



## **BEYOND HISTORY: THE ONGOING ASPECTS OF AUTONOMY**

BY STEVEN WEIMER

**JOURNAL OF ETHICS & SOCIAL PHILOSOPHY**

VOL. 4, NO. 1 | NOVEMBER 2009

URL: [WWW.JESP.ORG](http://WWW.JESP.ORG)

COPYRIGHT © STEVEN WEIMER 2009

## Beyond History: The Ongoing Aspects of Autonomy

Steven Weimer

CONCEPTIONS OF AUTONOMY DIFFER along a number of dimensions.<sup>1</sup> One of the more important of these concerns whether an agent's autonomy is taken to depend entirely on her present mental structure (beliefs, pro-attitudes, self-control abilities, etc.), or whether the history of that mental structure must also be taken into account. We can call accounts adopting the former approach *internalist* and those adopting the latter *historical*.<sup>2</sup> The historical approach is typically motivated by an example

---

<sup>1</sup> One such dimension I will not be addressing concerns whether autonomy is seen as, at its most basic level, a *local* property of particular aspects of an agent's life (her religious values, for example) or instead a *global* property of lives as a whole. On a *local* theory of autonomy, the autonomy of an agent's life as a whole would then be some function of the autonomy of particular aspects of her life. I will here follow Mele, whose work I will be examining and extending, in assuming a *local* approach to autonomy, on which the primary questions concern the requirements of autonomy relative to some particular aspect of one's life. However, I believe that the conditions of autonomy for which I argue could be adjusted to apply to a global approach as well. I will argue that in order to be autonomous in some area of her life, an agent must not only autonomously develop pro-attitudes in that area of life, but also autonomously maintain them and be capable of autonomously acting upon them. On a global theory of autonomy, these claims might be adjusted to require that in order to lead an autonomous *life*, an agent must not only autonomously develop a "plan of life," but also autonomously maintain that plan and be capable of autonomously acting on it. As we shall see below, Arneson does in fact include an autonomous-maintenance condition in his global account of autonomy, and suggests that a capacity for action condition might be necessary as well.

<sup>2</sup> This is not to say that these two general approaches to autonomy (i.e. the internalist and historical) exhaust all possibilities. Other theories of autonomy might deny the internalist claim that autonomy depends entirely on the agent's present mental structure, but appeal to other, *non-historical*, external conditions. For instance, Mele points out the possibility of a theory according to which the autonomy of a pro-attitude "depends on there being agent-external grounds or reasons for identifying with that value" (Alfred Mele, *Autonomous Agents* (Oxford: Oxford University Press 1995), 175, note 19). Such a theory would be neither historical nor internalist, but rather what we might call "reasons-externalist." I will ignore such accounts here. Mele is in my view correct to reject this approach on the grounds of its "departing significantly from the root notion of autonomy: self-government or self-rule" (Ibid, 160). Another, more plausible, non-historical externalist approach is what we might call the "counterfactually externalist" approach, according to which the autonomy of an agent's mental structure would depend on what that mental structure would be in some counterfactual scenario. The most promising such account of which I am aware is that offered by Christman. On Christman's theory, the autonomy of a pro-attitude depends in part on how the agent would respond were she to engage in sustained critical reflection on the pro-attitude, "and do so in light of the historical processes (adequately described) that gave rise to [it]." The pro-attitude is autonomous only if, following such reflection, the agent would not be alienated from it (John Christman, "Autonomy, History, and the Subject of Justice," *Social Theory and Practice* 33 (2007), 21). Christman's is thus a counterfactually externalist account of autonomy, but one that is historically sensitive. I have argued elsewhere that because the agent's response to the hypothetical scenario is dictated by her present mental structure, this counterfactual approach will have difficulty handling psychological twins

involving mind control, the most compelling of these perhaps being “psychological twins” cases, in which two agents come to have exactly similar mental structures, but in one case that structure is the result of ordinary character-development processes while in the other it was implanted against the agent’s will by a neurologist or team of brainwashers.<sup>3</sup> We intuitively want to say that only the first agent possesses her current mental structure autonomously, but this is a request, the advocate of the historical approach claims, that cannot be met by any internalist theory. Their present mental structures being exactly similar, the only way to distinguish between the autonomy of the two agents is, it seems, to employ a theory of autonomy which looks beyond an agent’s current beliefs, desires and abilities, and examines as well the way in which they came about. Historical accounts of autonomy will thus identify developmental processes which are incompatible with the autonomy of the eventual pro-attitude (e.g. those involving substantial coercion or manipulation, or a lack of rationality or self-control on the part of the agent).<sup>4</sup>

Although I will not defend this claim here, the historical is in my view the proper approach to autonomy. So far as I can tell, there simply need not be any relevant internal difference between the mind-controlled and those whose mental life is the result of more mundane causal forces which are intuitively compatible with autonomy. Thus if we are to distinguish between the two, as it seems we must, we will have to do so by appealing to their history.<sup>5</sup> My concern here lies instead with the question of how historical ac-

---

cases of the sort that trouble the internalist approach (see my “Externalist Autonomy and Availability of Alternatives,” *Social Theory and Practice* 35 (2009)). Because it is the drastically different developmental histories of the psychological twins that drive our intuition that their autonomy must differ as well, and because it is only *directly* historical accounts that can directly address agents’ developmental histories, it is my suspicion that it will only be historical accounts that can adequately handle such cases. I will not, though, revisit that issue here, but instead will simply set all such non-historical externalist approaches to autonomy to the side.

<sup>3</sup> See, for example, Mele, *Autonomous Agents*, 145-146.

<sup>4</sup> Different theorists will spell out the historical conditions necessary for development of autonomous pro-attitudes in different ways. For examples, see Mele, *Autonomous Agents*, Ch. 9; John Christman, “Autonomy and Personal History,” *Canadian Journal of Philosophy*, 21 (1991): 1-24; and Insoo Hyun, “Authentic Values and Individual Autonomy,” *Journal of Value Inquiry*, 35 (2001): 195-208. Although I am not here concerned with the historical conditions of autonomy, the historical requirements proposed by Mele and Christman are discussed briefly below. Raz also employs a historical approach, but his is a global rather than a local theory of autonomy. See Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press 1986), Ch. 14.

<sup>5</sup> Internalists would, of course, disagree. They can do so in two ways. First, some internalists deny that any distinction needs to be made between the autonomy of the two psychological twins. Albeit in a discussion of free action and responsibility rather than autonomy specifically, this appears to be Frankfurt’s response. See his “Three Concepts of Free Action,” in *The Importance of What We Care About* (Cambridge: Cambridge University Press 1988), 54. Other internalists are less willing to bite this bullet and attempt to locate a relevant self-control capacity that the mind-controlled twin, and she alone, will necessarily lack. David Shoemaker offers an example of this approach when he suggests that the mind-controlled are acting on motives which are not their own (i.e. not autonomous) not because their pro-

counts of autonomy can best be rounded out. Such accounts tend to focus the bulk of their attention on the historical requirements of autonomy – on specifying the conditions that enable agents to *develop* autonomous pro-attitudes. This focus is understandable, of course, given that it is the historical nature of such accounts that sets them apart, but it often comes at the cost of leaving other important aspects of personal autonomy inadequately addressed. This is unfortunate because although an agent’s autonomy in some aspect of her life does in my view turn “crucially on her *past*, on how and under what conditions her desires and values [in that aspect of her life] were formed,”<sup>6</sup> it does not entirely so turn. As Mele has argued, “there are at least *prima facie* differences among an agent’s autonomously *developing* a pro-attitude over a stretch of time, an agent’s autonomously *possessing* a pro-attitude during a stretch of time, and an agent’s being autonomous regarding the *influence* of a pro-attitude on his intentional behavior.”<sup>7</sup> It is plausible to think with Mele that the development of autonomous pro-attitudes in some aspect of one’s life – for instance, one’s occupation, religious values or artistic tastes – does not suffice for full, or as he puts it “robust,”<sup>8</sup> autonomy with respect to that aspect of life. On the contrary, and as will be argued in more detail below, it seems that an individual who previously developed autonomous pro-attitudes in some domain, but who now lacks intuitively important forms of control over the continued possession of those pro-attitudes and over the way in which they influence her behavior, for that reason lacks intuitively important forms of *autonomy* in that domain. A truly self-determining agent, it seems, will not only have had certain forms of control over the development of her personality and pro-attitudes, but will retain certain forms of control over both her present character and her conduct. If it is to accommodate this appearance, an account of autonomy will have to address not only developmental autonomy, but both ongoing aspects of autonomy as

---

attitudes have an external source, but rather because their implanted pro-attitudes are not subject to critical self-reflection and alteration. Without the ability for critical self-reflection, Shoemaker proposes, one cannot be an agent and thus cannot be an autonomous agent. See his “Caring, Identification, and Agency,” *Ethics* 114 (2003), 116-117. However, while some degree of critical self-reflective ability may well be necessary for autonomy, it is not clear that mind-controlled agents will necessarily fall short of any realistic, not overly demanding, threshold. It need not be the case that the mind-controller removes his victim’s capacity for self-reflection or otherwise renders her pro-attitudes impervious to evaluation and alteration. He might instead embed those pro-attitudes as central to his victim’s identity such that her self-evaluations will generally take those implanted values as their starting point rather than their subject. That is, she will much more often be judging other, more peripheral, aspects of her life by how well they accord with those implanted values, than vice versa. This is indeed what happens in the example Mele uses to motivate the historical approach to autonomy. See *Autonomous Agents*, 145.

<sup>6</sup> John Christman, “Introduction,” in Christman (ed.) *The Inner Citadel: Essays on Individual Autonomy* (Oxford: Oxford University Press, 1989), 9.

<sup>7</sup> Mele, *Autonomous Agents*, 138. See also Mele, “History and Personal Autonomy,” *Canadian Journal of Philosophy* 23 (1993), 273.

