



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Mental Muscles and the Extended Will

Citation for published version:

Vierkant, T 2014, 'Mental Muscles and the Extended Will' *Topoi*, vol. 33, no. 1, pp. 57-65. DOI: 10.1007/s11245-013-9188-5

Digital Object Identifier (DOI):

[10.1007/s11245-013-9188-5](https://doi.org/10.1007/s11245-013-9188-5)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Topoi

Publisher Rights Statement:

© Vierkant, Tillmann / Mental Muscles and the Extended Will.
In: *Topoi*, Vol. 33, No. 1, 04.2014, p. 57-65. The final publication is available at Springer via <http://dx.doi.org/10.1007/s11245-013-9188-5>

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Mental muscles and the extended will

Abstract:

In the wake of Clark & Chalmers famous argument for extended cognition some people have argued that will power equally can extend into the environment (e.g. Heath & Anderson, 2010). In a recent paper Fabio Paglieri (2012) provides an interesting argument to the effect that there might well be extended self control, but that will power does not lend itself to extension. This paper argues that Paglieri is right in claiming that previous attempts to extend the will are flawed. It then provides an argument for extending the will that does not fall foul of Paglieri's argument and actually provides us with an even stronger case for extension than the one that Clark & Chalmers provide for cognition.

Keywords:

Willpower, self control, managerial control, Holton, extended cognition, mental muscle

The ability to stop ourselves from being tempted is one of the things that is central to our agency. Humans can, some better than others, not give in to temptation if it presents itself, but stay the course and do the right thing. What exactly this ability consists in has been an age-old discussion. Nevertheless, what most people would want to agree on is that using external props that simply prevent one from giving in to temptation are an altogether different way of controlling the mind to real willpower. If you don't want to eat the cake in the fridge until tomorrow, you can lock the fridge and give the key to a friend, who will only come back tomorrow, but this is not the same as leaving your fridge unlocked and simply controlling your desires with your will. This paper sets out to argue against this intuition and contends that, in the way that matters, the two at first glance very different scenarios are about the same kind of control. Once this point is established, the paper will then argue that for the will we have even stronger grounds than for cognition (Clark & Chalmers, 1998) to believe that it is extended.¹

Before the main point can be tackled, however, some clarifications are in order. The claim defended in this paper is quite different from one that may seem superficially quite similar and which is defended by Anderson and Heath in their paper "Procrastination and the Extended Will" (2010). They argue that we underestimate the importance of self control using external props in our understanding of how humans manage to overcome temptation. They do not, however, deny that using props is different to using sheer will power. So, on their view, the will is only extended because, on a suitably coarse-grained level, self control by external props gets the job of overcoming temptation done just as well as control by internal will power. This is good enough to make it real will power, even if using such forms of self control is clearly operating with different *mechanisms*, because the two are *functionally* equivalent.

Nevertheless, this argument seems unsatisfactory because, as Paglieri (2012) points out, the disanalogies between using willpower and external props seem crucially important when describing their respective functional roles in our cognitive economies.

¹ Clark and Chalmers famously argue that cognitive processes extend into the environment. In their Otto example, they argue that information stored in a notebook under the right conditions can be a cognitive vehicle in exactly the same way as a biological memory trace in the brain. This is the case whenever the vehicle inside the brain (biological memory trace) plays the same functional role as the external vehicle (information stored in notebook).

Real will power is like a mental muscle. It gives one full control over a behaviour, and its resources are strictly limited, though trainable. Straining the muscle also requires effort. Self control by external props is in contrast effortless. Odysseus and the sirens as the archetype of this kind of control is a very good example here. All Odysseus has to do to listen to the tempting song of the sirens without risking that he will jump ship and die is to tie himself to the mast. Once the knot is tied, nothing needs to be done by the agent: But the downside of this is obviously that the agent has no control any more. One cannot undo the knot. Odysseus has to hope that the ship won't sink while he is tied, because he will not be able to change the tied state he is in by himself. On the plus side, self control by external props is limitless in the amount of temptation it can withstand (again Odysseus is a case in point), while on the minus side, such self control cannot be trained. It is, as it were, useless practising being tied up. Most importantly, as Paglieri points out, willpower does seem to be a *means* by which to achieve the end of self control. Hence, using external props and using will power are two very different means of achieving the same end. If this very plausible analysis is correct, then it seems to mean the end of all the extended will talk. Self control might well be something that we can extend, but willpower as an internal means simply does not have any functional equivalent in other forms of achieving self control.

