[302] Mele 2010, p. 190.

preserves *direct* standby control over his body. Thus, the distinction between direct and indirect control does not lead to the same counterintuitive results that Sehon's account leads to.

What this discussion suggests is that an anti-causalist about action explanation, like Sehon, could use the endorsement view to defend his account on metaphysical grounds. Since the endorsement view is based on an account of agential control, it does not run into the same difficulty as the anti-causal teleologist view.

7.4. Mele's objection to Frankfurt

We just examined Mele's critique of Sehon's anti-causal teleological view and we saw that the main issue concerns his anti-causalist metaphysics of agency (or his lack thereof). We also saw that the endorsement view is not subject to the same difficulty. For this reason, I concluded that an anti-causal teleologist like Sehon could use the endorsement view to defend his account of action explanation. In the present section, I will examine a possible objection that Mele might raise against my solution, one that is based on his critique of Frankfurt. In the next section, I will argue that Mele's potential objection is unsuccessful.

Mele has developed a critique of Frankfurt's non-causal account of action in a paper provocatively called "Passive Action" (1997). In his paper, Mele examines two cases based on Frankfurt's non-causal account and he argues that these two cases do not undermine the causal theory of action because they involve causation. If Mele's argument were successful, it would show that the endorsement view is in fact a causal account in disguise, and thus that my attempt at defending the non-causal view has failed.

In his critique of Frankfurt, Mele examines two cases that are supposed to undermine the causal theory of action. The first case is the case of Al, a driver coasting downhill. This case is a replica of Frankfurt's coasting scenario.[303] In that scenario, the car is coasting downhill in virtue of gravitational force alone and Al is satisfied with the movement of the car. Al is also prepared to intervene if there is interference, and he has the ability to do so more or less effectively. The second case is the case of Peter, who "awakes to find himself rolling down a snow-covered hill."[304] This case is based on the coasting scenario. The reasons why Mele develops this case is because it more obviously involves a bodily motion. But I take the two cases not to be significantly different, because the coasting also involves a bodily motion: in that example, Al's body is moving and the car can be seen as an extension of Al's body.[305] Moreover, Mele's critique of the two cases is the same so the distinction does not really matter for our purposes.

What the coasting and rolling scenarios are supposed to show, for Frankfurt, is that causation by mental elements is facultative for agency. Mele argues that this is mistaken, and he develops two arguments against that view. The first argument consists in showing that there is causation by mental items in these two scenarios. The second argument consists in showing that Frankfurt's counterfactual strategy, which consists in looking at what would happen in cases of interference, fails.

Let me start with the first argument, or the claim that there is, in fact, causation by mental states and events in the two scenarios. It is useful to distinguish three steps in Mele's argument. First, Mele argues that Al must have a desire or an intention to coast or that he must have

---

[303] See Frankfurt 1978.
[304] Mele 1997, p. 139.
[305] Zhu has argued, on the other hand, that "Frankfurt's example of the coasting car is used as an analogy: the coasting vehicle is the source analogue of one's body in motion" (Zhu 2004, p. 303).

decided to coast. Mele's point is based on Frankfurt's description of the case, in which the driver is *satisfied* with the speed and direction of the car: "A driver whose automobile is coasting downhill in virtue of gravitational forces alone may be entirely satisfied with its speed and direction."[306] As Mele claims, "In the absence of a desire or intention regarding 'the movement of the automobile,' there would be no basis for the driver's being 'satisfied' with the speed and direction of his car."[307]

Second, Mele argues that "it is natural to say that Al is coasting in his car… *because* he wants to, or intends to, or has decided to" and that "the 'because' here is quite naturally given a causal interpretation."[308] To develop his point, Mele uses the standard conditional analysis of causation. In general, when we say that A causes C, we mean that "if A had not occurred, C would not have occurred."[309] Schematically we get:

- If ¬A, then ¬C

On Mele's view, if the desire or intention or decision of the driver had not occurred, then the coasting would not have occurred: "if Al had not desired, or intended, or decided to coast, he would not have coasted."[310] What he suggests here is that if the driver had had another desire or intention, then he would have done something else instead and thus the coasting would not have occurred. Thus, according to Mele, we can conclude that the desire or intention or decision *causes* the coasting.

Third, Mele examines an exception that seems to falsify his analysis. In that scenario, if Al had not decided to coast, he would have been indifferent and thus the coasting would have

---

[306] Frankfurt 1978, p. 75.
[307] Mele 1997, p. 137.
[308] Mele 1997, p. 138.
[309] Menzies and Beebee 2020.
[310] Mele 1997, p. 138.

happened: “If Al had not decided to coast, he would have been utterly indifferent about the motion of his car - in which case he would have done nothing to alter the car’s course and the car would have continued coasting. In this scenario, it is false that if Al had not decided to coast, the car would not have continued coasting.”[311] This raises a problem for Mele’s analysis because it suggests that the decision did not cause the coasting.

To respond to that case, Mele suggests that this scenario is an unproblematic exception similar to a case of “preemption.” As it is well known in the literature on causation, the conditional analysis sometimes fails, in particular in those cases of preemption.[312] In a case of preemption there are two potential causes, A and B. It is assumed that A causes C in the actual scenario. But it is also supposed that if A had not occurred, then B would have caused C. Schematically, we get:

- If ¬A, then B causes C

Clearly, a case like this falsifies the conditional analysis because C occurs even when A does not. To illustrate this idea, consider Mele’s example: “X dialed Y’s phone number at *t,* but if X had not done so, Z would have done so (at *t*). X’s dialing is a cause of Y’s phone’s ringing at t_1, even though the phone would have rung at t_1 if *X’s* dialing had not occurred.”[313]

What those cases of preemption show is that the conditional analysis fails in some cases, not that there is no causation. Mele suggests that the case of would-be-indifferent-Al is a case of this sort. The idea is that if Al had not decided to coast, then gravity would have caused the coasting anyway. Thus, we get: “if ¬A, then B causes C”. But, for Mele, this simply shows that

[311] Mele 1997, p. 143.
[312] See for example Lewis 1973.
[313] Mele 1997, p. 143.

the conditional analysis fails in some cases, not that there is no causation. Thus, the case of would-be-indifferent-Al is not an issue for his analysis.