<sup>8</sup> Mele, “History and Personal Autonomy,” 279.

well. As Mele says, “An account of autonomy ‘relative to’ a desire that captures ‘the essence of self-government’ will illuminate each of these species (or aspects) of personal autonomy.”<sup>9</sup>

The non-historical aspects of autonomy have not, however, been adequately addressed by recent historical accounts. Sometimes, they are neglected altogether, and where they have been addressed, there is room for improvement. As a result, there is still some non-historical work to be done on the historical approach to personal autonomy. This paper aims to draw attention to, and hopefully go some way toward remedying, the need for further “illumination” of the two non-historical aspects of autonomy identified by Mele. As we shall see, Mele is a noteworthy exception to the general tendency to neglect the ongoing requirements of autonomy. Mele not only identifies and distinguishes the two non-historical aspects of autonomy with which I am here concerned, but has also offered conditions addressing each of them. His discussion and proposed conditions therefore offer useful places from which to start our examination of those aspects of autonomy.

The paper proceeds as follows. In Section One, I argue that in order for a pro-attitude to be autonomously possessed, it is not enough that it developed in an autonomous manner; it must also be maintained in an autonomous manner. I examine two “autonomous-maintenance” conditions that have been proposed, one by Mele, the other by Richard Arneson, and argue that, as those conditions stand, neither is satisfactory. What we need, I argue, is an autonomous-maintenance condition that adjusts and combines the requirements of those two conditions, such as that which I go on to propose. In Section Two, I argue that in order for an agent to be autonomous in some area of her life, it is not enough that her pro-attitudes in that area be autonomously possessed, she must also possess the capacity to autonomously act on those pro-attitudes. This behavioral requirement of personal autonomy is, however, often ignored or, in some cases, explicitly denied. Mele does acknowledge and develop such a requirement, but I argue that his proposal mistakenly equates autonomous action with effective action. The behavioral aspect of autonomy, I argue, requires merely that the agent be able to exercise her adequate self-control capacities in determining for herself whether and how to go about acting on her autonomously possessed pro-attitudes; those actions need not be successful. An individual who is able to do this – who at present both possesses autonomous pro-attitudes in some area of her life and is capable of autonomously acting on those pro-attitudes – is, I conclude, a “robustly” autonomous agent in that particular aspect of her life.

---

<sup>9</sup> Ibid.

## 1. Autonomous Maintenance

At the heart of a historical theory of autonomy will be an account of the conditions necessary for development of autonomous pro-attitudes. For instance, such an account might hold that a pro-attitude is autonomous only if during its development the agent was able to rationally reflect upon it, which would require that the agent both possessed adequate rationality and was not prevented from applying it to the pro-attitude in question by coercion, manipulation or other such forms of interference. Here, we will ignore the question of precisely what historical conditions are necessary for developmental autonomy and address instead the further question of whether the autonomous development of a pro-attitude is sufficient for its autonomous possession. According to what Arneson calls the “global” method of measuring autonomy, it is.<sup>10</sup> On that approach, if a pro-attitude developed autonomously, then that pro-attitude remains autonomous for as long as it exists. The “aggregative” method, on the other hand, holds that in order for a pro-attitude to be autonomously possessed, it must not only have been autonomously developed, but must also currently be autonomously maintained.<sup>11</sup> On the aggregative method, it is not enough for the autonomous possession of a pro-attitude that the agent had autonomous control over the way in which that pro-attitude originally came to be a part of her psychological life, she must also retain autonomous control over its continued presence within her psychological life.

Christman’s original (and since substantially revised) theory of autonomy provides an example of the global method.<sup>12</sup> If an agent approved (or would have approved) of the way in which a pro-attitude’s development induced change, and that approval was not (or would not have been) the result of a lack of rationality or self-awareness, then that pro-attitude was on Christman’s original model autonomous, and would always remain so.<sup>13</sup> In responding to Christman’s original account, Mele objects to its global approach to measuring autonomy.<sup>14</sup> Autonomy with respect to the possession of a pro-attitude is not, Mele argues, merely a matter of whether its development was conducive to self-control, but also of whether the agent has control over its continued possession. He has us consider the case of an addiction specialist, Alice, who “decides after careful reflection to make herself a heroin addict so

---

<sup>10</sup> Richard Arneson, “Autonomy and Preference Formation,” in J. Coleman and A. Buchanan (eds.) *In Harms Way: Essays in Honor of Joel Feinberg* (Cambridge: Cambridge University Press 1994), 50.

<sup>11</sup> *Ibid.*

<sup>12</sup> Christman’s original model was proposed in “Autonomy and Personal History.” As I have said, he has since abandoned the historical approach to autonomy for a counterfactual approach, albeit one that is sensitive to history. See his “Autonomy, History, and the Subject of Justice.”

<sup>13</sup> Christman, “Autonomy and Personal History,” 22.

<sup>14</sup> In Mele, “History and Personal Autonomy.”

that she can directly experience certain phenomena. She strives to develop irresistible desires for heroin.”<sup>15</sup> When Alice eventually does develop such desires, it is thus plausible to think that they have developed in an autonomous manner. Sometime later, however, she becomes “convinced that the experiment is more dangerous than she had realized and that it is time to start setting things right.”<sup>16</sup> On the basis of this realization, she “explicitly and rationally judges that it would be best to eradicate her standing desire for heroin, beginning immediately.”<sup>17</sup> Unfortunately, by this point, the grip of addiction is too strong, and Alice is incapable of eradicating her desire. “In this scenario,” Mele submits, “Alice plainly is not autonomous (or self-governing) with respect to her continued possession of the desire.”<sup>18</sup> “If she were self-governing with respect to her possession of that desire,” he tells us, “she would rid herself of it, as she judges it best. Instead, she is stuck with the standing desire and victimized by it, while rationally preferring its eradication.”<sup>19</sup>

Mele explains that what cases like this show is that “there are also non-historical constraints on autonomy relative to a desire; developmental autonomy is not sufficient.”<sup>20</sup> He goes on to propose the following non-historical necessary condition of autonomy with respect to the possession of a desire: “[an agent] *S*, who has throughout *t* a desire *D*, autonomously possesses *D* during *t* only if either (1) he is capable of shedding *D* during *t* or (2) he is incapable of shedding *D* during *t*, but he does not rationally judge it best during *t* to shed *D* during *t*.”<sup>21</sup> On this condition, then, an autonomously developed pro-attitude remains autonomous only if the agent is capable of shedding it should she rationally judge it best to do so. An account including this condition would thus employ an aggregative method for measuring autonomy.

As I have said, Mele’s criticism of the global method for measuring autonomy and his suggested non-historical, “autonomous-maintenance,” condition of autonomy both come in the context of a response to Christman’s original and exclusively historical model. Interestingly, in subsequently developing his own positive account of autonomy, Mele opts not to include that condition. On Mele’s full theory, “the autonomous possession of a pro-attitude requires authenticity regarding that pro-attitude,”<sup>22</sup> and “a necessary condition of an agent *S*’s authentically possessing a pro-attitude *P* (e.g. a value or preference) that he has over an interval *t* is that it be false that *S*’s hav-

---

<sup>15</sup> Ibid, 273. This case is discussed also, and in a bit more detail, in *Autonomous Agents*, 138-139.

<sup>16</sup> Mele, *Autonomous Agents*, 138.

<sup>17</sup> Mele, “History and Personal Autonomy,” 274.

<sup>18</sup> Ibid.

<sup>19</sup> Mele, *Autonomous Agents*, 138-139.

<sup>20</sup> Mele, “History and Personal Autonomy,” 274.

<sup>21</sup> Ibid, 279.

<sup>22</sup> Mele, *Autonomous Agents*, 156.

ing  $P$  over that interval is ... compelled\*.”<sup>23</sup> (The asterisk is being added to distinguish his conception from other possible understandings of compulsion.) Mele’s final approximation of a sufficient condition of compulsion\* holds that an agent  $S$  is compelled\* to possess a pro-attitude  $P$  if, among other things, “[A]  $S$  comes to possess ...  $P$  in a way that bypasses  $S$ ’s (perhaps relatively modest) capacities for control over his mental life; and [B] the bypassing issues in  $S$ ’s being practically unable to shed  $P$ .”<sup>24</sup> The self-control capacities relevant to (A) include such things as the abilities to rationally assess and revise one’s values and principles and to identify with values on the basis of informed, critical reflection.<sup>25</sup> Mele explains the notion of practical unshedability at work in (B) as follows: “Insofar as the conditions that would empower [an agent] to shed [a value] are ‘beyond his control’ – that is, insofar as his psychological constitution precludes his voluntarily producing those conditions – and the obtaining of those conditions independently of [his] voluntarily producing them is not in the cards, he is apparently stuck with the [value]. I will say that any agent who is stuck in this way with a value (during  $t$ ) is *practically unable* to shed it (during  $t$ ).”<sup>26</sup> On this account, then, in order to *autonomously* possess a pro-attitude, an agent must *authentically* possess it, and in order to authentically possess a pro-attitude, it must be true *either* that the pro-attitude did not come about in a way that bypassed the agent’s capacities for control over her mental life, *or* that she is currently able to rid herself of the pro-attitude. If you were unable to play any role in bringing a pro-attitude about, *and* if you are now unable to shed it, the idea seems to be, then that pro-attitude is in no way your own – i.e. is not autonomously possessed.

Mele’s developed theory thus does not include the necessary condition he earlier proposed: it is on that account not necessary for autonomy with respect to the possession of a pro-attitude that the agent be able to shed the pro-attitude, should she rationally judge it best to do so. For it might be true of a pro-attitude that the agent would not be able to shed it if she judged it best to do so, but nevertheless not the case that that pro-attitude is compelled\* because in order to qualify as compelled\* the pro-attitude must both have come about in a way that bypassed the agent’s self-control capacities and currently be practically unshedable. The unshedability of a desire is thus not sufficient for its non-autonomy on Mele’s account. While Mele’s

---

<sup>23</sup> Ibid, 166.