But if Heath and Anderson (2010) and their weaker form of merely functional equivalence between will power and control by external props is already misleading, how can it be at all plausible that the two are not only functionally equivalent, but employ the same mechanism?

In order to make the point that this seemingly very implausible claim is actually how we should think about willpower, it is now useful to introduce a strange character, who has a very specific problem with controlling some of his attitudes. He should help to sharpen our intuitions about what is important in self control. Here comes Karl...

Karl's problem

Karl has a nasty character streak. He just loves watching others suffer. This is not to say that Karl is all in all a nasty person. Apart from this character flaw, he is actually a quite moral person. He is a convinced Kantian and tries to organise his life in line with the categorical imperative. Karl has had fantasies about watching others when they are in pain since he was a little boy. During his early years there was almost nothing which gave him more pleasure than kicking cats and boiling frogs. When he grew up, he realised that these actions are morally

reprehensible – and, being a living argument for the power of moral judgment, this allowed him to refrain in adulthood from doing anything nasty to his fellow creatures.

Nevertheless, Karl still enjoyed watching suffering very much, but he thought that it was harmless, because he only ever satisfied this desire by watching gory movies. Karl was a happy man until one day during watching the news, he discovered that he actually really enjoyed seeing the pictures from the torture scenes in Abu Graib. It gave him great satisfaction to watch these people being humiliated and tortured in all kinds of horrible ways. He wanted to be disgusted by the scenes before his eyes, because he was very much aware that they were the sign of the utmost moral depravity, but to his horror he found that he actually quite enjoyed flicking through the pictures. Karl did not want to feel this pleasure, but he just could not help himself. It should be noted as well that Karl was very relieved to know that his love of torture did not extend to those he loved. Seeing people he knew suffer was for him as painful as for other people. It was just the anonymous faceless cruelty he enjoyed. Karl often told himself that this fact showed that he was not really cruel, because he did not enjoy suffering as soon as it became real for him. Nevertheless, he started to wonder whether he was actually responsible for what he felt. He was well aware that people at best reacted with a disbelieving grimace if he told them, and more often, were literally disgusted by him. Karl felt that this was very unfair, because he never did anything bad and, just as important, he could not do anything about these feelings anyway. He felt that you can only be responsible for something if you have agentive control over it.² But no matter how hard Karl tried to suppress the enjoyment, however much he said to himself sincerely: "this is bad", the feelings did not go away. But Karl argued that he had no control over his feelings, and that his feelings were not the result of a bad or in any way morally questionable lifestyle. On the contrary, he had used the power of moral judgment to transform himself from a nasty little boy, who did many bad things, into an upright citizen who never did anything bad, and who just had this small eccentricity in his character.³

Poor Karl

Karl leaves us with conflicting intuitions. On the one hand, we want to be morally outraged by his nasty enjoyment of suffering, but on the other hand, Karl seems to have a point when

² A thought that I want to agree with.

³ See Angela Smith (2005) for a similar argument on why tracking (or character formation arguments) does not work in cases like Karl's.

he claims that he can't do much about his feelings. We all know feelings that are unruly and don't seem to go away, even if we sincerely will them to do so.

But what drives the intuition that Karl is not responsible for his enjoyment? At first glance, it looks as if the answer here is as follows: As Karl cannot use willpower to make the attitude go away, he does not have direct control over it. He cannot just perform a simple deliberation about whether it is adequate to enjoy the suffering of others, and the outcome of that deliberation will determine which attitude he holds. As he has no control over it we should not blame him for holding it, because it seems unfair to blame someone for something they can't do anything about.

But is it really true that Karl has no control over his attitude at all? That does not seem right. So far, we only discussed willpower as a possible means to get rid of the unwanted attitude, but there might be other ways. Karl could e.g. visit a cognitive therapist. Let's suppose for our purposes that Karl could indeed get rid of his attitude by doing therapy. Would that change the intuition that he is not responsible for it? To a degree, it certainly would. If it was reasonable to expect of Karl that he could know that such therapies exist, then it no longer seems true that he could not have the means to get rid of his attitude.