Mele's second argument against Frankfurt's view consists in showing that his counterfactual strategy fails. To show that, he develops a scenario involving a mind-reading demon. He says:

> Imagine that, throughout the episode, Al was satisfied with how things went and did not intervene. He decided to coast and the coasting was purposive. Imagine further that although Al intended to intervene if necessary, an irresistible mind-reading demon would not have allowed him to intervene. If Al had abandoned his intention to coast or had decided to intervene, the demon would have paralyzed Al until his car ran its course. The coasting is purposive [i.e., an action] even though Al was *not* 'in a position to [intervene] more or less effectively.'"[314]

In this case, Frankfurt would have to say that Al is not acting or that his behaviour is not purposeful, because he is not in a position to intervene. But that seems counterintuitive for Mele. The behaviour is purposive; thus, it is an action.

What Mele wants to show with this case is that the counterfactual scenario does not really matter for determining whether an action occurs or whether a behaviour is purposive. What matters is what happens in the actual scenario. Thus, he compares that case with another one: the case of a person who is actually paralyzed while coasting. In that scenario, it is clear that the person is not acting or that her behaviour is not purposeful because she has no control over her body. But this is not the case in the mind-reading demon scenario.

7.5. A response to Mele

[314] Mele 1997, p. 138-139.

In this section, I show that Mele's arguments against Frankfurt fail, and thus that he does not succeed in undermining the non-causal view. The first argument fails for three reasons, which correspond to the three steps in his argument. The second argument, on the other hand, fails because Mele's case of the mind-reading demon involves *actual* paralysis.

Mele's first argument consists in showing that, in the coasting and rolling scenarios, a mental item causes the coasting or the rolling. This argument involves three sub-arguments. The first is that Al and Peter must have a desire or an intention to coast or to roll. Mele deduces this from the fact that Al and Peter are satisfied with the movement of the car and the movement of his body respectively.

Mele's first point, however, is mistaken. It is true that Frankfurt mention the driver's satisfaction. And it is also true that, in general, we assume that satisfaction comes with a desire or an intention: we are satisfied when we relieve a desire or the like. That being said, Frankfurt's notion of satisfaction is technical and departs from our colloquial notion. For him, to be satisfied with something is to have no desire or intention to change things as they are: "Being genuinely satisfied is… a matter of simply *having no interest* in making changes. What it requires is that psychic elements of certain kinds *do not occur*."[315] Thus, Mele's assumption seems to be unjustified.

Moreover, the important notion, on the endorsement view, is that of an *endorsement*. What an action requires is that a bodily or mental motion be endorsed. And, as we saw in Section 4.5.2, an endorsement cannot occur prior to the bodily or mental motion that is endorsed: we cannot endorse something that did not happen yet. For that reason, an endorsement cannot

---

[315] Frankfurt 1992, pp. 104-105 (his emphasis).

cause the bodily or mental motion that is endorsed. In this respect, an endorsement is very different from a desire or an intention, since we typically desire or intend states of affairs that have not happened yet. Thus, Mele's assumption that Al and Peter have a desire or intention to coast or to roll seems unjustified.

For the sake of the argument, let us suppose that we grant the first point to Mele and assume that Al has a desire or an intention to coast. Mele would argue, then, that the desire or the intention is the cause of the coasting. This is his second sub-argument. That point, however, also fails.

As we saw, Mele bases his view on the standard conditional analysis of causation. He argues that Al's desire or intention or decision causes the coasting, because "if Al had not desired, or intended, or decided to coast, he would not have coasted."[316] Schematically, we get:

- If ¬A, then ¬C

But Mele's argument is based on a series of implicit assumptions and, in fact, the picture looks more like this:

- If ¬A, then D and if D, then E and if E, then ¬C

To see why, imagine that we simply remove Al's desire or that we remove Al entirely from the car; then the car would have been coasting anyway (and thus we would obtain C). To reach the conclusion that the car is not coasting (¬C), Mele takes a different route and assumes a few things. First, he assumes that if Al had not had the desire to coast (¬A) then he would have had another desire (D); and if he had had another desire, then he would have done something else (E). And finally, if he had done something else (E), then he would not have been coasting (¬C).

---

[316] Mele 1997, p. 138.

It should be clear that this reasoning is problematic. For one thing, it is not the standard analysis of causation and Mele would need to defend his analysis better. But it is also problematic because the conclusion is not warranted. Suppose that if Al had not had the desire or the intention to coast, he would have had the desire to smoke a cigarette or the desire to read a book or the desire to play cards or the desire to text his friend. And, quite possibly, he could have done any of these things while being in the car. In these cases, the car would have been coasting anyway (C). Thus, Mele's conclusion that the car would not be coasting (¬C) if Al did not have the desire or intention to coast (¬A) is not warranted.[317]

Let us be charitable one more time and assume that this second point is valid. Then, Mele could examine a possible counterexample to his view, the case of would-be-indifferent-Al, and claim that this case does not show that there is no causation between the desire, the intention or the decision, and the coasting: rather, this case simply falsifies the standard conditional analysis of causation like a case of preemption. This is Mele's third sub-argument. That point, however, also fails.

The problem is that the case of would-be-indifferent-Al is not a case of preemption and that the comparison is mistaken. In cases of preemption, it is assumed that there are two causes that can operate independently. In the actual scenario, A causes C; in the counterfactual scenario, B causes C:

- **Actual scenario**: A causes C
- **Counterfactual scenario**: B causes C

[317] Zhu has made a similar point. He claimed that "the counterfactual condition that 'if Al had not desired, or intended, or decided to coast, he would not have coasted' does not entail that Al's continued coasting must have a mental cause, namely, his decision or intention. It only entails that Al continues to coast because he had not decided or intended to do otherwise" (Zhu 2004, p. 304).

Take Mele's example of the dialing. In the actual scenario, X dials Y's phone and Y's phone rings. In the counterfactual scenario, Z dials Y's phone and Y's phone rings.

The case of would-be-indifferent-Al is not like that: we do not have two independent causes. In the actual scenario, it is supposed that Al's decision to coast (A) causes the coasting *along with gravity* (B). Indeed, Al's decision is not sufficient to cause the coasting. Without gravity, the coasting would not have occurred. In the counterfactual scenario, on the other hand, gravity (B) alone causes the coasting. Schematically, we get:

- **Actual scenario**: (A and B) causes C
- **Counterfactual scenario**: B causes C

Thus, the case of would-be-indifferent-Al departs importantly from those cases of preemption. In this case, the actual scenario involves two causes.

