

Managerial Control and Free Mental Agency

TILLMANN VIERKANT

In this chapter it is argued that insights from recent literature on mental agency can help us to better understand what it is that makes us free agents. According to Pamela Hieronymi (2009), who developed Richard Moran's (2001) work on mental agency, there are two quite different ways in which we can be mental agents—either through “evaluative control” or through “managerial control.” According to Hieronymi, managerial control works very much like other forms of intentional action, whereas evaluative control is different and distinctive of mental agency. The first section of this chapter will discuss why the distinction introduced by Hieronymi is a good one, and will then go on to argue that Hieronymi nevertheless underestimates the importance of managerial control. This is because, as the chapter will argue, managerial control is central to free mental agency.

The chapter argues that managerial control is crucial for the will not because it enhances our understanding of our reasons, as one might easily assume, but because it creates an opportunity for the individual to change their beliefs and desires at will despite their own first-order rational evaluations. The discussion of the distinction between evaluative and managerial/manipulative control in Hieronymi will help us to see that there is no such thing as intentional rational evaluation, and what the intentional control of the mental is really good for. The last section of the chapter then tries to clarify what exactly is required for managerial control, in order for it to fulfill its function for the will and how this account compares to seemingly similar moves made by Michael Bratman and Richard Holton.

MENTAL ACTIONS

Hieronymi (2009) has argued that hierarchical accounts of mental agency fail to take into account what is at the heart of mental agency. This is because, according to Hieronymi, there are two distinct forms of mental agency. One, which she refers to as *managerial/manipulative control*,¹ works very much like bodily agency. The other form is referred to as *evaluative control* and lacks some of the most important

managerial vs.
evaluative
control

standard features of ordinary actions. Hieronymi nevertheless thinks that it is evaluative control that is the more fundamental form of mental agency.

To understand why she thinks this, let us have a short look at the two different forms of control. As stated, acts of managerial control work like ordinary actions. The agent forms an intention about an object in the world, and this intention is involved in bringing it about that the object is manipulated in a way that expresses the content of the intention. **One crucial feature of action so understood is the reflective distance between the agent and the object of the action.** In managerial control, we manipulate mental objects in exactly the same way as we would physical objects in bodily actions. We form an intention about a mental state that we want to manipulate, and our intention helps to bring it about that the relevant manipulation takes place.

But even though this form of mental agency clearly does exist, many have doubted that most of our mental activity works according to this model (see, e.g., Moran 2001; McGeer 2007; Hieronymi 2009; McHugh 2011). **It seems that, for example, the forming of a judgment or an intention does not normally work according to this model.**

It seems absurd to think that we always need an intention with the content “form an intention” in order to form one. Equally strange seems the notion that we normally should have to have intentions to form a judgment. The formation of judgments and intentions simply happens as a result of our ongoing evaluation of the world. We deliberate about what would be the best thing to do or what the facts of the matter are. As a result of this activity, we come to conclusions that can take the form of intentions and judgments, but our deliberations were not about mental attitudes,² they were about the world. We deliberated, for example, about whether we should go to the football match, or whether or not our football team would win the Champion’s League.

In fact, in many cases there seems to be something bordering on the incoherent about the idea that we should acquire a belief or an intention in a way that is not content-directed deliberation, but by forming an intention that is directed at acquiring an attitude.

This can be easily demonstrated, if one considers the following: if it were up to us whether or not we believe a proposition, then it should be quite possible for us to decide to acquire a false belief.

Famously, Moore points out that there is something very odd about this. Consider the sentence, “I believe that p, but p is false.” This sentence sounds paradoxical because whether or not a person believes something is not normally dependent on whether the person forms the intention to have that belief but on the plausibility of its content. Similarly for intentions, Kavka’s famous toxin puzzle seems to show that acquiring a desirable intention simply for the intention’s sake seems impossible, if the agent knows that she will have no reason to actually act on the intention. If, for example, you are offered €10,000 for having the intention at midnight to drink a mildly unpleasant toxin the next day, but you know that you will not actually have to drink it in order to receive the money, then it becomes puzzling to understand how you can acquire the intention to drink the toxin, given that you know that you will have no motivation to drink it after midnight has passed. It seems, then, that

normally at least we do not acquire attitudes like beliefs or intentions in the same way as we achieve our aims in bodily actions, because there does not seem to be the same element of intentional control or the reflective distance typical of that kind of control.