<sup>24</sup> Ibid, 172. Mele’s formulation of that compulsion\* condition includes two further qualifiers which can be set aside here. First, that a pro-attitude came about in a way that bypassed the agent’s self-control capacities is irrelevant if the bypassing was itself arranged or performed by the agent. Second, if the agent either “presently possesses [or] earlier possessed pro-attitudes that would support his identifying with [the pro-attitude], with the exception of pro-attitudes that are themselves practically unshedable products of unsolicited bypassing,” then the agent is not said to be compelled\* to possess the pro-attitude in question. (172)

<sup>25</sup> Ibid, 166-167.

<sup>26</sup> Ibid, 153. Mele addresses this explanation to a specific example of his. My adjustments serve only to make it apply more generally.



developed theory therefore does not include the autonomous-maintenance condition he earlier proposed, it does not preclude that condition either. Mele claims only that it is a necessary condition of an agent's autonomously possessing a pro-attitude that the pro-attitude not be compelled\*; although he has opted not to do so, it remains open to him to include an additional necessary condition requiring that the pro-attitude be maintained in an autonomous way. That is, Mele tells us only one way in which a pro-attitude may fail to be autonomously possessed – namely, its being compelled\* – there may well be additional ways in which a pro-attitude might be disqualified as well.

I would argue that we should indeed add some autonomous-maintenance condition to Mele's account because, owing to its lack of such a condition, that account does not, as it stands, deem pro-attitudes such as Alice's irresistible desire for heroin non-autonomous. As we have seen, Alice is autonomous with respect to the development of her desire for heroin. Mele's developed theory explains why this is: Alice having decided to develop that desire after careful reflection, it is clear that it did not come about in a way that "bypassed her capacities for control over her mental life," but was, rather, in an important sense brought about by those self-control capacities. Consequently, Alice's desire for heroin was not compelled\*, and as it is only compelled\* pro-attitudes that Mele's account deems non-autonomous, we do not on that account have the resources with which to say that Alice's desire is possessed non-autonomously. This is of course unacceptable. Mele would agree: he includes the Alice example in *Autonomous Agents*, taking from it the lesson that "the autonomous development of a pro-attitude *P* that one possesses throughout *t* is not sufficient for the autonomous possession of *P* throughout *t*. Alice (by hypothesis) autonomously developed her standing desire for heroin; but there was a time during which she possessed it nonautonomously."<sup>27</sup> As Mele's account stands, however, it does not imply that Alice's desire is at any time possessed non-autonomously.

Mele is correct to think that the autonomous development of a pro-attitude is not sufficient for its autonomous possession. In order to incorporate this thought into a historical account of autonomy, however, we must supplement the historical conditions necessary for the development of autonomous pro-attitudes with some present-directed autonomous-maintenance condition necessary for the possession of autonomous pro-attitudes. Although Mele does not include such a condition in his developed account, we find in his work two possible such conditions that will serve as useful starting points in our search for an appropriate autonomous-maintenance condition. I will argue, however, that, as they stand, neither of those possible conditions will quite do.

First, we might consider simply separating the "practical unsheddability" clause of Mele's compulsion\* condition from its "bypassing" clause and

---

<sup>27</sup> Ibid, 139.

make the former an additional, and present-directed, necessary condition of the autonomy of a pro-attitude. That is, whereas in its original form, Mele's compulsion\* condition holds that the practical unsheddability of a pro-attitude implies its non-autonomy only if that pro-attitude also came about in a way that bypassed the agent's self-control capacities, we might instead say that the present unsheddability of a pro-attitude on its own renders that pro-attitude non-autonomous. This, though, would give us an autonomous-maintenance condition that is much too strong. As Christman has said, "it is clear that the ability to do otherwise, or even to desire otherwise, is not in itself a plausible requirement for autonomy ... we are often in states that are not revisable by us without tremendous pain, effort, or even outside assistance."<sup>28</sup> To take Frankfurt's example, when Luther declared, "Here I stand; I can do no other," his inability to abandon or betray his values should not be taken to mean that those values are possessed non-autonomously.<sup>29</sup> This is, however, precisely what a practical unsheddability autonomous-maintenance condition would imply. During the relevant time period (say, just prior to and during his appearance before the Diet of Worms), Luther's reformist values are practically unsheddable – during that period, his psychological constitution precludes his bringing about conditions that would empower him to shed those values, and those conditions will not obtain independent of his voluntarily producing them. It is not at all clear, however, that this inflexibility is inconsistent with the autonomy of those values. On the contrary, it is reasonable to think that Luther's values were autonomously possessed.<sup>30</sup> Similarly, a deeply devoted father's parental values might be autonomously possessed despite the fact that those values are so thoroughly entrenched that he is at present unable to uproot or attenuate them.<sup>31</sup> An autonomous-maintenance condition extracted from Mele's unsheddability clause would deny this, and is thus too strong.

Since the present-directed aspect of Mele's developed theory of autonomy does not offer an adequate autonomous-maintenance condition, we might return to the condition he earlier proposed, which, unlike the unsheddability clause, was intended to provide a stand-alone necessary condition on the autonomy of pro-attitudes. According to that condition, recall, a pro-attitude is possessed autonomously (during some time period) only if (during that time period) the agent is either capable of shedding it, or is incapable of

---

<sup>28</sup> Christman, "Autonomy, History, and the Subject of Justice," 24.

<sup>29</sup> Harry Frankfurt, "The Importance of What We Care About," in *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988), 86. Frankfurt is not there discussing autonomy, but his point is, I think, plausible in this context as well.

<sup>30</sup> Whether or not this is true would of course depend upon the details of Luther's case and the account of autonomy to which we apply them. On a historical account of autonomy, this consistency would be attributed to the fact that Luther's inflexibility is presumed to be in some important sense self-imposed, that Luther made himself such that he cannot during the relevant time period uproot his reformist values.

<sup>31</sup> Mele, *Autonomous Agents*, 153.

shedding it, but does not rationally judge it best to shed it.<sup>32</sup> This condition Luther and the deeply devoted father might satisfy, for although neither agent can rid himself of his values, neither judges it best to do so. When we turn to examine an adjusted version of the Alice example, however, we see that this condition is too weak.

Mele's originally proposed condition requires for the autonomy of a pro-attitude that the agent be able to implement a rational judgment that it would be best to shed a pro-attitude, should she form one. This seems correct to me, so far as it goes. But it does not go quite far enough, for it says nothing about the inclination and ability to form such judgments, which are, I believe, equally necessary for ongoing autonomy with respect to possession of a pro-attitude. Imagine another experimental heroin addict, Beth. Unlike Alice, who rationally judged it best to shed her desire for heroin, but was incapable of implementing this judgment, Beth would have no problem implementing such a judgment, should she form one – she is possessed of exceptional will power, and has access to rehabilitation programs that would enable her to effectively employ that will power in ridding herself of her addiction. While her addiction to heroin has therefore not removed Beth's ability to effectively act on the basis of rational self-reflection, it has removed both her disposition and her ability to engage such reflection. Beth's addiction prevents her from ever thinking about whether she ought to end the experiment and get clean – if she is not introspecting on the pleasure she feels while on a high, she is entirely focused on getting her next fix. Moreover, even if she were to reflect on whether she should end the experiment, she could not do so rationally – any attempt at self-reflection would either end prematurely due to Beth's inability to remain focused on anything other than satisfying her desire, or issue in an irrational judgment that her experiment is not dangerous, that she has everything under control. Beth would nevertheless satisfy Mele's original autonomous-maintenance condition, for *if* she did judge it best to do so, she would be able to shed her desire.

Beth is, however, every bit as “stuck” with, and victimized by, her desire for heroin as is Alice. Neither Alice nor Beth, I submit, possesses the sort of ongoing control over her desire for heroin that is necessary for the ongoing autonomy of that desire. An autonomous agent in Beth's situation would, it seems, recognize that her circumstances have changed in an important way demanding, at the very least, that she reconsider her earlier decision to develop a heroin addiction, and likely that she reverse that decision as well – would recognize, like Alice does, “that the experiment is more dangerous than she had realized and that it is time to start setting things right.”<sup>33</sup> Beth has lost autonomous control over her desire because she lost the ability to recognize this. It is thus not simply the fact that Alice and Beth cannot in their present circumstances shed their desires that renders them non-

<sup>32</sup> Mele, “History and Personal Autonomy,” 279.

<sup>33</sup> Mele, *Autonomous Agents*, 138.

autonomous with respect to those desires, for neither can Luther and the devoted father, whose inflexibility is intuitively consistent with the ongoing autonomy of their values. What renders Alice and Beth non-autonomous with respect to their desires is instead the fact that they are incapable of appropriately responding to an important change in their circumstances. Alice and Beth's circumstances have changed in such a way that autonomy seems to require that they be able to adjust their pro-attitudes in light of that change. More specifically, an autonomous response to their altered circumstances seems to require that they be capable of both rationally recognizing that their desires for heroin should be shed, and then making the corresponding self-alterations. Beth is incapable of performing the first, judgment-formation, step of that autonomous self-alteration process. What renders Beth non-autonomous with respect to her desire for heroin is that she is incapable of rationally recognizing that her desire should be shed – she is in fact incapable both of recognizing that her circumstances have changed in an important way indicating the need to review that desire, and of performing such review in a rational manner. Put another way, Beth has lost autonomous control over her desire for heroin because she has lost the ability to appropriately recognize and process new and vitally important information – namely, the unexpected danger of her experiment. Alice, on the other hand, does appropriately recognize and process that information, for she does form a rational judgment that it would be best to end her experiment and shed her desire for heroin. What renders Alice non-autonomous with respect to her desire for heroin is instead, and as Mele saw, her inability to perform the second, judgment-implementation, step of the autonomous self-adjustment process. What these cases show, therefore, is that the ability to effectively implement a rational judgment that one should shed a pro-attitude is not sufficient for autonomous-maintenance of that pro-attitude; also necessary is the disposition and ability to form such judgments, to rationally recognize when one's pro-attitudes should be shed.