Mele's third sub-argument fails for two reasons. First, Mele needs to explain why his case of would-be-indifferent-Al is an exception to the conditional analysis of causation (like preemption), and not a case where causation is lacking. Mele simply fails to provide such an explanation. Second, the case of would-be-indifferent-Al suggests that the decision to coast is causally inert. Indeed, it is very possible that Al's decision is causally inert since gravity alone (B) can cause the coasting. Imagine that God suspended gravity for a moment. Then, Al's decision to coast would simply be ineffective: it would not cause the coasting. What this case suggests, then, is that there is no causation between the decision (A) and the coasting (C), not that we have another exception to the conditional analysis of causation.

The case of would-be-indifferent-Al can be compared to the following scenario. Imagine that Al kicked a bridge (A) at the same time that an earthquake happened (B) and imagine that the bridge collapsed (C). In the actual scenario, one might say that the kicking and the

earthquake caused the collapse of the bridge [(A and B) causes C]. And one might also say that if the kicking had not occurred, the bridge would have collapsed anyway: the earthquake would have caused the collapse (B causes C). But this case does not appear to be an exception to the conditional analysis of causation. Rather, it seems to suggest that there is no causation between the kicking (A) and the collapse (C): the kicking alone cannot cause the collapse and thus it appears to be causally ineffective.

I turn now to Mele's second main argument, the view that Frankfurt's counterfactual strategy fails. To show why it fails, Mele develops a case involving a mind-reading demon who would paralyze Al if Al were to decide to intervene. Frankfurt has to conclude that Al is not acting or that he does not behave purposefully because he cannot intervene. But this seems counterintuitive to Mele. To strengthen his case, he contrasts this scenario with a scenario where Al is *actually* paralyzed. Such scenario does not involve agency because Al lacks control over his body.

Mele's argument, however, seems to fail because his description of the case is misleading. When Al is under the influence of a mind-reading demon, he is not merely paralyzed in the counterfactual scenario; he is *actually* paralyzed. Indeed, his body is moving in virtue of gravitational force alone and if he were to decide to move it in some other way and to intervene, the demon would stop him and prevent him from doing it. Thus, this looks very much like a case of actual paralysis: a case in which a person is incapable of moving his body. I contend that there is no significant distinction between this case and a standard case of paralysis. Thus, Frankfurt's conclusion concerning this case is warranted: Al is not acting nor behaving purposefully because his body is paralyzed by the demon (in the actual scenario).

7.6. Conclusion

In this chapter, I examined the causalism and anti-causalism debate about action explanation. I focused specifically on the exchange between Sehon and Mele which centers on a metaphysical dispute. I argued that the endorsement view can provide resources for an anti-causalist like Sehon to respond to Mele's critique. Moreover, I argued that Mele's potential objection to the endorsement view fails.

Conclusion

The present thesis consisted of two parts. To each of these two parts there corresponded one main aim. In Part I, I argued that Frankfurt's hierarchical approach is best seen as an account of agency. I made three arguments to support that view, one argument in each of the first three chapters. In Chapter 1, I showed that it is more contextually sensitive to interpret Frankfurt's approach as an account of agency. In particular, I argued that Frankfurt developed his hierarchical framework to provide an event-causal account of decision, and to show, *contra* Chisholm, that moral responsibility does not require the freedom to decide otherwise.

In Chapter 2, I argued that it is more precise to see Frankfurt's approach as an account of agency than it is to see it as an account of free agency or autonomy – the views that are often defended in the secondary literature. That is so because Frankfurt construes constraints as passive happening and heteronomy as a form of mental passivity. This view is not an intuitive one and thus is prone to misunderstanding. For that reason, we gain a sharper understanding of what is going on in his work when we reformulate his approach in agential terms.

In Chapter 3, I argued that Frankfurt's hierarchical framework is more convincing when it is interpreted as an account of agency. To defend that point, I examined a common critique of his approach: the view that it is over-inclusive or that it counts as free and autonomous too many actions that should not count as such. I argued that this critique is correct, but that it loses its strength when we formulate Frankfurt's approach in agential terms.

In Part II, I developed a Frankfurtian account of agency which I call the endorsement view. The main aim of the second part was to show that the endorsement view, which is a non-causal account of agency, is a serious contender within the metaphysical debate. The

metaphysical debate is a debate over the nature of agential control that emerged after the publication of Davidson's influential essay "Actions, Reasons, and Causes," an essay in which Davidson developed the "standard account." Up until recently, the non-causal approach was not seen as a serious contender within that debate, because it has been seen as lacking an account of agential control. What I showed in the second part is that it is possible to develop a compelling non-causal account of agential control.

In Chapter 4, I argued that the endorsement view is based on the non-causal concept of standby control, and that this account of control is the most compelling. It is the most compelling account of control because it provides the best solution to the problem of internal deviance. As we saw, the alternative approaches, the event-causal and the agent-causal approaches, either fail to solve the problem or they introduce further difficulties.

In Chapter 5, I examined the problem of the disappearing agent, which suggests that the standard account fails to provide an account of *agential* control. I argued, *contra* Schlosser, that the disappearing agent objection is a real issue for the standard account. Moreover, I distinguished between two versions of the problem, the missing source of activity objection and the missing humanity objection.

In Chapter 6, I examined the most popular solutions to the disappearing agent objection and argued that the endorsement view, when properly developed, provides the strongest solution. I showed that the main event-causal and agent-causal accounts of the agent are problematic. I also argued that Frankfurt's non-causal account is problematic. But I showed that his account can be modified in such a way that we can provide a satisfactory account of the agent.

In Chapter 7, I examined the causalism vs. anti-causalism debate in the epistemology of agency, and more specifically the debate between Sehon and Mele. I showed that the

endorsement view provides resources to anti-causalists, like Sehon, to defend the non-causal account of action explanation.

To conclude, I will explore some questions that would require further thought. The present thesis was focused on the metaphysics of agency. While I showed that my account of agency can be used to strengthen the case for the non-causal account of action explanation, I did not myself take any position concerning the nature of action explanation or the nature of reasons for action. A further question to ask is: what is the proper way to explain actions? Should we deny, like anti-causalists do, that reasons are causes? This is obviously a much-debated question that would require an extensive discussion.