However, even in this case there still seems to be a case to be made for arguing that Karl is not responsible for his enjoyment *in the same way* as he would be if he could simply repress his enjoyment. If he could will the feeling away then, so it seems, he would have direct agentive control over it, but going to therapy is not like that. Karl might know that going to therapy will cure him, but it wouldn't be his actions that would cure him, but his actions would bring it about that some other causal force (the therapist) will act on his mind and thereby change his attitude.⁴

One explanation one could give for the intuition that this difference between direct and indirect attitude control matters can be constructed from the work of Richard Moran. Moran argues that the normal way of changing an attitude is by means of deliberation, and that attitudes that we change by means of external manipulation never become fully our own

⁴ This is true, even though the therapist will presumably ask Karl to help in his therapy, so that Karl is not fully passive. Karl will learn to do many different things, which are not aimed at directly controlling the attitude, but at doing something else, which will bring it about eventually that the attitude disappears. So even in this case it is not Karl's direct control of his attitude that brings about the change, but the things he does provide an environment which then in turn is causally efficacious in bringing about the change. Thanks to Jacq Cottrell and an audience member in Rome for pressing this worry.

(Moran 2002, p.117). Thus in such cases we remain in an important sense alienated from such attitudes, because they are not connected to our rational networks in the way e.g. beliefs normally are. We will have more to think about the difference between the different ways in which we can control our attitudes later on when we discuss Pamela Hieronymi's helpful work in the area. At this stage, it is enough to note that not only is there an intuitive difference between the different forms of attitude control, but that the intuition that direct deliberative agential control matters also has very influential philosophical back up.

Karl is also useful in helping us to understand what the supposed difference between direct control by willpower and indirect control by means of the environment comes down to. One might be tempted to contend that the difference between Odysseus's strategy of tying himself to the mast in order to not jump ship when the sirens start singing and a case of real willpower is that Odysseus does not control his mind at all. All he does is make sure that no harm will result if he changes his mind from what he thinks is the right thing to do now. Karl shows us that this is not the only difference that matters. Going to therapy obviously would change his mind, rather than just manipulate the environment, but that is not good enough to meet the conditions for direct agential control. What seems to matter for willpower versus control by props is not only that in the one case, the mind is controlled and in the other the environment; what matters as well is the way in which the mind is controlled.

After establishing what explains the intuition that Karl is not responsible for his attitude, we will now deal with the question whether this intuition is actually justified. The paper will try to establish that it is not. The main argument for this is going to be the idea that despite the strong appearance to the contrary willpower is actually not direct control of the mind either. If that is plausible, then either merely adding the ability to control his attitude by will power would still not be enough to make Karl responsible for it, because he still would not have direct control over it, which seems very implausible. The only alternative is that direct control is not necessarily required for responsibility, because simply adding willpower is good enough to make him responsible, but willpower is not direct control. If that should turn out to be true, however, then there is no principled difference between using willpower and other forms of indirect control of the mind – and in that case, the intuition that Karl is not to blame for his attitude because he lacks the ability to change it by willpower is shown to be false, because Karl can change it by other indirect means, and there is no obvious reason why these means are not sufficient for responsibility, if willpower is good enough.

With this point established, the obvious question to ask next is what kind of control is willpower actually, and does it qualify as direct control of the mind?

What is willpower?

Richard Holton (chapter 6, 2009) has recently given a very interesting and convincing account of how to think of willpower. According to Holton, willpower requires the ability to curb the desire to constantly evaluate and re-evaluate situations rationally. It requires the ability to not think about short term benefits in situations where short term benefits would be evaluated as more significant than long term benefits, if one considered them fully.

Holton's account is very rich, and we will not be able to fully discuss the details of the position, but here is a short version of what I take to be the main argument: The main claim Holton wants to defend is that willpower is the intentional maintenance of resolutions (p.112, 2009). Resolutions are a specific class of intentions formed specifically to withstand temptations at a later point.⁵ The crux of Holton's claim is the idea that resolutions need to be maintained actively. This is crucial, because it sets Holton's account from the standard alternative accounts that are available. Holton names these alternatives "augmented Humean accounts" and points out that in such accounts there does not seem to be any need for intentional activity in the exercise of willpower. On such an augmented Humean account, resolutions are states that help overcome temptation, because they provide the agent with an additional reason not to give in to temptation. Resolutions on this account are like promises, and if the agent is motivated not to break promises they have given to themselves, then the presence of a resolution represents an extra reason not to give in to temptation (because doing so would break the promise).