That it is nothing peculiar to these two cases (such as the fact that they involve addictive drugs, for instance) that drives this conclusion can be seen by considering adjusted versions of the examples involving Luther and the deeply devoted father. Imagine that in the evening prior to his Diet of Worms appearance, Luther was confronted with reliable evidence that the portions of the Bible on which his reformist views were based were not authentic, but had been inserted a generation earlier by a rebellious translator. If Luther's commitment to his reformist values was such that it rendered him unable to comprehend and sincerely evaluate that evidence, or unable to implement a rational judgment that he should adjust his values in light of the evidence, we may begin to doubt whether his inflexibility was consistent with autonomy after all. An autonomous agent would, it seems, recognize that such evidence represents new and crucially relevant information demanding, at the very least, serious consideration and, depending on the outcome of that consideration, perhaps the adjustment of some of his beliefs and val-

ues.<sup>34</sup> Or, suppose that, as in one of Mele's examples, CIA agents convince a deeply devoted father "that the CIA will ensure his children's flourishing if he uproots his parental values and that, otherwise, they will destroy his children's lives."<sup>35</sup> In that case, if the father is unable to see that this new information demands serious consideration, and indeed that it offers him a "decisive reason for shedding his parental values,"<sup>36</sup> we would, I think, doubt whether he retains autonomous control over those values. If the father cannot see that his parental values themselves demand that he attempt to shed those values in this case, then we would begin to think that his parental values now control him, he does not control them.<sup>37</sup> Or, if he does see that the CIA's threat provides him with a decisive reason to shed those values, but is unable to effectively implement the rational judgment to which that leads him, then he will be in the same, non-autonomous position as was Alice in Mele's original description of her case: he will be stuck with his parental values, and victimized by them, while rationally preferring their eradication.<sup>38</sup>

In each of these cases, an autonomous response to the agent's importantly altered circumstances requires that he or she be able both to rationally recognize that a pro-attitude ought to be adjusted, or shed, and then to make the corresponding self-alteration. Mele's original autonomous-maintenance condition addresses only the latter, judgment-implementation, step of this autonomous self-adjustment process and is thus in need of supplementation.

We can do this by adding to Mele's proposed autonomous-maintenance condition an adjusted version of the condition offered by Arneson. Arneson has suggested that,

To live an autonomous life an agent must decide on a plan of life through critical reflection and in the process of carrying it out, remain disposed to subject the plan to critical review if disturbing or unanticipated evidence indicates the need for such review. The level of counterevidence that would trigger a critical review is set by the agent. Finally, the autonomous agent must remain disposed to undertake critical review of the decision that set the level of counterevidence that would trigger review, if surprising evidence becomes available that calls into question that trigger-setting decision.<sup>39</sup>

---

<sup>34</sup> It appears that Luther did indeed possess the sort of flexibility I believe necessary for autonomy. His famous declaration, "Here I stand. I can do no other," if it was in fact made at all (which is disputed), was reportedly prefaced thus: "Unless I shall be convinced by the testimonies of the Scriptures or by clear reason ... I neither can nor will make any retraction, since it is neither safe nor honourable to act against conscience." See Vergilius Ferm, *An Encyclopedia of Religion* (Bloomington, Indiana: Littlefield Adams 1959), 830.

<sup>35</sup> Mele, *Autonomous Agents*, 153.

<sup>36</sup> *Ibid.*

<sup>37</sup> Mele, "History and Personal Autonomy," 273.

<sup>38</sup> *Ibid.*, 275

<sup>39</sup> Arneson, "Autonomy and Preference Formation," 49.

The idea here is that “If all goes well, the autonomous agent might sail through life without ever encountering evidence that indicates a need for critical review of her values and plans. But she is disposed to reflect if that course seems advisable.”<sup>40</sup>

We can, I think, tinker with a few of its details, but the thrust of Arneson’s condition captures well the sort of disposition for rational self-reflection that is necessary for ongoing autonomy, but was lacking in Mele’s originally proposed condition. Arneson’s condition will need to be adjusted, first of all, because it takes as its target an agent’s entire life. We can, though, easily apply it to individual pro-attitudes. Second, we should make clear that the agent must be able to critically review a pro-attitude in a *rational* manner. This is perhaps implied by another portion of Arneson’s discussion, but should be made explicit.<sup>41</sup> Finally, I believe we should remove the subjective trigger-setting portion of Arneson’s proposed autonomous-maintenance condition and focus directly on whether or not unanticipated evidence indicates the need for critical review of the pro-attitude itself. On Arneson’s formulation, it seems that while the agent herself determines the amount of counterevidence that would require critical review of her pro-attitude, the amount of evidence that would demand critical review of that decision (i.e. the decision as to how much counterevidence would require critical review of the pro-attitude) is not left for the agent to determine, but is instead set objectively. So, for instance, if Alice decides that she will reconsider her desire for heroin only if her husband asks her to – she is not confident in her own ability to make good judgments while under the influence of heroin, and trusts that her husband will competently look out for her best interests – then the ongoing autonomy of that desire will require that it be reconsidered only if and when her husband asks her to do so. However, it is not up to Alice to determine how much (or presumably what sort of) evidence would demand review of that decision, which is instead determined objectively. There is a fact of the matter as to whether or not the available evidence indicates that Alice needs to reconsider that decision. Presumably, Alice’s trigger-setting decision would be called into question by evidence that her husband had himself become addicted to heroin, say, or that he was planning to file for divorce and, in the ensuing custody battle, to argue that her addiction made Alice an unfit mother. If Alice is unable to recognize that such evi-

---

<sup>40</sup> Ibid.

<sup>41</sup> In summarizing his proposed condition, Arneson tells us that “the autonomous agent on this construal always is disposed to reflect should the trigger levels be reached and never destroys her rational capacities to engage in such reflection.” (“Autonomy and Preference Formation,” 49) Although this statement and an earlier footnote (note 13, 73-74) both suggest that autonomous maintenance is precluded only by the intentional destruction of one’s rational faculties, it seems clear that what matters here is simply whether the agent possesses the rationality necessary for autonomous self-reflection, and not why she does or does not. An individual who loses the capacity for rational critical review as a result of a car accident is no more able to autonomously maintain her pro-attitudes than is an individual who intentionally destroys her capacity to do so.

dence indicates the need to reconsider her trigger-setting decision, and/or unable to perform such reconsideration in a rational manner, then we would have reason to doubt that she has autonomous control over this aspect of her life. These are the kinds of evidence that an autonomous agent would recognize as providing a decisive reason to reconsider her decision to defer to her husband's judgment.

This seems to me correct. However, I see no reason not to apply the same sort of analysis to Alice's desire for heroin itself. Once we adopt a local approach to autonomy, we see that we really have two separate pro-attitudes here, each of which may or may not be autonomous: the desire for heroin, and an intention (based on a decision) to subject that desire to critical review only if her husband asks her to. Once we see this, we see that we have no reason to apply separate autonomous-maintenance conditions – one employing a subjective threshold, one an objective – to the two pro-attitudes. Just as there is a fact of the matter as to whether the available evidence indicates that Alice needs to subject her trigger-setting intention to critical review, there is a fact of the matter as to whether the available evidence indicates that she needs to subject her desire for heroin to critical review. On Arneson's approach, we are instructed to defer to Alice's judgment on the second, but not the first of these questions. This asymmetry is awkward and unnecessary. To see why, simply suppose that, as in Mele's original version of the case, evidence arises indicating that Alice's experiment is more dangerous than she had originally realized. On Arneson's formulation, the question we must ask is not whether this evidence indicates the need to review Alice's desire for heroin, as would seem natural, but rather whether it indicates the need to review her intention to reconsider that desire only if asked to do so by her husband. Should it? If we answer this question in the negative, then Arneson's account implies that Alice's desire for heroin is autonomously possessed, which, as Mele says, is implausible. But what grounds could we have for answering in the affirmative? Unlike evidence suggesting that her husband is himself addicted to heroin, or that he is planning to use Alice's addiction against her in a custody battle, evidence regarding the unexpected dangerousness of the experiment tells us nothing directly about the husband's ability or motivation to make good decisions on her behalf. Hence, if such evidence indicates the need to review her trigger-setting intention, as it must if we are to avoid the implausible conclusion that Alice's desire for heroin is autonomous, this could only be because we think that, were her husband competent and properly motivated, he would surely recognize that evidence as a reason to ask Alice to reconsider her desire for heroin. That is, we think that a good surrogate decision-maker would see the unexpected danger of the experiment as a sufficient reason to subject the entire addiction-development plan to critical review; we think that this is the appropriate response to such evidence. This, though, simply amounts to thinking that that evidence indicates the need to review her desire for heroin. The only way to avoid an implausible answer to Arneson's awkward and asymmetrical ques-

tion is thus to in effect ignore it by answering instead the more natural and more direct question of whether or not the available evidence indicates the need to review the relevant pro-attitude itself, as opposed to the intention concerning when to review that pro-attitude.<sup>42</sup>

When we make these adjustments we get the following: a pro-attitude is autonomous only if the agent is disposed and able to subject it to rational critical review if disturbing or unanticipated evidence indicating the need for such review becomes available. I believe this adequately captures the judgment-formation step of the autonomous self-adjustment process. Arneson's autonomous-maintenance condition is, however, the reverse of Mele's in that it does not address the second, and equally necessary, judgment-implementation, step. Our adjusted version of Arneson's condition requires only that the agent be able and disposed to subject her pro-attitudes to rational review; it says nothing about the consequences of such review. An agent may be disposed and able to recognize when her circumstances have changed in such a way as to necessitate the re-examination of her pro-attitudes and to perform that re-examination in a rational manner, but be unable to implement the judgment in which it issues. An addict like Alice, for example, may in the light of new evidence subject her desire for heroin to rational review, judge it best to shed that desire, but be unable to put that judgment into effect. As this sort of unwanted compulsion is the paradigm of a non-autonomous desire, our adjusted version of Arneson's autonomous-maintenance clearly cannot stand alone. What is needed is an autonomous-maintenance condition which accounts for both steps of the self-adjustment process, such as we get when we combine the requirements proposed by Mele and Arneson. The resulting condition would hold that it is a necessary condition of the autonomy of a pro-attitude that: (1) the agent is disposed and able to subject the pro-attitude to rational critical review if disturbing or unanticipated evidence indicating the need for such review becomes available, and (2) the agent is capable of shedding the pro-attitude should such review issue in a rational judgment that it is best to do so. On a historical account of autonomy including this autonomous-maintenance condition, it will not be enough for the ongoing autonomy of a pro-attitude that it was developed in an appropriate manner.<sup>43</sup> The agent must also remain disposed and

---

<sup>42</sup> To the question of exactly how we should go about answering this objective question – of what the relevant objective standards of evidence might be – I will not in this paper offer an answer. My aim here is merely to suggest that it is this objective question that we should be asking.