Another question worth exploring concerns the distinction between basic and non-basic actions. An action is *non-basic* when it is performed by means of another action and it is *basic* when it is not the case that it is performed by means of another action. Consider Davidson's well-known example. I flip the switch with my finger, turn on the light, illuminate the room, and alert the prowler.[318] On one reading, my flipping the switch is a basic action and everything that I do by means of that action counts as a non-basic action. Thus, all the following count as non-basic actions: turning on the light, illuminating the room, and alerting the prowler. The notion of non-basic action might raise a problem for the endorsement view, because non-basic actions cannot involve standby control. While I have standing control over my finger movement, I do not have that kind of control over the electrical process that my finger movement triggers.

A way to solve the problem of non-basic actions is to deny their existence. This is a position defended by Davidson.[319] But what allows Davidson to defend that position is his

[318] Davidson 1963.
[319] Davidson 1971.

conception of action individuation. On Davidson's view, when I say that "I flip the switch," "turn on the light," "illuminate the room," and "alert the prowler," I am in fact describing the same action in four different ways.[320] Thus, it is mistaken to say that there are four actions, one of which is basic and three of which are non-basic. Rather, there is one action, which is basic, and four ways to describe it. Non-basic actions do not exist. But that position leads to a question that has been extensively debated: the nature of action individuation.[321] Thus, another topic that is worth exploring is whether we should individuate action coarsely, in the way Anscombe and Davidson do,[322] or finely.[323]

Another much debated question that would be relevant for the present thesis concerns the ontology of dispositions. I have assumed throughout the thesis that dispositions are part of the natural event-causal order because they can be analyzed in terms of events and relations among them, or in terms of categorical properties, events, and relations among them.[324] But this is a much debated question. Many counterexamples have been raised against these analyses, such as finking[325] and masking[326] cases. Thus, another question worth exploring concerns the best way to analyze dispositions.

Finally, there is a plethora of types of action that have been analyzed in the literature, including akratic actions,[327] free and autonomous actions, collective actions,[328] mental actions,[329]

---

[320] Davidson 1963.
[321] See for example Sandis 2010 and Hornsby 1979.
[322] See Anscombe 1963 and Davidson 1963, 1967.
[323] For an example of this view, see Goldman 1970.
[324] See Ryle 1949, Goodman 1954, and Quine 1960, Armstrong 1968, and Lewis 1997.
[325] Martin 1994.
[326] Johnston 1992 and Bird 1998.
[327] See, for example, Davidson 1970, Stroud and Tappolet 2003.
[328] See, for example, Searle 1990 and Pettit 2003.
[329] See, for example, O'Brien and Soteriou 2009.

animal actions,[330] and so on. Another question worth exploring is whether the endorsement view can contribute to our understanding of these other forms of actions. Does the endorsement view fare better than causal accounts in the analysis of these types of action? My intuition is that it does, but this would require further development.

[330] See, for example, Glock 2010.

Bibliography

Adams, F. and A. Mele (1989), "The Role of Intention in Intentional Action," *Canadian Journal of Philosophy*, vol. 19.

Aguilar, J. (2010), "Agential Systems, Causal Deviance, and Reliability," in J. Aguilar and A. Buckareff (eds), *Causing Human Actions: New Perspectives on the Causal Theory of Action*, Cambridge, MA: MIT Press.

———— (2012), "Basic Causal Deviance, Action Repertoires, and Reliability," Philosophical Issues, vol. 22, pp. 1–19.

Aguilar, J. and A. Buckareff (2010), "The Causal Theory of Action: Origins and Issues," in J. Aguilar and A. Buckareff (eds), *Causing Human Actions: New Perspectives on the Causal Theory of Action*, Cambridge, MA: MIT Press.

Allen, C. and M. Bekoff (1997), *Species of Mind: The Philosophy and Biology of Cognitive Ethology*, Cambridge, MA: MIT Press.

Alston, W. (1986), "An Action-Plan Interpretation of Purposive Explanations of Actions," *Theory and Decision*, vol. 20.

Alvarez, M. (2013), "Explaining Actions and Explaining Bodily Movements," in G. D'Oro and C. Sandis (eds.), *Reasons and Causes: Causalism and Anti-Causalism in the Philosophy of Action*, Palgrave Macmillan.

———— (2017), "Reasons for Action: Justification, Motivation, Explanation," *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL <https://plato.stanford.edu/archives/win2017/entries/reasons-just-vs-expl/>.

Alvarez, M. and J. Hyman (1988), "Agents and their Actions," *Philosophy*, vol. 73, pp. 218-245.

Andrews, K. (2016), "Animal Cognition," *The Stanford Encyclopedia of* Philosophy, E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2016/entries/cognition-animal/>.

Anscombe, G. E. M. (1957), *Intention*, Cambridge: Harvard University Press.

Armstrong, D. (1968), *A materialist Theory of the Mind*, London: Routledge.

Aquinas, St. T. (1265-1268), *Summa Theologiae*, Fathers of the English Dominican Province (trans.), Allen, TX: Christian Classics, 1981.

Aune, B. (1967), "Cans and Ifs: An Exchange," *Analysis*, vol. 27, pp. 191-195.

Austin, J. L. (1956), "Ifs and Cans," *Proceedings of the British Academy*.

Bach, K. (1978), "A Representational Theory of Action," *Philosophical Studies*, vol. 34.

Bekoff, M. and J. Pierce (2009), *Wild Justice: The Moral Lives of Animals*. Chicago: University of Chicago Press.

Benson, P. (1994), "Free Agency and Self-Worth," *The Journal of Philosophy*, vol. 91, no. 12.

Beran, M., J. Brandl, J. Perner, and J. Proust (eds.) (2012), *The Foundations of Metacognition*, Oxford: Oxford University Press.