Obviously, however, there are nonstandard cases that complicate the picture. Kavka carefully rules out hypnosis in order to get the puzzle going, and in the belief case there is among many others the famous example of Pascal's wager. Pascal argued that it is rational to acquire the belief that God exists, even if there is little evidence for that belief. Faced with Moore's paradox, Pascal advises that one can still acquire the belief by going to mass, praying rosaries, and the like. What Pascal advises here in effect is basically a form of self-conditioning. So, it is quite possible to acquire mental attitudes for their own sake. Nevertheless, when an agent does acquire an attitude in a managerial way, she bypasses her own rationality, and her mind becomes a simple psychological object to manipulate. This is exactly what we should expect, if mental agency is modeled on bodily agency, but it seems clear that this is not the ordinary way of acquiring beliefs or intentions.

Because of this, and because of the very strong intuition that deliberating is something that we do rather than something that merely happens to us, Hieronymi argues that we should introduce a second form of mental agency that better describes the deliberative process, even if it fails to exhibit many of the characteristics that we ordinarily associate with agency. This is why she introduces evaluative control. Hieronymi has a positive and a negative reason for insisting that evaluative control really is a form of agency. The positive reason is that evaluative control is nothing else than the agent's rational machinery in action. The reflective distance that is so important for bodily action simply does not seem adequate when talking about the activity of the mind. Our deliberations are not something external to us, but express our understanding of the world. When we want to find out whether we judge that p we do not introspect to find our judgment there, but we look at the evidence for or against p.

The negative reason Hieronymi gives is that there simply is no alternative adequate account that would allow us to understand most of our judgments and intendings as actions. This is because, even though she acknowledges that there is a second form of mental agency—that is, the managerial control mentioned earlier—she does not believe that this managerial control can explain most of our judgments or intendings, nor does she believe that managerial control is in any case a completely independent form of mental agency. It is easy to see why: managerial control, like ordinary bodily action, requires an intention that can help to bring about the desired effect in the world. In managerial control the relevant intention would have as its content an intention or judgment that the agent would like to acquire. The problem is obviously that the intention that is controlling the managerial act is itself in need of being formed. Even in the highly unlikely case where the formation of this intention was also done in a managerial way, we have obviously entered a vicious circle. In the end, there will have to be an intention that has been brought about without the use of a previous intention to acquire the intention, and this intention will presumably be acquired by an act of evaluative control. In effect, then, every instance of managerial control will require at the very least one instance of evaluative control to get off

against
managerial
control as
explaining
judgment

the ground.³ The agent has to form the intention evaluatively to bring it about that she will acquire the relevant attitude. Pascal, for example, has to evaluatively acquire the judgment that it would be best, all things considered, to have the belief in God. Similarly, the agent in the toxin puzzle has to judge evaluatively that she should hypnotize herself in order to acquire the intention to drink the toxin.

Evaluative control, then, so Hieronymi's conclusion, is the basic form of mental agency. It is the way in which we ordinarily acquire attitudes like beliefs and intentions, and we should not be worried about this because in contrast to managerial control, in evaluative control we express ourselves as rational deliberating beings.

REFLECTIVE CONTROL

Evaluative control is indispensable, because on Hieronymi's picture there is no alternative account of mental agency that could fulfill the functions of evaluative control and have at the same time the same features as ordinary bodily agency. This claim, one might think, is wrong, though: there seems to be an alternative, which she labels *reflective control*. It seems to be quite possible to intentionally reflect and thereby change mental attitudes. An agent might, for example, think about placing a bet on her team winning the Champion's League. She might then form the intention that she should reexamine her reasons for her belief. It seems that she can now easily and fully intentionally think about the individual reasons she has for her belief. For example, are the forwards really that strong? Are the opponents really vulnerable on the left wing? She might, again fully intentionally, go through a list of things that are important for success and of which she might not have thought so far (proneness to injury, whether the referee likes the aggressive style of her team, etc.). Now the great thing is that even though it seems that these are things that an agent clearly can do intentionally, they seem as well to exhibit the same characteristics as evaluative control. If the agent were to find upon reflection that her team was not quite as strong as she originally thought, then she would change her belief that the team will win the Champion's League. It does not seem that this has happened in a way that bypasses rationality as in the managerial scenario, but that reflection makes the agent more rational because she now does not purely deliberate about the content of her belief, but explicitly about the question whether or not the attitude in question is justified.

But even though reflective control does seem very tempting, there is an obvious problem. How exactly does reflection bring about a change in attitude? Obviously, we can intentionally reflect on our reasons for a specific belief, but whether or not this reflection will bring about a change in attitude depends on our rational evaluation of the reasons, and it is not up to us (in the intentional sense) how that evaluation goes. Hieronymi therefore suspects that at the heart of reflective control we will find an exercise in evaluative control, which is doing all the work and which obviously does not have the intentional characteristics that seemed to make reflective control so attractive.