<sup>43</sup> The developmental autonomy of a pro-attitude would, though, presumably be necessary for its autonomous possession. The proposed autonomous-maintenance condition is presented as necessary for autonomy with respect to the possession of a pro-attitude, but not sufficient, because I presume that such a condition would on a historical account of autonomy be combined with some historical condition(s) addressing the autonomous development of a pro-attitude. This is because a brainwashed agent who played no role whatsoever in the development of a pro-attitude may well satisfy the autonomous-maintenance condition I have proposed. Although the historical requirements of autonomy lie outside the



able to rationally review her pro-attitude if she detects appropriate signals,<sup>44</sup> and to shed the pro-attitude, should that review issue in a judgment that it would be best to do so.

I will briefly explain what I take this condition to imply. First, the requirement is not that the “autonomous person must exercise critical reason continuously, or even frequently,” which as Arneson says, would be a mistake.<sup>45</sup> Nor is it necessary that the agent be continually able to shed her pro-attitudes, which we have seen to be mistaken as well. Rather, the proposed autonomous-maintenance condition requires merely that the agent remain properly disposed to subject her pro-attitude to critical review<sup>46</sup> and, if such review issues in a judgment that it would be best to shed the pro-attitude, that she be able to effectively implement that judgment. Consequently, the proposed condition allows for an agent to be on “auto pilot” with respect to a pro-attitude, merely monitoring in a passive way for evidence relevant to that pro-attitude, for long stretches of time without losing autonomy with respect to that pro-attitude. It is only if and when evidence indicating the need for critical review becomes available that autonomy requires the agent to retake active control of that aspect of her life by rationally reconsidering, and perhaps adjusting or shedding, her pro-attitude(s).

Second, although I will not here attempt to spell out precisely when it is that evidence indicates the need for critical review of a pro-attitude, it is clear that the trigger for such review will vary from pro-attitude to pro-attitude and from agent to agent. To take a trivial example, the evidence provided by an agent’s tasting a new flavor of ice cream would not without some extravagant story indicate the need to review her moral values, but very well may indicate

---

scope of this paper, I assume that on a historical account of autonomy, such an agent will be said to lack autonomy with respect to her brainwashed pro-attitudes, and that autonomous-maintenance is thus not sufficient but merely necessary for autonomy with respect to the possession of a pro-attitude.

<sup>44</sup> Arneson, “Autonomy and Preference Formation,” 50.

<sup>45</sup> *Ibid.*, 48.

<sup>46</sup> A less demanding alternative to this first half of the proposed condition might hold that it is a necessary condition of the autonomy of a pro-attitude that, if evidence indicating the need for rational critical review is available, the agent does in fact subject the pro-attitude to such review. This alternative would not require an ongoing disposition to subject the pro-attitude to critical review and, as such, offers a historical rather than a present-directed condition of autonomy. The requirement would in fact be (roughly) that the agent has not neglected evidence indicating the need for rational critical review of a pro-attitude, that if and when such evidence has been available, the agent responded to it in an appropriate manner. Hence, if no such evidence has ever been available, it would not on this condition matter whether the agent is or ever has been capable of subjecting her pro-attitude to rational critical review. An account of autonomy employing this alternative would thus employ a global method for measuring autonomy, for it would allow that “an agent could be autonomous at this moment even though she is incapable of subjecting her [pro-attitudes] to critical review at this moment.” (Arneson, “Autonomy and Preference Formation,” 50) I follow Arneson in rejecting such an approach. In my view, the autonomous agent is disposed to appropriately respond to important changes in her circumstances, regardless of whether or when last such changes have occurred.

the need to review her ice cream preferences. For another person already familiar with that flavor, tasting it again likely would not indicate the need to reopen questions regarding her ice cream preferences, for her preferences were formed in light of a familiarity with that flavor. The present-directed, autonomous-maintenance condition I have proposed is therefore historically sensitive in an important way. Whether or not some evidence indicates that an agent needs to critically review a pro-attitude depends in large part upon her history. Evidence with which an agent is already familiar and which has been taken into account in the development of her pro-attitudes can by that agent be thoughtlessly, yet autonomously, dismissed. If that same evidence is entirely new to another agent, however, such a dismissal might exhibit a cognitive inflexibility which is incompatible with the ongoing autonomy of her pro-attitudes.<sup>47</sup> An implication of this is that the more evidence taken into account in the development of a pro-attitude, the less likely it becomes that that pro-attitude will later be called into question. Owing to Luther's extended theological studies, the likelihood of his encountering evidence indicating the need for critical review of his reformist values is far lower than it would be for someone who possesses the same values, but on the basis of considerably less information. Luther need not re-examine his values in light of an argument for priestly indulgences which he has already considered and rejected. If that argument is new to the latter, less knowledgeable, reformer, however, the ongoing autonomy of his values may require that he re-examine those values.

## 2. Capacity for Action

This section deals with the second of the ongoing aspects of autonomy with respect to a pro-attitude identified by Mele: autonomy with respect to

---

<sup>47</sup> We might worry here about evidence that is new to an agent, but should not be. If evidence that now indicates the need for review of an agent's pro-attitude was available during the development of that pro-attitude, but she neglected to consider it, we might be reluctant to say that its present availability demands a re-examination of that pro-attitude. For example, although we would need to know more of the details of her case, we might think that Alice should have foreseen the danger of her experiment – should have seen that her plan of addicting herself to heroin could go wrong in just the sort of way it in fact did. On the basis of this thought, we might be reluctant to say that when she later comes to realize the danger she should have seen all along, autonomy requires that she be able to reconsider her desire in light of that evidence. If we are inclined to think this way, we could adjust the proposed autonomous-maintenance condition to exclude from “evidence indicating the need for rational critical review of a pro-attitude” any evidence which should have been taken into account in the development of that pro-attitude. We would of course have to spell out the relevant “should.” An alternative way to handle such cases might be to say that the agent's failure to consider available information relevant to a pro-attitude has implications for the autonomous *development*, rather than the autonomous *maintenance*, of that pro-attitude. That is, we could say that as a result of the agent's failure to consider such evidence, she is less autonomous with respect to the development of her pro-attitude than she otherwise would have been. I make no stand on which of these two approaches is superior. Hence, while I will not include an exception of the sort described above here, I am open to the possibility.

the influence the pro-attitude has on an agent's behavior. Some theorists deny that this is in fact an aspect of autonomy. On their view, personal autonomy consists entirely in psychological autonomy. Hence, while the psychological abilities "to form intentions of a sufficiently complex kind, and plan their execution," might on this view be necessary for autonomy, the ability to in fact execute those intentions is not.<sup>48</sup> An agent's ability to control her behavior, to bring her conduct into line with her pro-attitudes, is on this view a matter of *freedom*, and says nothing about her autonomy.<sup>49</sup>

I reject this view. I believe that personal autonomy comprises not only psychological autonomy, but autonomy with respect to one's behavior as well, and this latter aspect of autonomy clearly requires more than the possession of autonomous pro-attitudes and the ability to plan or intend to act on them. It requires also that the individual be able to in fact act on those pro-attitudes, to carry out at least some of her intentions and thus to pursue at least some of her ends. As Robert Young puts it, "to be autonomous requires the capacity to *do* those things which figure prominently in one's life-plan."<sup>50</sup> This is because "it is a platitude that autonomous agents ... possess and exercise some degree of control over their lives,"<sup>51</sup> and there is more to agents' lives than their psychological lives. There is also their physical life, their conduct in the world, and our account of personal autonomy should address control over this aspect of life as well. Put another way, autonomous

---

<sup>48</sup> Raz, *The Morality of Freedom*, 373

<sup>49</sup> For example, Christman argues that "There are a variety of skills and abilities that are fundamental to acting intentionally that must be included as part of the requirements for autonomy. The ability to effectively form intentions to act, then, along with the various skills that this requires, must be seen as necessary for autonomy. It must be noted further that the ability to act – successfully and as planned – cannot be what we mean here. I often am prevented from acting or completing my plans because of the happenstance of my surrounding circumstances. Such circumstances certainly make me less free (in a certain sense of freedom), but they do not make me less autonomous, at least if this latter term is to retain any of its conceptual distinctiveness. So the competence conditions for autonomy merely refer to the effective ability to form intentions to act but not to complete such actions" ("Autonomy, History, and the Subject of Justice," 20). Rejection of this view need not deprive autonomy of its "conceptual distinctiveness," I would argue, because while some sense of freedom will on the alternative approach I will adopt be necessary for autonomy, they will not be sufficient for it. Although I will not here present a complete account of autonomy, I will in the conclusion suggest that to be autonomous in some area of one's life one must both possess autonomous pro-attitudes in that area and be capable of autonomously acting on them. Hence, even if the latter requirement roughly amounts to (some conception of positive) freedom in that area of one's life – the freedom *to* act on one's relevant pro-attitudes – this is not enough for autonomy with respect to that aspect of one's life, which requires also that those pro-attitudes are autonomously possessed. And the freedom to act on a pro-attitude tells us nothing about whether or not it is autonomously possessed. The autonomous possession of pro-attitudes requires instead, to put it roughly, that the agent both had control over their development and retains control over their continued presence in her mental life.

<sup>50</sup> Robert Young, *Personal Autonomy: Beyond Negative and Positive Liberty* (New York: St. Martin's, 1986), 49, emphasis added.