Holton accepts that the augmented Humean account provides a workable picture of how resolutions might function, but thinks that it is phenomenologically implausible and not supported by empirical evidence.

The augmented Humean account does still seem to be fairly what has been described as "hydraulic"⁶. If this account were correct, it would seem to indicate that temptation should not feel like a struggle (p.118, 2009), but like a simple calculation of whether or not the value

⁵ This difference between ordinary intentions and resolutions is crucial for Holton's disagreement with Bratman's account of resolution (see chapter 7, 2009). Holton thinks that resolutions work like intentions on a Bratman-like two tier model, but Bratman thinks that his model for intentions does not work for resolutions.

⁶ See e.g. Velleman (1992).

ascribed to the resolution swings the overall calculus towards staying the course, or whether it does not. Either way, the outcome should be fairly automatic and should not involve any significant mental effort. But this just seems phenomenologically inadequate. Willpower seems to be all about trying, and, as the fascinating work by Baumeister et. al. (2008) shows, it does seem to be like straining a muscle, which can eventually become depleted (what they refer to as “ego depletion”). So Holton’s account does what augmented Humean account has failed to do: It puts the intentional agent back into the picture.

In order to make this difference clearer, let’s have a look at an example: Suppose a woman wants to drive to the pub in the evening. She knows in advance that alcohol will be consumed, but forms the resolution that she will drink no more than the legal limit. Once she is close to the legal limit, she sees new (and morally not very helpful) reasons that speak in favour of another drink: she does not have to drive very far to get home, she could drive on quiet roads, etc. In this situation, on the augmented Humean account, the knowledge that she has formed the resolution, and a commitment to it, is what potentially helps the woman not to drink more. She should simply consider what is more important: having another drink, or all the reasons that speak against having a drink, plus the fact that she is committed to her resolution. Now crucially, it seems that whatever the outcome of that process is, that is what she is going to do. There is no need for a mental struggle.

Contrast this with the Holton proposal. On Holton’s account, what the woman does is to intentionally disallow re-evaluation. She simply doesn’t allow her mind to perform the relevant calculation. As long as she succeeds in doing that, she will not order a new drink. There are a couple of advantages to this account: on the one hand, it explains the phenomenology of the struggle, and on the other, it explains why willpower so often fails to achieve its goal in the long run. It seems less plausible to say, as the augmented Humean account would have to, that there are constantly new reasons that force the agent to make a new calculation. It seems much more straightforward to argue that, because an evening is long, at some point the energy required to prevent the re-evaluation is depleted and the agent just gives in and allows herself to newly appraise the situation, with foreseeable and in the long run, very unwanted consequences.

Assuming that Holton is right about willpower being at bottom an intentional activity, which seems plausible, why is that important for Karl? It still seems the case that Karl’s enjoyment of torture and death is quite innocent, because he can strain his mental muscles as much as he

likes, but the enjoyment does not go away. In order to move the argument forward, we now have to have a look at the relationship between intentional and other mental actions.

Evaluative and managerial control

To discuss the question of what role intentional activity in general plays in mental agency, it is useful to bring in another player. Pamela Hieronymie (2009) works within the Moran (2002) framework, but has very much helped to clarify his ideas on mental agency. Hieronymie argues that there are two types of mental agency, which she calls evaluative and managerial control. Managerial control is the locus of all intentional mental agency, according to Hieronymie, but in order to understand managerial control it is first necessary to understand its non-intentional counterpart, evaluative control.

Evaluative control is the control an agent exercises when they form a judgment or an intention in the normal way. Judgments and intendings are at the heart of what it is to be a rational agent, but they seem not to fit the criteria for an intentional action. The two crucial differences are: First, it is not up to the agent to judge or intend at will; and second, there does not seem to be a reflective difference between the agent and what the agent is acting on. When we intentionally throw a ball, we have an intention to act on an object (the ball) and we then manipulate the ball according to our intention, but when we form a judgment, we don't have an intention to manipulate our mind in such a way that we judge what we intend to judge. Rather, we come to judge the way we do, because the evidence we see before us compels us rationally to judge accordingly. Evaluative control is not about manipulating your mind intentionally, because you have judged that it would be a good thing to do so, but about the rational mind coming to hold an attitude by using its rational resources. Evaluative control is nothing more or less than control by means of the rational process itself.