So is there a way, then, to make sense of reflective control without falling back on evaluative control? The most promising account, according to Hieronymi, would be a hierarchical account. According to such an account, when one reflects on one's

reasons for a belief that p and finds that those reasons are not sufficient, one will then form a second-order belief that the first-order belief is unjustified. Once one has this belief, all that needs to be the case for the first-order belief to change is that the first-order belief is sensitive to that second-order belief. **This sounds like a good account, but Hieronymi's next move notes an obvious problem. The account does not give us a story about how it is that the second-order belief will change the first-order belief.** Now, once we look closer at how this could happen, it becomes clear that this will not happen in a way that resembles intentional control.

Let's take stock of what we discussed so far. We followed Hieronymi's convincing defense of evaluative control as a specific form of mental agency that is importantly different from ordinary bodily agency. We saw as well that there is a second form of mental agency (managerial control), and that this form of mental agency can be modeled successfully on ordinary intentional actions. However, as Hieronymi pointed out, managerial control requires at the very least one instance of evaluative control in order to get off the ground and is therefore not a completely independent alternative to evaluative control. We then wondered whether reflective control, understood as a higher-order account of mental agency, could not fulfill the same function as evaluative control, while at the same time being a form of intentional control. A closer look revealed that reflective control necessarily has at its heart acts of evaluative control, and that these obviously cannot be modeled as intentional control.

MANAGERIAL CONTROL AND SELF-CONTROL

The hierarchical account of free mental agency seemed attractive because it provided us with a way of combining deliberative mental agency with intentional mental agency, but as Hieronymi convincingly argues, this combination does not work. The deliberative part in higher-order accounts looks on all consistent models very much like evaluative control. If one is convinced by this line of reasoning, higher-order accounts do lose their intuitive appeal as being able to combine rational with intentional control. One obvious consequence one could draw from this would be to abandon such higher-order accounts and to focus instead on developing a fuller picture of mental agency along the lines of evaluative control.

Alternatively, one could pursue the idea that perhaps higher-order accounts of free mental agency are correct after all, but that this is not related to the alleged features of reflective control. This is the direction this chapter will pursue. The idea starts from the thought that perhaps higher-order beliefs and desires do not make the agent more rational but rather allow her valuable room for maneuver with regard to her own evaluative processes. Perhaps this additional wiggle room is so important that it makes sense to argue that the ability to manipulate one's mind is a necessary condition for *free* mental agency.

That managerial control has its uses has already been discussed. Only by means of managerial control can Pascal overcome his skepticism, and the toxin puzzle does not seem very puzzling any longer as soon as we allow the agent to exercise managerial control over herself. But admitting that managerial control can be useful is obviously by far not enough to justify the much stronger claim that managerial control

is a necessary condition for free mental agency. Two immediate objections might be raised here.

First of all, these are rather contrived examples. It is obviously true that agents can manipulate themselves in the way the examples suggest, but the overwhelming majority of our mental acts are not like that. We do not normally think about how we can influence our attitudes, but we think about first-order content. We are interested in what is true about our world and the best thing to do in the world we live in. **A focus on our own attitudes sounds as unlikely as it sounds strangely narcissist.**

Second, even if one were convinced of the importance of managerial control, it would still be the case that managerial control is always parasitic on at least one act of evaluative control. Pascal goes to mass because he evaluated that this is the best thing to do if he wants to acquire the desired belief, and the agent in the toxin puzzle evaluatively forms the intention to hypnotize herself to get the cash.

Let us look at this second objection first. If it really were the case that the proposed account suggested that managerial control was a completely independent form of mental agency, then this would be a knock-down argument, but this is not what the claim amounts to. Rather, **the account accepts that the most basic form of mental agency is evaluative control, but it wants to add that evaluative control on its own is not enough for free mental agency.**

Let us now move on to the first objection. In answering this objection, we will as well encounter the central argument for the main claim of the chapter, that is, that managerial control is necessary for free mental agency. In order to get the answer off the ground, it will be helpful to have a closer look at an idea that Victoria McGeer (2007) explores in her essay “The Moral Development of First-Person Authority,” because her account is in many important ways similar to the one developed here. In addition, some of the topics McGeer discusses will prepare us for the crucial discussion of what it is that is important about managerial control.