<sup>51</sup> Mele, *Autonomous Agents*, 9.

self-governance involves governing (in some sense and to some degree) not only one's mind, but one's behavior as well. An individual who possesses a robustly autonomous set of pro-attitudes and who is competent to form intentions to act on the basis of those pro-attitudes, but is plagued by a chronic inability to execute his intentions, thus does not lead a robustly autonomous life. Consider, for instance, Mele's example of "a hospitalized man who was wounded in battle and now is totally paralyzed and incapable even of communicating. Imagine also that the man's cognitive faculties are in good working order ... and that he devotes some time to the question of how he ought to lead his life."<sup>52</sup> While his condition therefore might have little effect on his *psychological* autonomy, it does, I submit, radically affect his *personal* autonomy. By rendering him incapable of overt action, his paralysis drastically reduces the control he has over his life. Putting it roughly, whereas prior to his injury he was presumably able to play a significant role both in determining his ends (his values, preferences and other such pro-attitudes) and in determining whether and how to go about pursuing those ends, he has now to a significant extent lost his capacity for the latter sort of self-determination.<sup>53</sup> To in this way lose a significant portion of one's control over one's life is to lose a significant portion of one's autonomy.<sup>54</sup> In order to accommodate this judgment, our account of autonomy must not only specify the conditions necessary for autonomy with respect to the development and ongoing possession of pro-attitudes, but also the conditions necessary for autonomy with respect to the influence of pro-attitudes on one's behavior.

Mele's initial proposal in this regard was to make it a necessary condition of autonomy with respect to the influence of a pro-attitude on an agent's behavior (what he calls *I-autonomy*) at some time  $t$  that either (1) the agent is capable of refraining from acting on that pro-attitude at  $t$  and capable of acting on it at  $t$  or (2) the agent lacks exactly one of these powers, but holds no rational judgment concerning what it would be best to do regarding action on the basis of the pro-attitude that is thwarted by her lacking that power.<sup>55</sup> Alice

---

<sup>52</sup> Ibid, 191.

<sup>53</sup> He need not have lost the ability to pursue all of his ends, for he may be able to mentally act on the basis of some of his pro-attitudes. In a portion of Mele's description of the case I omitted, he says that the paralyzed man "spends much of his time pondering philosophical and mathematical problems and developing solutions." (*Autonomous Agents*, 191) We can assume that in so doing he acts on the basis of some of his pro-attitudes – most obviously, a desire to solve those problems, or perhaps a more general desire to stay active mentally, or to develop and exercise his intellectual capacities.

<sup>54</sup> It seems that Arneson would agree. He tells us that "the autonomous person's critical reflection is not an idle wheel spinning aimlessly. Her critical reflection has a shaping power over her values, preferences, *and conduct*." ("Autonomy and Preference Formation," 50, emphasis added) Arneson does not, however, specify the sort of shaping power over conduct relevant to autonomy. It is this that I attempt to do in what follows.

<sup>55</sup> Mele, "History and Personal Autonomy," 278-29. As Mele formulates it, this condition is applied to a desire to do some particular action at  $t$ . I have adjusted it to apply to pro-attitudes more generally.

thus lacks I-autonomy with respect to her desire for heroin because she is unable to refrain from acting on that desire, and this inability thwarts her rational judgment that it would be best not to act on that desire.

As was the case with Mele's initial autonomous-maintenance condition, this I-autonomy condition is not included in his developed theory of autonomy. In *Autonomous Agents*, Mele's discussion of the behavioral aspect of autonomy with respect to a pro-attitude focuses instead on the capacity to autonomously act on a pro-attitude which, as we shall soon see, is not equivalent to I-autonomy with respect to a pro-attitude, but a more demanding behavioral requirement. I believe that this shift in focus is a step in the right direction and that Mele's account of the requirements of autonomous action on a pro-attitude represents an improvement on the I-autonomy condition he earlier offered. I will argue, however, that that account is misguided in one important way.

According to Mele, a deliberative, intentional action  $A$  is autonomous if it issues directly from an intention to  $A$  that was autonomously formed; an intention is autonomously formed if it issues directly from a decisive evaluative judgment that it would be best to  $A$  now that was autonomously made; and a decisive evaluative judgment is autonomously made if it issues directly from deliberation that was autonomously conducted.<sup>56</sup> In simple cases, autonomous action therefore "requires nothing more than an autonomous psychological process ... the ability to execute the intention in which it concludes, and an intentional action constituting the execution of that intention."<sup>57</sup> The question whether an action was autonomously performed thus typically leads us to the question of whether the psychological processes, and more specifically the deliberative processes, giving rise to that action were autonomously conducted.<sup>58</sup> Mele tells us that there are three ways in which an agent might lack autonomy with respect to her deliberative processes. First, the pro-attitudes on the basis of which she deliberates might be non-autonomous. Since this will obviously depend upon our account of the conditions necessary for autonomy with respect to the possession of pro-attitudes, this possibility can be set aside here.

---

<sup>56</sup> Mele, *Autonomous Agents*, 177-178. Mele's ensuing discussion suggests that one mental state "issues directly from" another if it is the "nondeviant causal product, in part" of that mental state. (Ibid, 193)

Mele makes clear that his focus on deliberative, intentional action should not be taken to mean that he thinks that it is only such actions that can be autonomous. That focus is motivated instead by the fact that "wanting it to be true that we sometimes act autonomously, we want most that this be true in the sphere of full-blown, deliberative actions. An autonomy limited, for example, to relatively spontaneous actions performed in the absence of deliberation would not amount to much." (Ibid, 193) I will follow Mele in restricting my attention to deliberative intentional action, and it is to actions of this sort to which "action" should be taken to refer.

<sup>57</sup> Ibid, 191.

<sup>58</sup> Ibid, 178.

Second, an agent might be “informationally cut off” from autonomous deliberation (and hence autonomous action). Mele illustrates this possibility with the case of a king, George,

whose only access to the state of his kingdom is through his staff of advisors. George’s primary concern in life is to do what is best for his kingdom as a whole. The staff provide him with information, and he takes legislative and other measures on that basis. They also provide George with feedback about his measures, which information he takes into account in drafting further legislation and in modifying previous measures. ... However, the staff have their own agenda, namely, their own enrichment at the expense of the populace. Knowing the king’s preferences, they systematically provide him with such false information as will lead him to make decisions that will further their aims. ... George plainly is not ruling autonomously. Hence, he is not autonomous in the sphere of his life that is most important to him.<sup>59</sup>

George is, Mele proposes, “informationally cut off” from autonomous action in this area of his life: “A sufficient condition of [an agent] *S*’s being informationally cut off from autonomous action in a domain in which *S* has intrinsic pro-attitudes is that *S* has no control over the success of his efforts to achieve his ends in that domain, owing to his informational condition.”<sup>60</sup> An agent lacks control over the success of his efforts to *A*, Mele continues, “If *S* is unable to construct a plan for *A*-ing such that his flawlessly executing that plan would increase the likelihood that he will *A* above that provided by mere chance and if *S*’s *A*-ing intentionally is contingent upon his constructing a plan for *A*-ing.”<sup>61</sup>

Finally, autonomous deliberation can also be denied by a lack of the “executive qualities,” or “skills, capacities, and habits,” necessary for reliable deliberation.<sup>62</sup> An agent whose deliberative practices encompass “unreliable cognitive procedures (e.g. procedures driven by the anchoring effect or confirmation bias)” might thus lack autonomy with respect to her deliberative processes and, hence, with respect to the intentions and actions in which those processes result.<sup>63</sup>

An agent who is able to engage in informed and reliable deliberation on the practical implications of an autonomous pro-attitude is able to engage in *autonomous* deliberation on the practical implications of that pro-attitude. An agent who is able to execute the intention in which such deliberation concludes is capable of autonomously acting on that pro-attitude. Although Mele has not to my knowledge discussed the relation between his earlier I-autonomy condition and his later account of the conditions necessary for autonomous action, it seems that the two are not equivalent, for an agent

---

<sup>59</sup> Ibid, 180.

<sup>60</sup> Ibid, 181.

<sup>61</sup> Ibid.

<sup>62</sup> Ibid, 182-184.

<sup>63</sup> Ibid, 184.

might satisfy the former, but not the latter. This is in fact true of King George, who possesses I-autonomy with respect to his desire to improve his kingdom, but according to Mele is incapable of autonomously acting on that desire. Since George is obviously capable of acting on his desire to improve his kingdom, and presumably capable of refraining from acting on that desire as well (we have no reason to think that this is an irresistible desire of George's), he likely would satisfy the first disjunct of Mele's I-autonomy condition. Regardless of whether that is true, he would certainly satisfy the second disjunct, for even if he lacks the power to refrain from acting that desire, he holds no rational judgment regarding what it would be best to do that is thwarted by his lacking that power – George is doing precisely what he rationally judges it best to be do with regard to his desire to improve his kingdom. According to Mele's suggested I-autonomy condition, George therefore is autonomous with respect to the influence that desire has on his behavior. Yet he is on Mele's account of autonomous action informationally cut off from ruling autonomously.

While I-autonomy with respect to a pro-attitude thus does not suffice for the capacity to autonomously act on that pro-attitude, the latter would appear to entail the former – it seems that an agent who possesses the capacity to autonomously act on a pro-attitude would necessarily possess I-autonomy with respect to that pro-attitude as well. If an agent who is able to autonomously act on a pro-attitude is also able to refrain (whether autonomously or not) from acting on that pro-attitude, then she possesses I-autonomy with respect to that pro-attitude. If she is not able to refrain from acting on that pro-attitude, then she will nevertheless possess I-autonomy with respect to it so long as she holds no rational judgment concerning what it would be best to do regarding action on the basis of that pro-attitude that is thwarted by her lacking that power (i.e. the power to refrain from acting on the pro-attitude). Presumably, this would have to be a rational judgment that it would be best not to do what she is unable to refrain from doing, such as an unwilling addict's judgment that she ought not take the drug for which she has an irresistible desire. However, it would seem that an individual could not hold a rational judgment that it would be best not to do some (deliberative, intentional) action *A* and at the same time be capable of autonomously intending to do *A*, which is a necessary condition of her being capable of autonomously doing *A* on Mele's account.<sup>64</sup> In order to be capable of autonomously intending to *A*, an agent must on Mele's account be capable of making an autonomous decisive evaluative judgment that it would be best to *A* now. In order for such a judgment to be autonomously made, the deliberation on which it is based must be autonomously conducted. And in order for deliberation to be autonomously conducted, it must be driven by the agent's adequate executive capacities – that is, adequate mental skills, abilities and habits. Although Mele does not elaborate on the necessary capacities, they

---

<sup>64</sup> Ibid, 191.

presumably include or imply some meaningful degree of rationality. Hence, it is reasonable to think that in order for an agent to be capable of autonomously intending to *A*, she must be capable of making a *rational* decisive evaluative judgment that it would be best to *A* now. Presumably, an agent capable of making such a judgment could not at the same time hold a rational judgment that it would be best *not* to do *A* now – could not make a rational judgment that it would be best to *A* while holding a rational judgment that it would be best not to *A*. If this is correct, then an agent possessing the capacity to autonomously *A* could not also hold a rational judgment that it would be best not to *A*, and hence could not lack I-autonomy with respect to the desire to *A*. The requirements of the capacity to autonomously act on a pro-attitude thus appear to be more demanding than those of I-autonomy with respect to a pro-attitude, which they entail and extend beyond.