The details of Hieronymie's proposal are controversial, but in the field of mental agency, it is commonly accepted that there is a distinctive non intentional element in the mental episodes of judging and intention formation.⁷ Understanding this in turn clarifies the relationship

⁷ The whole debate in the mental agency literature really seems to be about the relative importance of intentional shepherding versus non intentional evaluative components. Moran (2001) and Strawson (2003) are here united in claiming that shepherding is not very central, Mele (2009) and Hieronymie (2009) are somewhere in the middle, while McGeer (2007) and myself (2012, 2013) hold positions that emphasize the intentional part of the mental (albeit for very different reasons).

between the two forms of mental agency. Most importantly, Hieronymi makes it clear that all intentional control of the mind is always dependent on prior acts of evaluative control. The reason for this is simple: Intentional action is normally about making the world fit an intention. But how is the intention itself formed? Most of the time at least, it seems very implausible that the forming of the intention is itself an intentional action. So either we are most of the time passive in intention formation, or the intention formation itself is agentive in a different sense – and that sense is presumably evaluative control.

But even though Hieronymi argues that evaluative control is the most fundamental form of mental agency she is not negative about the role of managerial control for our mental lives. Importantly, she makes clear that managerial control exists in a very wide spectrum of cases. When we normally influence our mental lives intentionally, we do this by bringing about conditions for our evaluative processes which will support successful evaluation. For example, when we write down notes to remind us how to solve a logic puzzle, we do something intentionally that will influence which mental state we will be in when we attempt to solve the puzzle again a couple of days later. However, we don't do this, as it were, by simply imprinting the solution in our mind, but by providing props for the evaluative processes to work properly.⁸ This kind of mental shepherding is very widespread: All forms of reminding ourselves, focussing, attending to detail, providing notes or other external props for the evaluative processes themselves, can be described in this way.

Holton and Hieronymi

The distinction between evaluative control and managerial control can now help us to gain new insights and understand better what Holton has to say about willpower.

The first thing to remember is that willpower on Holton's picture requires the forming of special intention-like states (resolutions). Forming these states is obviously not an intentional activity, so, as in standard cases of intention formation, Hieronymi's point – that all mental agency ultimately has an evaluative component – applies. However, this does not mean that the intentional component of willpower is unimportant or somehow defective – it is simply one form of managerial control. As in other forms of managerial control, the agent uses his intentional capacities to help his evaluative processes to function better, which in this case

⁸ Hieronymi contends that this is the case for managerial control. She does allow for cases of simple imprinting as well, but uses the special term "manipulative control" for this form of attitude control.

means that she tries to disable them for the time when they are bound to produce the wrong conclusions.

Obviously, what is slightly different here from the logic puzzle case is that the intentional action does not lead to more or better evaluative processing, but to less. Nevertheless, this seems fine, because the agent judges (i.e. exercises evaluative control) at the time she forms the resolution that the kind of evaluative processing that is likely to happen under tempting circumstances is not in her long-term interest and should therefore be avoided.

There is one important caveat to claiming that willpower on the Holton account clearly is a form of managerial control: Holton himself does not use the term managerial control. Can we assume that managerial control really is what Holton has in mind? There is at least one reason why we might have to be careful here. Holton suggests that the intentional maintenance of resolutions works by means of what he calls ‘rehearsal’ (p.123, 2009). When faced with temptation, the agent does not simply avert her attention, but neither does she reopen a deliberative process about whether it would be worth giving in to it. Instead, she repeats to herself the reasons she has for sticking to her resolution and thereby blocks a shift towards open deliberation. Holton argues that doing so is not arational (p147, 2009), because rehearsal does not mean that our rational abilities are fully disabled (we could use them if new information urgently required us to do so) and there is ongoing rational activity (e.g. about implementing the resolution) during rehearsal. Does this not mean that Holton thinks ‘rehearsal’ *is* rational and therefore different from managerial control, where the intentional part of the action is *not* supposed to be rational?