McGeer, in contrast especially to Moran, does think that managerial control is extremely important for the moral development of an imperfect rational creature. She identifies two problems for such an agent. **On the one hand, there is the problem of rampant rationalization.**⁴ Agents might be able to rationalize their behavior even though it is perfectly obvious to a neutral observer that their behavior is actually controlled by different motives from the ones that they ascribe to themselves. **Second, even if an agent is aware of the right thing to do most of the time, this does not mean that there cannot be situations where their judgment is changed in undesirable ways because of the strong affordances of the situation.** McGeer discusses the example of a *Middlemarch* character who sincerely believes in a marriage between two friends of his, but who has the problem that he is quite fond of the girl himself. In order to stop himself from giving in to temptation, the character reveals his feelings for the girl to his male friend. By confessing his feelings, the agent makes it impossible for himself to pursue his desire for the girl. In other words, the agent uses his knowledge about his potentially changing psychology in order to prevent the feared changes.

Taking McGeer’s musings about her *Middlemarch* characters as a starting point, we can now return to the worry that managerial control—even though clearly

useful—is simply not of enough relevance in our lives to justify the strong claim of making it a necessary condition for free mental agency.

The crucial point here is that even though Pascal and the toxin puzzle describe very unusual situations, McGeer's story does not. In effect, McGeer describes a case of self-control. The character binds himself to the mast in a way that is structurally very similar to the archetype of all self-control stories. Like Odysseus, the character knows that he cannot be sure that his rational evaluation of the situation will remain constant if the wrong circumstances should arise. Self-control delivers the wiggle room mentioned earlier because it allows the agent to keep believing p , even under circumstances where the agent normally would tend to reevaluate and believe $\neg p$.

Now if it were the case that managerial control is necessary for us to be able to exercise future-directed acts of self-control, then it seems very plausible to maintain that it is a necessary condition for free mental agency, because I take it to be uncontroversial that the ability for future-directed self-control is at least a necessary condition for free mental agency. Before we move on, one very important difference between the account defended here and McGeer should be pointed out. McGeer argues that an ideally rational creature might not need these self-control techniques, and this is one important point where the accounts differ. Ideal rationality would not help Pascal or the agent in the toxin puzzle. Even for ideally rational agents, a knowledge of their own psychology is important, because sometimes mental states matter to us as states, rather than because of their content.

The most obvious problem with this account is that it seems simply false to say that future-directed self-control necessarily requires managerial control. The next two sections will flesh out this worry, first by discussing exactly what ability is required for managerial control and why this ability might be important for future-directed self-control, and second by looking at Michael Bratman's and Richard Holton's work on self-control. Because both their accounts do not seem to require managerial control, the rest of the chapter will then try to justify why it is nevertheless necessary.

WHAT EXACTLY DOES MANAGERIAL CONTROL INVOLVE?

In order to answer the question of whether managerial control is necessary for future-directed self-control, some more has to be said about what managerial control is. In particular, it is important to point out one ability that I take to be necessary for the intentional control of the mind. This requirement is not discussed in Hieronymi, but as I will explain now, it seems like a necessary condition in order to make sense of managerial control. Managerial control requires a theory that allows one in general to understand that one's mind is to a degree independent of the actual state of the world and can be manipulated to represent states even if they do not obtain. In other words, what is required is a basic understanding of the nature of representation⁵ or the ability to metarepresent.⁶

This basic understanding allows the agent to do two very important things. One, the agent can now understand that it is possible to acquire a false belief if that is useful for the agent, and two, she can understand that it is possible that she will hold

a false belief in the future, even if she has a true belief about the same matter at the moment.

Both abilities seem crucial for future-directed self-control. **As long as an agent cannot understand that a belief which they hold might nevertheless in the future be considered by themselves as false, self-control seems pointless.** Only an understanding of the nature of misrepresentation allows an agent to be aware of the need for self-regulation. A gambler who has a habit of betting large sums of money on her team to win the Champion's League might in a quiet moment know that the team is not really good enough, but whenever she goes past the bookmakers, she might be overcome by the irrational belief that this year things will be different. Now, as long as she does not understand that beliefs can change from true to false, she will not be able to understand that in the moment where she understands that her team has no chance of winning, she might have to take precautions against acquiring the false belief again. She will see no need to do anything as she knows at that moment without any doubt that betting on the team would be the wrong thing to do and has absolutely no intention of doing so. Only if she does not only think about the *evidence for the belief* but about *the belief itself as a mental attitude that can misrepresent the state of the world* will she realize that beliefs are vulnerable to misleading evidence or situational effects. Only then can she understand that she has to be worried about doing the wrong thing in the future, even though she knows what the right thing to do now is and can take steps to avoid doing the wrong thing.

DOES SELF-CONTROL HAVE TO BE MANAGERIAL AND DOES IT NEED METAREPRESENTATION?

If future-directed self-control really is centrally dependent on managerial manipulations of the mind, which in turn requires metarepresentation, then it would seem to make the argument very plausible that the ability to perform these manipulations is necessary for free mental agency. But is this claim really correct?