Again, Mele has not, so far as I am aware, indicated how he takes his account of the capacity for autonomous action on a pro-attitude to relate to his earlier I-autonomy with respect to a pro-attitude. Nevertheless, we can, I think, safely interpret him as regarding the former as the behavioral aspect of personal autonomy with which we should mainly be concerned. There would appear to be two ways in which these notions might be related: the capacity to autonomously act on a pro-attitude might either supplant I-autonomy with respect to a pro-attitude as the sole behavioral aspect of autonomy (relative to pro-attitude), or it might instead join I-autonomy as a second, and separate, behavioral aspect of autonomy. The fact that, despite its obvious relation to the capacity for autonomous action, I-autonomy receives no mention in Mele's positive account, would seem to suggest the former reading – would seem to suggest that the I-autonomy condition is no longer in play and that the capacity to autonomously act exhausts the behavioral aspect of autonomy with respect to one's pro-attitudes on his account. However, even if this is incorrect and Mele instead takes I-autonomy to remain a relevant behavioral aspect of autonomy, it is clearly in his eyes the less important such aspect. Mele tells us that because George is not capable of autonomously acting on his desire to improve his kingdom he is “not autonomous in the sphere of his life that is most important to him.”<sup>65</sup> The I-autonomy with respect to that desire that George *does* possess thus appears to be of little importance. On Mele's account, it is instead the capacity to autonomously act on a desire that is necessary for autonomy in a domain in which an agent holds intrinsic pro-attitudes.<sup>66</sup> Since the capacity to autonomously act on a pro-attitude appears to entail I-autonomy with respect to that pro-attitude, and since it is the former, more demanding notion that is necessary for autonomy in some area of an agent's life, it would seem to be that notion with which we should be concerned.

---

<sup>65</sup> Ibid, 180.

<sup>66</sup> Ibid, 180-181.



We therefore have good interpretive reason to focus our attention on Mele's more recent account of the requirements of autonomous action on a pro-attitude, rather than his earlier I-autonomy condition. More importantly, we have good theoretical reason to do so as well, for Mele's account of the requirements of autonomous action improves upon his earlier condition in an important way, and consequently represents a more promising approach to the behavioral aspect of autonomy relative to a pro-attitude. The improvement consists of the fact that, whereas Mele's I-autonomy condition required only that the agent possess certain abilities or judgments, his account of the requirements of autonomous action requires also that the agent be autonomous with respect to the psychological processes underlying the relevant abilities and judgments. George's inability to rule autonomously is, after all, not the result of any lack of ability on his part or the presence of a judgment regarding what he should do that he is unable to implement. Again, George's actions are entirely in line with his relevant judgments. What renders George incapable of ruling autonomously, according to Mele, is instead the fact that his judgments are the result of non-autonomous deliberation. Although I disagree with Mele on this particular case, and will argue below that George is capable of ruling autonomously, it is clearly an advantage of Mele's later treatment of the behavioral aspect of autonomy that it addresses the psychological processes giving rise to the agent's actions and judgments. For it is easy to imagine cases in which an agent possesses judgments of the sort required by Mele's original I-autonomy condition, but lacks autonomy with respect to the psychological processes giving rise to those judgments and hence intuitively lacks autonomy in that area of her life. An agent might be incapable of refraining from acting on some pro-attitude and lack any rational judgment that is thwarted by that inability, but lack such judgments only because her "executive capacity" to form rational judgments regarding what it would be best to do has been destroyed or circumvented. For instance, Beth holds no rational judgments that it would be best not to act on her desires for heroin, but this is because her addiction has removed both her disposition and her ability to engage in rational self-reflection. That she holds no rational judgments that are thwarted by the irresistibility of her desire should not in this case be taken to mean that she has autonomous control over the effect that desire has on her behavior. What this shows is that although personal autonomy does not consist entirely in psychological autonomy – but must include as well some condition that addresses an agent's ability to act on her pro-attitudes – that condition cannot ignore the psychological processes on the basis of which she does or does not employ that ability. The behavioral aspect of autonomy relative to a pro-attitude requires not merely the ability to act or not act on that pro-attitude, but the ability to do so on the basis of autonomous psychological processes.

Mele is thus in my view correct to think that the behavioral aspect of autonomy requires not only abilities or judgments of the sort identified by his original I-autonomy condition, but also autonomy with respect to the

processes giving rise to those abilities or judgments. However, I believe that Mele's account of the requirements of autonomous action, and more specifically autonomous deliberation, are in need of revision at one point.

On Mele's account, an agent is capable of autonomously acting on the basis of some pro-attitude if she is able to form and execute an intention to act on that pro-attitude which is the result of autonomous deliberation. An agent can be blocked from autonomous deliberation, recall, in three ways: through a lack of autonomous pro-attitudes, through a lack of necessary information and through a lack of necessary executive qualities. In its broad outlines, this seems to me a promising approach to the capacity for autonomous action on some pro-attitude. As I have said, such action clearly requires not only the ability to translate one's autonomous pro-attitudes into certain behaviors, but the ability to do so on the basis of autonomous psychological processes, and more specifically, autonomous deliberation. And as Mele argues, such deliberation clearly requires not only the possession of certain executive qualities – mental skills, dispositions, etc. – but also some supply of information. Without information regarding the practical implications of her pro-attitude and the details of her circumstances, an agent will be unable to employ her relevant self-control capacities in translating that pro-attitude into action. The informational condition Mele proposes, however, requires not merely that the agent has access to information enabling her to act on the basis of her pro-attitudes, but that she has access to information which enables her to act with some level of success, and as such, I will argue, confuses the ability to autonomously pursue one's ends with the ability to achieve them.

Recall that, according to Mele, an agent is "informationally cut off" from autonomous action in some area of her life if, owing to her informational condition, she has no control over the success of her efforts to achieve her ends in that area of her life.<sup>67</sup> And an agent lacks control over the success of her efforts to perform some action *A* if she "is unable to construct a plan for *A*-ing such that [her] flawlessly executing that plan would increase the likelihood that [she] will *A* above that provided by mere chance and if [her] *A*-ing intentionally is contingent upon his constructing a plan for *A*-ing."<sup>68</sup> This approach to the informational aspect of autonomous agency seems to me misguided, for it mistakenly associates autonomous agency with effective

---

<sup>67</sup> Ibid, 181.

<sup>68</sup> Ibid. Although I will not pursue the matter too far here, as it is irrelevant once we abandon this success-based approach to the behavioral aspect of autonomy, which I will do below, it is not clear how we are to identify the baseline for success against which we are to measure plans for *A*-ing. Mele's discussion suggests that the relevant baseline is the success of some attempt to *A*. This may not be problematic in the gambling cases Mele uses by way of illustration, in which there is typically only one way in which to make an attempt (e.g. by buying a lottery ticket or placing chips on the roulette table) and success is clearly determined by chance. In many other cases, however, things are not so simple. There are many ways in which one might attempt to impress a particular member of the other sex, for example, and it is not clear which of these gives us the likelihood of success provided by mere chance.

agency. Mele tells us that “a distinction is in order between having control over whether one acts with the goal of *A*-ing and having control over whether, given that one acts with the goal of *A*-ing (i.e., given that one attempts to *A*) one succeeds in *A*-ing.”<sup>69</sup> I agree that such a distinction is necessary, but in my view it is the former that is relevant to the behavioral aspect of autonomy relative to a pro-attitude. To be an autonomous agent is to be a self-directed agent, and this requires only that one has some meaningful degree of control over one’s life, including whether – and I would suggest *how* – one pursues one’s ends, and not necessarily that one has control over the success of such pursuits. Consider, for instance, a 14<sup>th</sup> century alchemist’s desire to turn lead into gold. Probability, in the form of quantum mechanics, provides some infinitesimal chance that his attempts to transform lead into gold will in fact succeed, or at least let us so assume. Given the state of scientific knowledge and methods at that time, his informational condition is such that he will be unable to construct any plan that would improve upon those meager odds. According to Mele’s condition, he is thus informationally cut off from autonomous action in that area of his life; he is not autonomous with respect to his alchemic pursuits. This, though, seems wrong. We can suppose that the alchemist employed his self-control capacities not only in the development and maintenance of his transmutation desire, but subsequently in gathering and rationally reflecting upon the relevant information which was available to him, in reasoning, on the basis of that information, about how best to go about acting on the basis of that desire, and in constructing and carrying out some experimental plan for doing so. Such an agent is in my view significantly self-directed with respect to that sphere of his life, for he has played a meaningful role both in determining the end he now seeks and in determining whether and how to go about doing so. He is, I submit, a substantially autonomous agent in that aspect of his life. The fact that Mele’s informational condition denies this represents a reason to reject that condition.