On one reading, this could indeed be the case, e.g. if we understand the intentional rehearsal itself as a rational (evaluative) activity.⁹ If this were the case, Holton would believe in something that Hieronymi calls “reflective control” – but this is in itself problematic, because it is unclear whether reflective control is possible, or even a coherent concept. The issue is complicated and a fuller discussion is beyond the scope of the paper (but see Hieronymi 2008), but the main argument is easy to see. The very point of the distinction between evaluative and managerial control is that evaluations are *never* intentional. We can do many things in order to prepare evaluations intentionally (concentrating, attending, rehearsing, etc),

⁹ One reason to think that this might be the reading that Holton has in mind, in spite of my claiming that it is an implausible reading, is the fact that it seems to be what Baumeister (2008) thinks is correct, who clearly has been very influential for Holton’s position.

but it seems that the evaluative part of any rational act is always non-intentional, and it seems that we would have to subscribe to a very implausible fully fledged doxastic voluntarism to be able to accept that rehearsals can be intentional and evaluative at the same time. For this reason, this reading of Holton's claim is one that does seem to come at a very high price.

On another reading, all that Holton says here is that the intentional activity does not fully disable the rational abilities of the agent (unlike e.g. the taking of a sleeping drug to resist temptation would). On this reading, it seems justified to describe Holton's rehearsal as a form of managerial control. Obviously, managerial control does not disable evaluative processes, but, because of a prior evaluation, manages them without being evaluative itself. Thus, on this reading, the agent performing rehearsals is clearly still a rational agent, but the rationality of the agent is influenced by an arational process (the rehearsal). Nonetheless, this arational process itself is the result of an evaluative act (the forming of the resolution) earlier on.

It is not clear which of the two readings Holton actually endorses, but as it seems that the only one that works is the latter interpretation, I will take this to be the correct understanding of Holton's work in the following sections.

Back to Karl

The result that willpower is simply a form of managerial control now makes it possible as well to see the Karl case in a new light. Karl felt uneasy about going to CBT, because he was worried that he would somehow change his attitude in a way that is second best. However, Karl is in full agreement with the common intuition that using willpower to control the attitude would be legitimate. As Holton's account of willpower shows, using willpower is nothing else than using intentional mental shepherding strategies (rehearsal) to prevent oneself from evaluating at all, if the agent knows in advance that in certain situations the result of the evaluation is bound to go the wrong way. From here, it is only a small step to the Karl case, where the intentional action of going to a CBT therapist will (hopefully) lead to the disappearance of an evaluative process that is unwanted by the agent. Admittedly, this does entail what Moran refers to as the passive installation of an attitude by brute intentional force, but it seems difficult to detect a crucial difference here to the processes described in Holton's

willpower case: The agent simply intentionally disables unwanted evaluative processes. This leaves us with the option of either claiming that willpower alienates us from ourselves, or to accept that not all intentional control of attitudes is alienating. The latter option seems so much more plausible than the former that the discussion below will only take this second option into account.

So far, it has been established that going to CBT and pure willpower (at least on the plausible reading of the Holton account) can both be thought of as forms of managerial control. At least in one respect then, willpower and other forms of intentional control of the mind should be regarded as the same. Nevertheless, this is clearly not the same as showing that they are the same in all respects. The rest of the paper will try to deal with likely worries about pushing the similarity between the two attitude control strategies too far.

Possible other relevant differences between willpower and other indirect intentional control of the mind

There are countless different forms of managerial control, ranging from willpower to far more indirect strategies. Obviously, these strategies differ in many respects, so it will not be possible to demonstrate that willpower is exactly like all other forms of managerial control, because there clearly is diversity. What can be done, however, is to point out that some of the dimensions that traditionally have been used to emphasise the difference are actually less dramatic than is commonly assumed.

Most important in this regard is the intuition that willpower is direct intentional control of the mind, while using external tools is somehow indirect and manipulative. This intuition is very strong and it is also simply false. It is false because, as the Hieronymi discussion shows, there is no direct intentional control of the mind – at least, not if direct intentional control is supposed to be a mental act that does the job of evaluating and managing at the same time. Even in the intuitively purest cases of willpower, the intentional part of the mental action always follows an independent evaluative act. The intentional part of the action is not part of the evaluating, but makes sure that the right environment is created before a new evaluative act can take place. In a way, this simply repeats the argument presented in the last section. It is important to repeat this argument here though, because if willpower is an *indirect* control of

the mind, just like other forms of managerial control, and *not* a combination of rational and intentional control, then it is far less obvious that it should have a special status.