Future-directed self-control has received a lot of attention in the philosophical literature. Probably the most influential player in this literature is Michael Bratman. According to him, intentions are the mental states designed to facilitate future-directed self-control (e.g., Bratman 1987).

For Bratman, intentions are all about making yourself more rational in the long term. They do that by providing extra reasons to do something that seems not attractive anymore at the moment of temptation. If I form the intention to go to the cinema tonight and on my way there I realize that there is as well an interesting football match on, then my intention to go to the cinema provides me with an extra reason to stick with my original plan. Basically, intentions change the evaluative situation during temptation, and this in turn helps ensure that the right (diachronically consistent) evaluations are made.

Obviously, Bratman's account is extremely rich, and one cannot even begin to do justice to it in a short paragraph, but for our purposes here only two things matter. First, it is not clear that one needs to understand the nature of mental states in order to form an intention and to commit to it, and so Bratman's account seems to be in contrast to the claim defended here that metarepresentation is required for

self-control. Second, on Bratman's account, intentions do their job by influencing the process of rational evaluation. In fact, intentions seem, on this account, in line with what we discussed in the Hieronymi section, very much part of evaluative control and do not normally involve intentional action. Again, this seems to be bad news for the claim defended here that future-directed self-control requires managerial (i.e., intentional) control.

HOLTON ON SELF-CONTROL

One reason to doubt that Bratman's account of intentions is all that there is to self-control can be constructed from Richard Holton's (2009) work on self-control. According to Holton, forming a resolution⁷ in order to prepare against future temptations is not about providing new reasons for action, as Bratman's accounts would have it (these new reasons are on Bratman's account nothing more than the intentions themselves), but simply reduces the ability of the agent to take new reasons into account. It makes the agent in effect less judgment sensitive. Judgment sensitivity here means the ability to reevaluate one's beliefs if the environment provides reasons to do so.

Even more important, according to Holton, this making oneself less judgment sensitive is something that the agent does, and it is clear that what Holton has in mind here is intentional action rather than evaluative control.

This seems very much in the spirit of the account here. On Holton's account, as on the one defended here, following through on one's resolutions requires the agent to be able to break free of her natural evaluative tendencies by means of intentional control of her mind.

Holton's main argument for his account is the phenomenology of battling temptation. If Bratman's account were right, it ought to be the case that battling temptation feels like evaluating two desires and then naturally going with whichever turns out to be stronger. In reality, though, fighting temptation really does seem to involve a constant intentional trying to keep one's mind from reevaluating. One has to be quite revisionist to deny that the phenomenology of fighting temptation does not involve intentional tryings.

As the account defended here also insists on the importance of the intentional control of the mind for self-control, is my account simply a version of Holton's view?

The answer to this question is not at all, because even though, like Holton, this account does emphasize the role of intentional action for self-control, there is one decisive difference. On Holton's account, trying to actively reduce judgment sensitivity does not imply that we form resolutions in order to manipulate our minds.

One might, for example, form the resolution to stop smoking. It seems quite possible to form this resolution without ever thinking about one's mental states. In fact, this seems to be the norm. One will think about the associated health risks and then vow to not smoke another cigarette ever again. This seems like a very fair point to make, but does this not fatally undermine the claim defended here that self-control requires manipulating mental states as states? This is a crucial worry and requires a new section.

REFLECTIVE CONTROL REVISITED

In the last section, we claimed that Holton is a partial ally for the account developed here because, like Holton, this account places a lot of emphasis on the difference between judgment sensitivity and the will. Like Holton, it argues that self-control is about making yourself immune to being too sensitive to circumstances that could change your judgment. However, as already pointed out, there are important differences in the two conceptions as well. The account defended here insists on the ability to metarepresent in order to be able to self-control in the relevant way. Holton not only disagrees with the idea that metarepresentation is crucial for self-control but also fundamentally disagrees with the idea that self-control by means of self-manipulation has anything to do with the will at all.

Holton is very impressed by the work done in psychology, especially by Baumeister (2008), according to whom the will is like a mental muscle. Holton thinks that willpower is something active and that the tying to the most strategies that were discussed here, though clearly useful in self-control, are not exercises of the special mental organ that is the will. This disagreement is especially interesting because, again, it looks as if the exercise of the mental muscle might be a form of mental agency that does not sit easily with the either-or distinction between evaluative control and managerial control. The crucial idea behind Holton's account is that controlling impulses to act by means of deliberating about the right way to act is an effortful activity. This is very much what Baumeister and others have found in many empirical studies, and it is intuitively plausible. But does that not mean that agents can reflectively control their behavior? It seems right that deliberation does contain intentional tryings to remember, to attend to, and so on. Does that not mean that reflective control of the mind exists in contrast to what we claimed in line with Hieronymi? Hieronymi's answer to that kind of scenario was that it is still the case that the actual judgments and intendings will be evaluative and not intentionally controlled. Once she had established the need for evaluative control, she happily admitted that the rest was down to managerial control.