Bird, A. (1998), "Dispositions and Antidotes," *The Philosophical Quarterly*, vol. 48, pp. 227–234.
Bishop, J. (1983), "Agent causation," *Mind*, vol. 92, pp. 61-79.
———— (1989), *Natural Agency*, Cambridge: Cambridge University Press.
Brand, M. (1984), *Intending and Acting: Toward a Naturalized Action Theory*, Cambridge, Ma: MIT Press.
Bratman, M. (1987), *Intention, Plans, and Practical Reason*, Cambridge, MA: Harvard University Press.
———— (2001), "Two Problems About Human Agency," *Proceedings of the Aristotelian Society*, vol. 101, pp. 309-326.
Call, J. and M. Carpenter (2001), "Do Apes and Children Know What They Have Seen?" *Animal Cognition*, vol. 4, pp. 207–220.
Candish. S. and N. Damnjanovic (2013), "Reasons, Actions, and the Will: The Fall and Rise of Causalism," in M. Beaney (ed.), *The Oxford Handbook of the History of Analytic Philosophy*, Oxford: Oxford University Press.
Chisholm, R. (1964), *Human Freedom and the Self*, Lawrence: University of Kansas.
Christman, J. (1987), "Autonomy: A defense of the Split-Level Self," *The Southern Journal of Philosophy*, vol. 25, no. 3.
———— (1989), "Introduction," in J. Christman (ed.), *The Inner Citadel*, Brattleboro: EPBM.
———— (1991), "Autonomy and Personal History," *Canadian Journal of Philosophy*, vol. 21, no. 1, pp. 1-24.
———— (2004), "Relational autonomy, liberal individualism, and the social constitution of selves," *Philosophical Studies*, vol. 117, no. 1–2, pp. 143–64.
———— (2018), "Autonomy in Moral and Political Philosophy," *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL =<https://plato.stanford.edu/archives/spr2018/entries/autonomy-moral/>.
Choi, S. and M. Fara (2018), "Dispositions," *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/dispositions/>.
Clarke, R. (2003), *Libertarian Accounts of Free Will*, Oxford: Oxford University Press.
———— (2010), "Agent Causation," in T. O'Connor and C. Sandis (eds.), *A Companion to the Philosophy of Action*, Wiley-Blackwell.
Crystal, J. D. and A. L. Foote (2009), "Metacognition in animals," *Comparative Cognition & Behaviour Reviews*, vol. 4, pp. 1–16.
D'Oro, G. (2007), "Two Dogmas of Contemporary Philosophy of Action," *Journal of the Philosophy of History*, vol. 1, pp. 11-26.
D'Oro, G. and C. Sandis (2013), "From Anti-Causalism to Causalism and Back," in G. D'Oro and C. Sandis (eds.), *Reasons and Causes: Causalism and Anti-Causalism in the Philosophy of Action*, Palgrave Macmillan.
Descartes, R. (1641), *Meditations on First Philosophy,* G. Heffernan (trans.), Notre Dame: University of Notre Dame Press, 1990.

Dancy, J. (1993), *Moral Reasons*, Oxford: Blackwell.
———— (1995), "Why There is Really No Such Thing as the Theory of Motivation," *Proceedings of the Aristotelian Society*, vol. 95, pp. 1-18.
———— (2000), *Practical Reality*, Oxford: Oxford University Press.
Davidson, D. (1963), "Actions, Reasons, and Causes," in D. Davidson, *Essays on Actions and Events*, Oxford: Oxford University Press, 2001.
———— (1967), "The Logical Form of Action Sentences," in D. Davidson, *Essays on Actions and Events*, Oxford: Oxford University Press, 2001.
———— (1970), "How Is Weakness of the Will Possible?," in D. Davidson, *Essays on Actions and Events*, Oxford: Oxford University Press, 2001.
———— (1971), "Agency," in D. Davidson, *Essays on Actions and Events*, Oxford: Oxford University Press, 2001.
———— (1973), "Freedom to Act," in D. Davidson, *Essays on Actions and Events*, Oxford: Oxford University Press, 2001.
———— (1982), "Rational Animals", *Dialectica*, vol. 3, no. 4, pp. 317–327.
Davies, M. (1983), "Function in Perception," *Australian Journal of Philosophy*, vol. 61, pp. 409-426.
Davis, W. A. (2010), "The Causal Theory of Action," in T. O'Connor and C. Sandis (eds.), *A Companion to the Philosophy of Action*, Wiley-Blackwell.
Dennett, D. (1984), *Elbow Room: the varieties of free will worth wanting*, Cambridge: MIT Press.
Doris, J. (2002), *Lack of Character*, Cambridge: Cambridge University Press.
Dretske, F. (1988), *Explaining Behaviour: Reasons in a World of Causes*, Cambridge, MA: MIT Press.
Dworkin, G. (1970), "Acting Freely," *Noûs*, vol. 4, no. 4, pp. 367-383.
———— (1976), "Autonomy and Behaviour Control," *Hastings Center Report*, vol. 6, pp. 23-28.
———— (1988), *The Theory and Practice of Autonomy*, Cambridge: Cambridge University Press.
Enç, B. (2003), *How We Act: Causes, Reasons, and Intentions*, Oxford: Oxford University Press.
Fischer, J. M. (2005), "General Introduction," in J. M. Fischer (ed.), *Free Will, Critical Concepts in Philosophy*, New-York: Routledge.
———— (2012a), "Responsibility and Autonomy: The Problem of Mission Creep," *Philosophical Issues*, vol. 22, no. 1, pp. 165–84.
———— (2012b), "Semicompatibilism and Its Rivals," *Journal of Ethics*, vol. 16, no. 2, pp. 117–43.
Fischer, J. M. and M. Ravizza (1998), *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.