Conversely, if we lose sight of this crucial element and accept that willpower can be direct intentional and rational control of the mind at the same time, i.e. it can be managerial and evaluative, then immediately, there is a very powerful reason to think of willpower and self control as fundamentally different mechanisms. In fact, it is very likely that precisely because we do not ordinarily carefully disentangle evaluative and intentional aspects of our mental agency, that the specialness of willpower is so incredibly intuitive.

Secondly, willpower itself is far less of a homogenous category than one might think. Clearly, locking the fridge and flexing the mental muscle while staring at the cake are two extremes, but in many cases it is far more difficult to know what exactly is willpower and what is a different form of intentional control. If the agent puts the TV on to distract herself, or simply does not look at the cake, is that willpower or another form of control? What if the agent looks at the cake, but focuses only on one particular bit that looks rather like disgusting slime? What if she does some difficult algebra instead, etc.?

In everyday contexts, we very often would assume that someone who has used these strategies is showing willpower, and this plausibly is a large part of the intuition that willpower is such a fundamental human ability (a point made forcefully by Heath & Anderson, 2010), but these examples are nevertheless clearly exercises in indirect intentional control of the mind.

Philosophers like Paglieri (2012) deny therefore that employing these strategies is about real willpower and place distraction and fridge-locking firmly in the “mental tricks” category. But the question remains: Are there really means of self control that are fundamentally different from these cases?

It is here that we have to come back to the Paglieri’s point about the nature of willpower again. He argued that willpower is a means to an end, whereas self control is the end. If we accept this analysis, it seems easy to explain the difference between distraction cases and real willpower. In distraction cases, the agent achieves her end by performing an action which then robs her of the ability to control her behaviour for as long as the distraction lasts, whereas in real willpower, the self control is achieved by a constant exercise of willpower, so that the agent always stays in control. Both methods achieve self control, but one does it by

temporarily disabling agentive self control, whereas the other achieves self control by exercising it.

Using Hieronymi, we can now see what the problem with that analysis is. Paglieri tells us that willpower is a means, but he does not tell us how it works. If we now look at Holton's analysis of how willpower works, then it turns out that willpower itself has, as it were, a tying part and a tied part. Willpower achieves self control by performing an intentional action (the tying) that leads to a cognitive environment that prevents re-evaluation. The intentional action, as it were, ensures that the evaluative agent is tied. But this means the evaluative agent is not free to re-decide. She is entirely passive as long as the ties (the manipulated cognitive environment) hold. Just as in the Odysseus case, the cognitive ties prevent the evaluative agent from reacting flexibly to the new input.

The obvious objection to this argument could be that even here willpower consists only in the intentional action, but not in the cognitive environment that is its consequence, and which forms the "tie" for the evaluative agent. But, if that were the case, then what is special about this intentional action in contrast to any other intentional action that produces ties, as in the case of distraction? It seems that the difference between the two actions is not that they are different forms of agency, but that they have different effects: Distraction in the one case and reduced sensitivity to alternative reasons in the other. But if this is true, the difference between the two is a difference in the nature of the ties, but not the tying action itself.

Finally, this leads us directly to the third intuitive crucial difference between willpower and indirect forms of intentional control of the mind: flexibility. Once you have locked the fridge and given your friend the key, it is locked, even if your starving relatives come for a surprise visit that evening. On the other hand, if you simply control yourself with willpower, then you can now simply re-evaluate the new situation and feed them the cake, and if politeness requires it, even have a piece yourself.

However, despite its surface plausibility, this argument does not stand up to any scrutiny. Even locking the fridge need not be irreversible. One can e.g. call the friend who has the key and make them aware of the situation, and if distracting by television is the preferred strategy, then the mere fact that your relatives ring the bell should be enough to snap you out of your distraction.

Still one might think that even if it is true that willpower, like other intentional control of the mind, works by introducing ties and making the (evaluative) agent less flexible, that willpower nonetheless remains the most flexible form of self-binding. But even this is not true. Distraction can be used in exactly the same way as a short term tie, with the added advantage that it is possible to calibrate much more exactly how long the ties are supposed to last before they disintegrate. One can e.g. pick a piece of distracting music with the exact length required.