But there is a problem lurking here because, as we have seen with Holton, it does not seem true that these intentional doings are always directed explicitly at mental attitudes. To be sure, there are cases where management is clearly attitude-directed. The agent might tell herself: concentrate or think of the reasons you had for your decision, or try to remember, and so on.

But as described in the Holton scenario, it does not seem to be the case that an agent, when trying to convince herself, will always think about mental states in cases that seem very similar. An agent who wants to stop smoking might not try explicitly to evoke specific mental states. She might, for example, try to focus her attention on reasons for not smoking, but she will not think of these as psychological states. She might think: Why exactly did I want to stop, were there not health reasons?, and so on. What is more, this kind of thing seems ubiquitous.⁸

This shows that there is a very important distinction to be made within what Hieronymi calls managerial control. There are managerial actions that are attitude-directed in the sense that the agent is treating the attitude as an object, and there are managerial actions where the agent is bringing about a change in the

attitude by changing the conditions under which her evaluative processes are taking place, but without an attitude-directed intention. In such cases the agent is obviously not interested in acquiring the specific attitude for the attitude's sake, but does know that certain intentional behaviors have desirable first-order effects.⁹ These effects can be things, like getting the bigger reward, not smoking, and so on. The agent does not have to know that these effects are obtained by means of the acquisition of a mental state.¹⁰

For Hieronymi's purposes, this distinction might not be crucial because she is mainly interested in showing that evaluative control is a specific form of mental agency, and it is certainly true that the intentional part of this—unaware managerial control—would not bring about any change in attitude without the evaluative component.

But the distinction matters here. It matters because it is necessary in order to clarify the claim made in this chapter. **The form of managerial control we are interested in has to be one where Hieronymi's statement that we are intentionally manipulating mental attitudes like ordinary objects is literally true, because only once this is the case will the agent be able to understand that attitudes can be false, can change over time, and so on—and, as we argued earlier, these are necessary elements of self-control in the sense that we are after. So, if self-control can be exercised by means of unaware managerial control, then our claim that the intentional targeting of attitudes is a necessary condition for self-control collapses.**

In addition, once we have introduced this distinction, we obtain as well an explanation of where exactly the difference between Holton and Baumeister and the position defended here lies. Holton and Baumeister argue that willing is intentional and effortful, but the scenarios they describe are clearly not ones where subjects are manipulating their attitudes. As mentioned earlier, bringing about a mental state that will easily allow you to master the self-control task on Holton's model is not about the will at all, because as soon as the manipulation is successful, the characteristics of effort and depletion will vanish. **It seems clear that Holton and Baumeister, in the terminology used here, think of the will mainly as a form of unaware managerial control. In these cases, subjects are trying intentionally to evaluate a first-order proposition in a specific way.** Obviously, as Hieronymi told us, that is impossible; you can intentionally focus or repeat reasons for a specific action, but you cannot intentionally evaluate. Attempting to do it does, however, have an attitude-directed effect. It can help to bring it about that the agent will evaluate the situation differently.¹¹ It does this not by bringing new material to the evaluation but by changing the evaluator (e.g., by making it less interested in new evidence, as in Holton's scenario of self-control).¹²

However, the agent in this scenario is not aware of what it is that they are doing. And that means that such tools are much less flexible and effective than the tools used in managerial control that is aware. If that is the right way to understand such acts of behavior control, then in one sense they are simply less sophisticated versions of real self-control. They achieve their aims by changing an attitude, rather than by providing new evidence for content evaluations. However, they are obviously not intentionally directed at the attitude itself. If that is right, then it seems implausible to exclude the more sophisticated versions from the will and to describe them as mere

tricks. On the other hand, however, there obviously is a major difference between the two mental tools. Obviously, **once you understand what it is that you are doing, the level of control and flexibility is many times higher than before, and that is why the claim is justified that this very general form of theoretical self-awareness is necessary for free mental agency, while the ability to control behavior with the Holton tool on its own is not good enough.**