Forrest, P. (2016), "The Identity of Indiscernibles," *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/identity-indiscernible/>.
Frankfurt, H. (1969), "Alternate Possibilities and Moral Responsibility," in H. Frankfurt, *The Importance of What We Care About*, Cambridge: Cambridge University Press, 1988.
———— (1971), "Freedom of the Will and the Concept of a Person," in H. Frankfurt, *The Importance of What We Care About*, Cambridge: Cambridge University Press, 1988.
———— (1973), "Coercion and Moral Responsibility," in H. Frankfurt, *The Importance of What We Care About*, Cambridge: Cambridge University Press, 1988.
———— (1975), "Three Concepts of Free Action," in H. Frankfurt, *The Importance of What We Care About*, Cambridge: Cambridge University Press, 1988.
———— (1976), "Identification and Externality," in H. Frankfurt, *The Importance of What We Care About*, Cambridge: Cambridge University Press, 1988.
———— (1978), "The Problem of Action," in H. Frankfurt, *The Importance of What We Care About*, Cambridge: Cambridge University Press, 1988.
———— (1987), "Identification and Wholeheartedness," in H. Frankfurt, *The Importance of What We Care About*, Cambridge: Cambridge University Press, 1988.
———— (1989), "Concerning the Freedom and Limits of the Will," in H. Frankfurt, *Necessity, Volition, and Love*, Cambridge: Cambridge University Press, 1999.
———— (1992), "The Faintest Passion," in H. Frankfurt, *Necessity, Volition, and Love*, Cambridge: Cambridge University Press, 1999.
———— (2002), "Reply to Michael E. Bratman," in S. Buss and L. Overton (eds), *Contours of Agency: Essays on Themes from Harry Frankfurt*, Cambridge, Mass: MIT Press.
———— (2004), *The Reasons of Love*, Princeton: Princeton University Press.
Friedman, M. (1986), "Autonomy and the split-level self," *The Southern Journal of Philosophy*, vol. 24, no. 1.
Ginet, C. (1990), *On Action*, Cambridge: Cambridge University Press.
———— (1996), "In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Argument Convincing," *Philosophical Perspectives*, vol. 1996, pp. 403-417.
Glock, H.-J. (2010), "Animal Agency," in T. O'Connor and C. Sandis (eds.), *A Companion to the Philosophy of Action*, Wiley-Blackwell.
Goldman, A. (1970), *A Theory of Human Action*, Princeton: Princeton University Press,
Goodman, N. (1954), *Fact, Fiction and Forecast*, Cambridge, Mass.: Harvard University Press.
Gorman, A. (2019), "The Minimal Approval Account of Attributability," in D. Shoemaker (ed.), *Oxford Studies in Agency and Responsibility Volume 6*, Oxford: Oxford University Press.
Hacker, P. M. S. (1996), *Wittgenstein: Mind and Will. Analytical Commentary on the Philosophical Investigations*, vol. 4, part 1, Oxford: Wiley-Blackwell.
———— (2009), "Agential Reasons and the Explanation of Human Behaviour," in C. Sandis (ed.), *New Essays on the Explanation of Action*, Palgrave-MacMillan, pp. 75-93.

Haji, I. (1998), *Moral Appraisability: Puzzles, Proposals, and Perplexities*, Oxford: Oxford University Press.
———— (2002), "Compatibilist Views of Freedom and Responsibility," in R. H. Kane (ed.), *The Oxford Handbook of Free Will*, Oxford: Oxford University Press, pp. 202–28.
Hampton, R. R. (2001), "Rhesus Monkeys Know When They Remember," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, pp. 5359–5362.
Harcourt, E. (2010), "Action Explanation and the Unconscious," in T. O'Connor and C. Sandis (eds.), *A Companion to the Philosophy of Action*, Wiley-Blackwell.
Hieronymi, P. (2008), "Responsibility for Believing," *Synthese*, vol. 161, pp. 357-373.
———— (2009), "The Will as Reason," *Philosophical Perspectives*, vol. 23, *Ethics*, pp. 201-220.
———— (2014), "Reflection and Responsibility," *Philosophy and Public Affairs*, vol. 42, no. 1, pp. 3-41.
Hobbes, T. (1640), *Human Nature and De Corpore Politico: The Elements of Law, Natural and Politic*, Oxford: Oxford University Press, 1994.
Hornsby, J. (1979), "Actions and Identities," *Analysis*, vol. 39, no. 4, pp. 195–201.
———— (2004), "Agency and Actions," in J. Hyman and H. Steward (eds.), *Agency and action*, Cambridge: Cambridge University Press.
Hutto, D. (1999), "A Cause for Concern: Reasons, Causes and Explanations," *Philosophical and Phenomenological Research*, vol. 59, no. 2, pp. 381-401.
Johnston, M. (1992), "How to Speak of the Colors," *Philosophical Studies*, vol. 68, pp. 221–263.
Lehrer, K. (1966), "An Empirical Disproof of Determinism," in K. Lehrer (ed.), *Freedom and Determinism*, New-York, pp. 175-202.
Lewis, D. (1973), "Causation," *Journal of Philosophy*, vol. 70, pp. 556–567.
———— (1997), "Finkish Dispositions," *Philosophical Quarterly*, vol. 47, pp. 143.
Locke, D. (1975), "Three Concepts of Free Action," *Aristotelian Society Supplementary Volume*, vol. 49, no. 1, pp. 95–126.
Locke, J. (1689), *An Essay Concerning Human Understanding*, P. H. Nidditch (ed.), URL=< doi:10.1093/actrade/9780198243861.book.1/actrade-9780198243861-book-1>, 1975.
Lowe, E. J. (2008), *Personal Agency*, Oxford: Oxford University Press.
———— (2010), "Action Theory and Ontology," in T. O'Connor and C. Sandis (eds.), *A Companion to the Philosophy of Action*, Wiley-Blackwell.
Mackenzie, C. (2014), "Three Dimensions of Autonomy: A Relational Analysis," in A. Veltman and M. Piper (eds.), *Autonomy, Oppression, and Gender*, Oxford: Oxford University Press.
Mackenzie, C. and N. Stoljar (2000), "Introduction: Autonomy Refigured," in C. Mackenzie and N. Stoljar (eds.), *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, New-York: Oxford University Press.
Mayr, E. (2011), *Understanding Human Agency*, Oxford: Oxford University Press.