The extended will

Karl is responsible for his enjoyment, because he can do something about it. Even if it is true that simply willing it to go away does not help, he can still use other methods that will do the trick. Importantly, Karl should not worry that using these methods is somehow worse for him – at least if he buys into the idea that willpower is all about the intentional control of the mind, as Holton argues.

By introducing Hieronymi, it was possible to keep some of Moran's central insights. Karl can accept that evaluative control is central for mental agency, but this is not to say that managerial control is unimportant. This allows for the intentional control of the environment for evaluative acts.

Finally, if the argument presented here is right, then we are now in a position to justify the claim that the will is extended far more clearly than cognition. Clark & Chalmers argue that cognition loops into the environment, in cases where the environment plays a role that we would not hesitate to call cognitive, if it happened within the skull. One thing one might worry about in the cognition case is that the process of internal cognition is very likely to be quite different to the process of using an intentional action in order to achieve the cognitive goal, or that the level at which the argument works has to be so coarse-grained as to allow very implausible extensions.¹⁰

In the case of the will, this worry disappears if the arguments presented in this paper hold. In this case, exercising the will is always an intentional action, which helps to manage evaluative

¹⁰ In fact, countless people do worry exactly about this or similar points. See prominently e.g., Rupert (2004), Adam & Aizawa (2008), Sprevak (2009).

processes. The will is not about these evaluative processes themselves, so it would not matter here, if it turned out that cognition (which is what evaluative processes presumably are) only happens inside the skull. Instead, the will is about acting intentionally as a consequence of the evaluative acts and the relevant intentional action may or may not involve the manipulation of objects outside the skull. In contrast to the case of cognition, the process involved here is always the same. It is, as Moran has taught us, not the operation of the rational mind as in deliberation, but the intentional manipulation of parts of the mind as an object distinct from the agent. One might go so far as to say that the will is not only sometimes, but always extended, because it always involves the intentional manipulation by the mental agent of an environment that is distinct from her, even if this environment sometimes happens to be within the skull.

Literature:

Adams, F. and K. Aizawa. 2008. *The Bounds of Cognition*. Malden, MA and Oxford: Blackwell.

Baumeister, R. (2008). Free will, consciousness and cultural animals. *Are we free*. J. Baer, OUP: 65-85.

Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis* 58:10-23,

Heath, J., & Anderson, J. (2010). "Procrastination and the extended will". In C. Andreou, M. White (Eds.), *The Thief of Time: Philosophical Essays on Procrastination* (pp. 233-252). New York: Oxford University Press.

Hieronymi, P. (2009). Two kinds of mental agency. *Mental actions*. M. S. Lucy O'Brien. Oxford, OUP.

Holton, R. (2009). *Willing, wanting, waiting*. Oxford ; New York, Oxford University Press.

McGeer, V. (2007). The Moral development of first-person authority. *European Journal of Philosophy* 16(1): 81–108.

Mele, A. (2009) *Mental Action: A case study*. *Mental actions*. M. S. Lucy O'Brien. Oxford, OUP.

Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton, Princeton University Press.

Paglieri, F. (2012). "Ulysses' will: Self control, external constraints, and games". In F. Paglieri (Ed.), *Consciousness in Interaction: The Role of the Natural and Social Context in Shaping Consciousness* (pp. 179-206). Amsterdam: John Benjamins.

- Rupert, R., 2004. Challenges to the hypothesis of extended cognition. *Journal of Philosophy* 101 (8): 389-428.
- Smith, A. (2005). "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115: 236-271.
- Sprevak, M. (2009). Extended cognition and functionalism. *Journal of Philosophy*, 106, 503–527.
- Strawson, G. (2003). *Mental ballistics or the involuntariness of spontaneity*. Meeting of the Aristotelian Society, London.
- Velleman, D. (1992). "What Happens When Someone Acts?" *Mind* 101
- Vierkant, T (2013) "Managerial Control and free Mental Agency " Vierkant, T, Clark, A, Kiverstein, J (ed) *Decomposing the will*, Oxford University Press".
- Vierkant, T (2012a) "What Metarepresentation is For". In Michael Beran, Johannes Brandl, Josef Perner & Joelle Proust (eds.), *The Foundations of Metacognition*. Oxford University Press.
- Vierkant, T & Paraskevaides, A (2012b) "Mindshaping and the intentional control of the mind" in Paglieri, F Castelfranchi, C (ed) *Consciousness in Interaction*, John Benjamin.