Finally, one common objection to seeing self-control by manipulation in contrast to behavioral control by sheer effort as part of the will has to be discussed here. This objection states that effortful control is active, while in manipulation the agent gives up control and becomes passive. As soon as the manipulation is successful, the agent cannot go back. On closer examination, this is really quite a weak argument. On the one hand, it is obviously not true that self-manipulations cannot be reversible or conditional, and on the other, control by sheer effort obviously does make the agent more passive, in the sense that she will be less judgment sensitive to good reasons as well as to temptations. **Both forms of control are about introducing a certain element of passivity—that is in fact the very point of them.** How durable that intentionally introduced passivity should be depends obviously on the situation, but it is again true that understanding managerial control as just that will help to optimize strategies. Once we see this, the argument is now turned on its head. Once the agent understands what it is she is doing, it will be much more easily possible to calibrate the right mixture between flexibility and rigidity. Once again, it makes sense to argue that only aware managerial control is good enough for the kind of self-control that is intuitively a necessary condition for free mental agency.¹³

SUMMING UP, FURTHER SUPPORT FOR THIS POSITION, AND OUTLOOK

The chapter started off by accepting Hieronymi's argument that at heart mental agency is not something voluntary but a different form of agency. In line with her account, this form of agency was labeled evaluative control. The chapter agreed as well that the alternative of reflective control, which combines the features of evaluative control and managerial control, does not work because reflective control can be broken up into the two distinct elements of managerial and evaluative control. The chapter disagreed with Hieronymi, however, in arguing that managerial control is at the heart of free mental agency nevertheless. The argument for this was that only managerial control allows the agent to become free from her own evaluations of the world and to begin to construct the kind of psychology that she might think desirable. It was claimed that this form of managerial control requires the ability to metarepresent.

The last couple of sections then clarified why this account is different from other accounts of self-control, especially Holton's view. The crucial point here was that Holton's form of self-control is not really intentional control of the mind at all, because either we understand it as intentional evaluation, which is nothing else than the reflective control shown to be impossible earlier, or it is really just behavioral control that has nonintended side effects in the mind of the agent. So, if Holton's idea is right—namely, that intentional control of the mind is crucial for

self-control—then it was argued that the only way to achieve this coherently is to put forward the account defended here.

The chapter concentrated on presenting the main idea behind the account and discussed some necessary clarifications and obvious objections, but there are many more things that one could add in favor of the account. Here is a loose collection of them.

If the account is right, it would give us an explanation for why free agency is something that we intuitively think only humans can do. As yet there seems to be no clear evidence that any other species other than humans is able to metarepresent—and metarepresentation is a necessary condition for aware managerial control.

The account also has a story to tell about what the function of making people responsible for their mental states might be. It is true, we do tell our criminals that they should understand the error of their ways, but this has always been a big ask. Philosophers and most ordinary people struggle to find a fault-proof rational way of arguing that doing or being good is also being rational. So why do we think that criminals should be able to do it? However, what has a chance of succeeding is an exercise in managerial attitude acquisition, which helps the potential reoffender to overcome her reasoning, which had seemed previously to make the offense rational for her. Aware managerial control is something that we can teach people to do.

Interestingly, it is a sociological fact that the genre of books that is supposed to help people to exercise self-control is already one of the biggest sellers on the market.¹⁴ Many people look down on the self-help genre, but many more swear by it. This is not that surprising actually, because the advice given in these books maps quite nicely on the findings in serious cognitive science labs like Gollwitzer's, that is, it works by helping people to exercise managerial control.

Finally, the account has some interesting consequences. It was argued that self-blindness is not a problem for the account, as long as the agent understands the nature of representation, but obviously, new knowledge in the sciences does allow us to be far more effective in this form of mind and self-creation. This already has led to enormous changes in the way we manipulate our minds, for example, psychoactive pharmacy or cognitive behavioral therapy. In this respect, the account is in the end about breaking down the boundary between forms of self-control that are supposed to be internal to the agent, like the mental muscle phenomena that Holton and Baumeister describe, and the use of external scaffolding that humans use to aid their self-control. This chapter shows that both forms use the same mechanism and that, if anything, the aware use of external scaffolding is a more sophisticated form of the will than the simple straining of the supposed mental muscle.¹⁵

NOTES

1. Managerial and manipulative control differ only insofar as in managerial control the agent influences the environment in such a way that a normal evaluative process brings about the desired result, whereas in manipulative control the bringing about of the judgment does not depend on a normal functioning of the evaluative machinery. From here on, I will label both forms managerial.
2. They were not looking under the hood, in Moran's apt phrase (Moran 2001).