Martin, C. B. (1994), "Dispositions and Conditionals," *Philosophical Quartely*, vol. 44, no. 1.
McCann, H. J. (1998), *The Works of Agency: On Human Action, Will, and Freedom*, Ithaca: Cornell University Press.
McKenna, M. (2011), "Contemporary Compatibilism: Mesh Theories and Reasons-Responsive Theories," in R. Kane (ed.), *Oxford Handbook of Free Will, 2nd ed.*, New York: Oxford University Press, pp. 175–198.
McKenna, M. and J. Coates (2018), "Compatibilism," *The Stanford Encyclopedia of Philosophy*, E. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2018/entries/compatibilism/>.
McKennam M. and C. Van Schoelandt (2015), "Crossing a Mesh Theory with a Reasons-Responsive Theory," in A. Buckareff, C. Moya, and S. Rosell (eds), *Agency and Responsibility*, Basingstoke: Palgrave Macmillan, pp. 44–64.
Melden, A. I. (1961), *Free Action*, London and New-York: Routledge Library Editions.
Mele, A. (1988), "Effective Reasons and Intrinsically Motivated Actions," *Philosophy and Phenomenological Research*, vol. 48, no. 4, pp. 723-731.
———— (1992), *Springs of Action*, Oxford: Oxford University Press.
———— (1997), "Passive Action," in G. Holstrom-Hintikka and R. Tuomela (eds.), *Contemporary Action Theory*, Kluwer Academic Publishers.
———— (2000), "Goal-directed Action: Teleological Explanations, Causal Theories, and Deviance," *Noûs*, vol. 34, no. 14, pp. 279-300.
———— (2003), *Motivation and Agency*, Oxford: Oxford University Press.
———— (2010), "Teleological Explanations of Actions: Anticausalism versus Causalism," in J. Aguilar and A. Buckareff, *Causing Human Actions: New Perspectives on the Causal Theory of Action*, Cambridge, MA: MIT Press.
———— (2013), "Actions, Explanations, and Causes," in G. D'Oro and C. Sandis (eds.), *Reasons and Causes: Causalism and Anti-Causalism in the Philosophy of Action*, Palgrave Macmillan.
———— (2019), "On Action Explanation and Causal Deviance," in G. Schumann (ed.), *Explanation in Action Theory and Historiography: Causal and Teleological Approaches*, New York and London: Routledge.
Mellor, D. H. (1995), *The Facts of Causation*, London: Routledge.
Menzies, P. and H. Beebee (2020), "Counterfactual Theories of Causation," *The Stanford Encyclopedia of Philosophy*, E. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2020/entries/causation-counterfactual/>.
Meyers, D. T. (1989), *Self, Society, and Personal Choice*, New-York: Columbia University Press.
Moore, G. E. (1911), *Ethics*, London: Williams & Norgate.
Morton, A. (1975), "Because He Thought He Had Insulted Him," *Journal of Philosophy*, vol. 74, pp. 261-301.

Nagel, E. (1977), "Goal-Directed Processes in Biology," *The Journal of Philosophy*, vol. 74, no. 5, pp. 261-279.
Nagel, T. (1986), *The View from Nowhere*, Oxford: Oxford University Press.
O'Brien, L. and M. Soteriou (eds) (2009), *Mental Action*. Oxford: Oxford University Press.
O'Connor, T. (2000), *Persons and Causes: The Metaphysics of Free Will*, Oxford: Oxford University Press.
O'Connor, T. and C. Franklin (2020), "Free Will," *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2020/entries/freewill/>.
Oshana, M. (1998), "Personal Autonomy and Society," *Journal of Social Philosophy*, vol. 29, no. 1, pp. 81-102.
Papineau, D. (2020), "Naturalism," *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/entries/naturalism/>.
Peacocke, C. (1979), *Holistic Explanation, Action, Space, Interpretation*, Oxford: Oxford University Press.
Pereboom, D. (2014), *Free Will, Agency, and Meaning in Life*, New York: Oxford University Press.
Pettit, P. (2003), "Groups with Minds of Their Own," in F. Schmitt (ed), *Socializing Metaphysics: the Nature of Social Reality*, Lanham, MD: Rowman & Littlefield, pp. 167–93.
———— (2012), "The Inescapability of Consequentialism," in U. Heuer and G. Lang (eds), *Luck, Value, and Commitment: Themes from the Ethics of Bernard Williams*, Oxford: Oxford University Press.
Pink, T. (2010), "Reason, Voluntariness, and Moral Responsibility," in L. O'Brien and M. Soteriou (eds), *Mental Actions*, Oxford: Oxford University Press.
Pollard, B. (2003), "Can Virtuous Actions Be Both Habitual and Rational?" *Ethical Theory and Moral Practice*, vol. 6, pp. 411-425.
———— (2005), "Naturalizing the Space of Reasons," *International Journal of Philosophical Studies*, vol. 13, pp. 69-82.
———— (2006), "Explaining Actions with Habits" *American Philosophical Quarterly*, vol. 43, pp. 57-68.
———— (2010), "Habitual Actions," in T. O'Connor and C. Sandis, *A Companion to the Philosophy of Action*, West-Sussex: Wiley-Blackwell.
Proust, J. (2010), "Mental Acts," in T. O'Connor and C. Sandis, *A Companion to the Philosophy of Action*, West-Sussex: Wiley-Blackwell.
———— (2013), *The Philosophy of Metacognition: Mental Agency and Self-Awareness*, Oxford: Oxford University Press.
Quine, W. V. (1960), *Word and Object*, Cambridge: MIT Press.
Rachels, J. (1975), "Active and Passive Euthanasia," *New England Journal of Medicine*, vol. 292, pp. 78–86.

Raz, J. (1999), *Engaging Reason: On the Theory of Value and Action*, Oxford: Oxford University Press.
Reid, T. (1788), *Essays on the Active Powers of Man*, K. Haakonssen and J. A. Harris (eds.), Edinburgh: Edinburgh University Press, 2010.
Roth, A. (2000), "Reasons Explanation of Actions: Causal, Singular, and Situational," *Philosophy and Phenomenological Research*, vol. 59, pp. 839–74.
Ruben, D.-H. (2009), "Con-Reasons as Causes," in C. Sandis (ed.), *New Essays on the Explanation of Action*, Palgrave-Macmillan, pp. 62-74.
Rundle, B. (1997), *Mind in Action*, Oxford: Oxford University Press.
Ryle, G. (1949), *The Concept of Mind*, London and New-York: Routledge.
Sandis, C. (2009), "Introduction," in C. Sandis (ed.), *New Essays on the Explanation of Action*, Palgrave-Macmillan.
———— (2010), "Basic Actions and Individuation," in T. O'Connor and C. Sandis (eds.), *A Companion to the Philosophy of Action*, Wiley-Blackwell.
Scanlon, T. M. (1998), *What We Owe to Each Other*, Cambridge: Harvard University Press.
Schlosser, M. E. (2007), "Basic Deviance Reconsidered," *Analysis* 67: 186–94.
———— (2011), "Agency, Ownership, and the Standard Theory," in J. H. Aguilar (ed), *New Waves in the Philosophy of Action*, London: Palgrave Macmillan.
———— (2019), "Agency," *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2019/entries/agency/>.
Schroeder, S. (2001), "Are Reasons Causes? A Wittgensteinian Response to Davidson," in S. Schroeder (ed.), *Wittgenstein and Contemporary Philosophy of Mind*, 1st ed., Basingstoke: Palgrave Macmillan, pp. 150–170.
Schroeter, F. (2004), "Endorsement and Autonomous Agency," *Philosophy and Phenomenological Research*, vol. 69, pp. 633–59.
Schueler, G. F. (2003), *Reasons and Purposes*, Oxford: Clarendon Press.
———— (2009), "Interpretive Explanations," in C. Sandis (ed.), *New Essays on the Explanation of Action*, Palgrave-Macmillan.
———— (2019), "Why and How? Teleological and Causal Concepts in Action Explanation," in G. Schumann (ed.), *Explanation in Action Theory and Historiography: Causal and Teleological Approaches*, New York and London: Routledge.
Schumann, G. (2019), "Introduction," in G. Schumann (ed.), *Explanation in Action Theory and Historiography: Causal and Teleological Approaches*, New York and London: Routledge.
Searle, J. (1990), "Collective Intentions and Actions," in P. Cohen, J. Morgan and M. Pollack (eds), *Intentions in Communication*, Cambridge: MIT Press, pp. 401–415.
Sehon, S. R. (1994), "Teleology and the Nature of Mental States," *American Philosophical Quarterly*, vol. 31, no. 1, pp. 63-72.
———— (1997), "Deviant Causal Chains and the Irreducibility of Teleological Explanation," *Pacific Philosophical Quarterly*, vol. 78, pp. 195-213.