3. Hieronymi argues that actually two acts of evaluative control are required. The second act consists in the evaluation that the intentionally brought about circumstances cause. This seems plausible enough for managerial control (see distinction in note 2), but there is an ambiguity here for manipulative control, where the bringing about of the attitude does not seem to necessarily require evaluative control at this stage. Imagine, e.g., that the belief is surgically implanted. It is not clear that in such a scenario there has to be initially a second act of evaluation.
4. I owe this term to Andreas Paraskevaides.
5. I.e., the ability to pass the false belief task. See Perner 1993.
6. This is obviously most fitting for belief, but arguably it works for intention as well. **Intentions always contain a judgment about what is the best thing to do**, and obviously this judgment can go wrong. Understanding this is crucial if one wants to implant an intention for an intention's sake, rather than for the sake of its content.
7. Holton's term for an intention formed in order to ensure that one sticks to one's plans in the face of temptation.
8. Even though the ability to intentionally guide deliberation is no mean feat. In fact, **there is good reason to think that this controlled deliberation is what gives humans a form of thought regulation that other animals do not have. However, it is still true that this form of controlled thinking does not require metarepresentation**. I discuss the role of intentionally controlled deliberation in detail in a forthcoming paper (Vierkant 2012).
9. Unaware managerial self-control is itself a very broad term. In one sense, it includes most intentional behaviors that there are, because most intentional behaviors have consequences for the mental states of the agents. However, there are some forms of unaware managerial control that are far more sophisticated and effective in controlling minds as a side effect than others. There is no room to elaborate on the various forms of unaware managerial control here, but I do develop this point in (Vierkant & Paraskevaides 2012)
10. There is no space here to expand on this distinction, but it would seem to be a worthwhile undertaking. There has been a very lively debate on which mental actions can be performed intentionally (e.g., Strawson 2003; Pettit 2007). In most of these debates, however, it is presumed that we know what it is that we are doing when we manage our attitudes or bring it about that we have better conditions for our evaluations. Obviously, this knowledge is theoretically available in humans, but it is far less clear whether it plays a role in many of these managerial acts. In fact, it seems not that unlikely that the intuition of reflective control is created exactly by the fact that very many managerial acts are not understood as such by the agent.
11. This very short sketch of nonaware managerial control only scratches the surface of a huge fascinating field of cognitive science. There are probably many stages on the way to making an animal aware of its own mentality.
12. **There is an interesting link here to the discussion on metacognition in animals**. See, e.g., Smith et al. 2003.
13. How knowledge of our psychology could enable us to optimize self-control can be seen in the chapter by Hall and Johansson, this volume.
14. For an interesting analysis of the self-help genre as the contemporary form of talking about the will, see Maasen et al. 2008.
15. For a way of fleshing out this idea of how we could use external scaffolding to support the will, see Hall and Johansson, this volume.

REFERENCES

- Baumeister, R. (2008). Free will, consciousness and cultural animals. In J. Baer (ed.), *Are we free*, 65–85. New York: Oxford University Press.
- Bratman, M. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.
- Faude-Koivisto, T., Würz, D., and Gollwitzer, P. M. (2009). Implementation intentions: The mental representations and cognitive procedures of IF-THEN planning. In W. Klein and K. Markman (eds.), *The handbook of imagination and mental simulation*, 69–86. New York: Guilford Press.
- Hieronymi, P. (2009). Two kinds of mental agency. In M. S. Lucy O'Brien (ed.), *Mental actions*, 138–162. Oxford: Oxford University Press.
- Holton, R. (2009). *Willing, wanting, waiting*. New York: Oxford University Press.
- Maasen, S. Sutter, B., and Duttweiler, S. (2008). Wille und Gesellschaft oder ist der Wille ein soziales Phaenomen. In T. Vierkant (ed.), *Willenshandlungen*, 136–169. Frankfurt: Suhrkamp.
- McGeer, V. (2007). The moral development of first-person authority. *European Journal of Philosophy* 16: 81–108.
- McHugh, C. (2011). Judging as a non-voluntary action. *Philosophical Studies* 152: 245–269.
- Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton, NJ: Princeton University Press.
- Perner, J. (1993). *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Pettit, P. (2007). Neuroscience and agent control. In D. Ross (ed.), *Distributed cognition and the will*, 77–91. Cambridge MA, MIT Press.
- Smith, J. D., Shields, W. E., and Washburn, D. (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behavioural and Brain Sciences* 26: 317–373.
- Strawson, G. (2003). Mental ballistics or the involuntariness of spontaneity. *Proceedings of the Aristotelian Society* 103:227–257.
- Vierkant, T. (2012). What metarepresentation is for. In J. Brandl, J. Perner, and J. Proust (eds.), *Foundations of metacognition*. Oxford: Oxford University Press.
- Vierkant, T., and Paraskevaides, A. (2012). How social is our understanding of minds? In F. Paglieri and C. Castelfranchi (eds.), *Consciousness in interaction*, 105–124. Amsterdam: John Benjamins.
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.