———— (2005), *Teleological Realism: Mind, Agency, and Explanation*, Cambridge, MA: Bradford Book/MIT Press.
———— (2016), *Free Will and Action Explanation: A Non-Causal, Compatibilist Account*, Oxford: Oxford University Press.
Shettleworth, S. J. and J. E. Sutton (2006), "Do Animals Know What They Know?" in S. Hurley and M. Nudds (eds.), *Rational Animals?*, Oxford: Oxford University Press, pp. 235–246.
Shoemaker, D. (2011), "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility," *Ethics*, vol. 121, no. 3.
———— (2015), *Responsibility from the Margins*, New York: Oxford University Press.
Smith, A. (2015), "Responsibility for Attitudes: Activity and Passivity in Mental Life," *Ethics*, vol. 115, no. 2, pp. 236–271.
Smith, J. D., J. Schull, et al. (1995), "The Uncertain Response in the Bottlenose Dolphin," *Journal of Experimental Psychology General*, vol. 124, pp. 391–408.
Steward, H. (2013), "Processes, Continuants and Individuals," *Mind*, vol. 122, no. 487, pp. 781–812.
Stout, R. (2010), "Deviant Causal Chains," in T. O'Connor and C. Sandis, *A Companion to the Philosophy of Action*, Wiley-Blackwell.
Strabbing, J. T. (2016), "Attributability, Weakness of Will, and the Importance of Just Having the Capacity," *Philosophical Studies*, vol. 173, no. 2, pp. 289–307.
Stroud, S. and C. Tappolet (eds.) (2003), *Weakness of Will and Practical Irrationality*, Oxford: Clarendon Press.
Stump, E. (1988), "Sanctification, Hardening of the Heart, and Frankfurt's Concept of Free Will," *Journal of Philosophy*, vol. 85, no. 8, pp. 395-420.
Talbert, M. (2019), "Moral Responsibility," *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2019/entries/moral-responsibility/>.
Tanney, J. (1995), "Why Reasons May Not Be Causes," *Mind and Language*, vol. 10, pp. 103-126.
———— (2005), "Reason-Explanation and the Contents of the Mind," *Ratio*, vol. 18, no. 3, pp. 338–51.
———— (2009), "Reasons as Non-Causal, Context-Placing Explanations," in C. Sandis (ed.), *New Essays on the Explanation of Action*, Palgrave-Macmillan.
Tappolet, C. (2016), *Emotions, Values, and Agency*, New-York: Oxford University Press.
Taylor, J. S. (2005), "Introduction," in J. S. Taylor (ed.), *Personal Autonomy*, Cambridge: Cambridge University Press.
Taylor, R. (1966), *Action and Purpose*, Englewood Cliffs: Prentice-Hall, Inc.
Thalberg, I. (1978), "Hierarchical Analyses of Unfree Action," *Canadian Journal of Philosophy*, vol. 8, no. 2, pp. 211-226.

———— (1984), "Do our Intentions Cause Our Intentional Actions?" *American Philosophical Quarterly*, vol. 21, no. 3, pp. 249-260.
Velleman, D. (1992), "What Happens When Someone Acts?," *Mind*, vol. 101, no. 403.
Vihvelin, K. (1994), "Are Drug Addicts Unfree?" in S. Luper-Foy and C. Brown (eds), *Drugs, Morality and the Law*, New York: Garland, pp. 51–78.
Von Wright, G. H. (1971), *Explanation and Understanding*, Ithaca: Cornell University Press.
Wallace, R. J. (1999), "Three Conceptions of Rational Agency," *Ethical Theory and Moral Practice*, vol. 2, no. 3, pp. 217-242.
Watson, G. (1975), "Free Agency," *Journal of Philosophy*, vol. 72, pp. 205-220.
Westlund, A. C. (2009), "Rethinking Relational Autonomy," *Hypatia*, vol. 24, no. 4.
Widerker, D. (1995), "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *The Philosophical Review*, vol. 104, no. 2, pp. 247-261.
Wilson, G. (1980), *The Intentionality of Human Action*, Amsterdam: North-Holland Publishing Company.
Wilson, G. and S. Shpall (2016), "Action," *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/action/>.
Wolf, S. (1988), "Sanity and the Metaphysics of Responsibility," in F. Schoeman (ed.), *Responsibility, Character, and the Emotions*, Cambridge: Cambridge University Press.
———— (1990), *Freedom Within Reason*, Oxford: Oxford University Press.
Woollard, F. and F. Howard-Snyder, (2016), "Doing vs. Allowing Harm," The Stanford Encyclopedia of Philosophy, E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/doing-allowing/>.
Wu, W. (2016), "Experts and Deviants: The Story of Agentive Control," *Philosophy and Phenomenological Research*, vol. 92, no. 2, pp. 101-126.
Zhu, J. (2004), "Passive Action and Causalism," *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, vol. 119, No. 3, pp. 295-314
Zimmerman, M. J. (2010), "Chisholm," in T. O'Connor and C. Sandis (eds.), *A Companion to the Philosophy of Action*, Wiley-Blackwell.