It is better, I believe, to have the informational aspect of the capacity for autonomous action require merely that the agent has access to information from which what she does (or would) rationally take to be a promising and viable plan for action on the basis of a pro-attitude could be constructed. A few explanatory remarks are in order. First, an agent has access to information if her circumstances are such that her information-acquisition abilities are sufficient for the collection and consideration of that information.<sup>70</sup> Such information enables an agent to construct a plan if her information-retention and -processing abilities are sufficient for the formation of that plan. The precise nature of the necessary abilities will thus vary with the agent’s cir-

---

<sup>69</sup> Ibid, 180.

<sup>70</sup> I say that access requires consideration as well as collection because an agent might collect and possess information, but be incapable of bringing it to mind, in which case it would not of course enable construction of a plan for action on the basis of a pro-attitude.

cumstances as well as with the pro-attitude in question. Note that the requirement here is not that the agent in fact possesses information from which a promising and viable plan could be constructed, but merely that she has access to such information. Suppose that King George lacks the information necessary to construct a promising plan for acting on his desire to improve his kingdom simply because he has neglected to review the reports he has available. In this case, George does not know how to go about acting on the basis of his desire to improve his kingdom, but he has not been denied the capacity to rule autonomously. Assuming that George's failure to review the relevant reports was not the result of coercion, manipulation or the lack of some relevant self-control capacity – he had ample opportunity to act on his desire to rule well. This opportunity to employ one's relevant self-control capacities in determining whether and how to act on the basis of one's autonomous pro-attitudes is the essence of autonomous agency and, in any particular instance, it may or may not issue in an action, depending on whether or not the agent takes advantage of the opportunity. I propose that the informational aspect of such opportunity requires merely that the agent has access to the necessary information, which in this adjusted version of the case, King George clearly does.

Second, an agent does (or would) rationally take a plan for action on the basis of a pro-attitude to be promising if she does (or would) judge it to be a worthwhile way of pursuing that pro-attitude, and that judgment is not (or would not be) the result of irrationality. This need not be an all-things-considered judgment that she should in fact pursue that plan. What matters is simply whether the agent does (or would) rationally take the plan to be recommended by the pro-attitude in question, not whether it is recommended, all things considered. If an agent has access to information from which she can construct a plan that she rationally judges to be a worthwhile means for pursuing one of her pro-attitudes (and that she regards as viable, as explained in the following paragraph), but her circumstances are such that she rationally judges it best not to pursue that pro-attitude at present, her informational condition has not denied her the opportunity to autonomously act on the basis of that pro-attitude. I will not attempt to specify the relevant standard of rationality, but simply stipulate that, as with all rationality requirements on autonomy, it must not be overly demanding. Autonomy does not require perfect rationality, or anything approaching it, but merely some minimal threshold of rationality. An agent who judges some plan to be promising, but does so in a way that would fail even a (plausibly specified) minimal rationality requirement, lacks the sort of control necessary for autonomy. For instance, if Bob takes walking into the Atlantic Ocean to represent a promising means of renting a movie, he may control his behavior in some sense, but clearly not the sense relevant to personal autonomy.

Finally, if a plan is to provide a meaningful opportunity to act on the basis of a pro-attitude, it must be one the construction and execution of which the agent regards as viable. This requires both that the agent takes herself to

have access to the resources necessary for execution of the plan and that she can bring herself to in fact construct and execute it. An agent who is able to construct a plan that is promising in principle, but appears to her impossible to put into practice, lacks a meaningful opportunity to act on the basis of that pro-attitude. Imagine that a woman who loves to paint has been captured by a serial kidnapper who imprisons and psychologically torments his victims. To this end, he places in her cell a locked glass case containing a large supply of paints, brushes and canvasses. She knows enough about the hardness of glass, and about her own strength, to know that with a decent-sized hard object, she would be able to break the case, get at the painting supplies and put her desire to paint into action. That is, she possesses (and is able to recall and appropriately process) information from which what she would rationally take to be a promising plan for acting on that pro-attitude could be constructed. However, as there appear to be no hard objects in her cell, and no way out of it, she takes herself to lack a resource essential to execution of that plan.<sup>71</sup> Indeed, it may be that the impossibility of executing that plan precludes its ever occurring to her – given the lack of any object with which she might break the case, there is no reason for the relevant information to be recalled. In this case, that plan clearly does not provide her with an opportunity for action. The same is true if, although she takes herself to have access to the resources necessary for execution of a plan for action, she finds execution of the plan unthinkable. Suppose that the woman’s captor convinces her that she can have the key to the case if she pushes a button that will result in the death of the woman in the neighboring cell. In this case, she has access to the necessary resources – namely, the button. On the assumption that she is a morally decent person for whom pushing the button is not a viable possibility, however, she nevertheless lacks the capacity to autonomously act on that desire. The option of pushing the button will be dismissed out of hand, and thus does not provide her with a meaningful opportunity to employ her self-control capacities in determining whether and how to translate her desire into action.<sup>72</sup> In order for a plan to enable autonomous action, it must be one the agent regards as viable.

The condition I have proposed differs from Mele’s in that the necessary information need not enable construction of a plan with some objectively determined likelihood of success, but only one the agent does (or would) rationally judge to have some promise, one she does (or would) rationally deem worthy of pursuit. This condition the alchemist satisfies, for while from our perspective we see that his informational situation renders effective action on

---

<sup>71</sup> This is true even if, unbeknownst to her, one of bricks in the wall of her cell could be removed and used for this purpose – what matters is that she takes herself to lack an essential resource.

<sup>72</sup> This would not be true of another painter who could bring herself to cause the death of her neighboring captive. Such a painter would not lack the capacity to autonomously act on her desire to paint, for she would have a meaningful opportunity to employ her self-control capacities in determining whether and how to act on the basis of that desire.

the basis of his desire to turn lead into gold (virtually) impossible, it does not appear so to him. He takes himself to have a worthwhile opportunity to put that desire into action and, given the body of knowledge available to him, does so in an at least minimally rational manner. The alchemist is thus able to employ his adequate self-control capacities in determining whether and how to go about acting on that desire. The improbability of successfully satisfying a desire need not preclude the autonomous construction and execution of a plan for acting on it.

Note, however, that Mele's King George satisfies the informational condition I have suggested as well. George has access to information from which what he takes to be a promising plan for acting on his desire to benefit his kingdom could be, and indeed is, constructed. Because that information was false, and although he is led to believe the contrary, George's plan has of course failed to achieve his goal. As the alchemist case shows us, however, an inability to achieve one's ends must not be taken to preclude their autonomous pursuit. Both George and the alchemist have been prevented from satisfying their desires due to the falsity of the information available to them. In neither case, however, has the falsity of that information in any way prevented them from exercising their adequate self-control capacities in putting those desires into action. Both George and the alchemist have had, and have taken advantage of, a meaningful opportunity to determine for themselves whether and how to go about translating their autonomous goals into action, and as I have said, such opportunity is the essence of the behavioral aspect of personal autonomy.

Indeed, the only significant difference between the two cases appears to be that, whereas the explanation of the alchemist's possessing false information includes no immorality, George's possession of false information is the result of his advisors having intentionally deceived him. What matters to personal autonomy, however, is simply whether or not an agent is able to engage in meaningful self-direction. The moral legitimacy of the reasons why he is or is not so able are irrelevant. Mele would agree: he recognizes that autonomy can be denied by impersonal, and hence morally neutral, forces as well as intentional interference. If an agent's values are erased and replaced by "strange, randomly occurring electromagnetic fields at the center of the Bermuda Triangle," then according to Mele, "her capacities for control over her mental life are bypassed by the value-changing mechanism or process, but the bypassing is utterly nonintentional."<sup>73</sup> Such an agent is every bit as non-autonomous with respect to the values with which she has been implanted as she would be were she to undergo the same transformation at the hands of an intentionally manipulative "team of new-wave brainwashers."<sup>74</sup> Mele is correct to in this way think that, in questions of developmental autonomy, what matters is simply whether or not the agent was able to play a meaning-

---

<sup>73</sup> Mele, *Autonomous Agents*, 168.

<sup>74</sup> *Ibid.*, 145.

ful role in the development of a pro-attitude. I would, though, recommend a similar response to questions of the capacity for autonomous action: what matters should simply be whether or not an agent is able to exercise her relevant and adequate self-control capacities in acting on the basis of her pro-attitudes, and not why she is or is not so able. Both George and the alchemist do this successfully. Both have access to information from which what they take to be promising plans could be constructed and both have utilized their adequate information-acquisition and -processing, and intention-formation and -execution capacities in rationally constructing and executing such a plan. Both have therefore determined for themselves what action their desires recommend in their present circumstances and have successfully executed those actions. This form of self-direction is all that the capacity for autonomous action should be made to require. It is thus not autonomy that George's life lacks, but success, and these two notions should not be confused.

## **Conclusion**

Although I have not defended this claim here, an agent's autonomy in some area of her life does in my view depend in part upon her history – upon the way in which her pro-attitudes in that area of life were developed. It does not, however, entirely so depend, for as Mele has argued, the development of autonomous pro-attitudes in some aspect of one's life does not suffice for full or “robust” autonomy with respect to that aspect of one's life.<sup>75</sup> An agent is more robustly autonomous with respect to some area of her life if she not only developed her pro-attitudes in that area of life in an autonomous manner, but also maintains them in an autonomous manner and is capable of autonomously acting on them. For such an agent will not only have had control over the way in which she originally came to have those pro-attitudes, but will retain control over her continued possession of them, and over the influence they have on her behavior. These non-developmental aspects of autonomy have not, however, been adequately addressed by recent historical accounts of autonomy. I have in this paper attempted to remedy this by critiquing and improving upon, first, the autonomous-maintenance conditions suggested by Mele and by Arneson, and second, Mele's two separate discussions of the behavioral aspect of autonomy with respect to a pro-attitude. Although the “autonomous-maintenance” and “capacity for autonomous action” conditions I have proposed are admittedly still quite rough, I believe that they better “illuminate” the ongoing aspects of autonomy than

---

<sup>75</sup> Another topic in need of more thorough treatment than it has hitherto received concerns how an agent's autonomy with respect to particular areas of her life relates to her overall autonomy. It is plausible to think that the extent to which an agent is autonomous with respect to more central or important areas of her life (e.g. her religious values) matters more to her personal autonomy than does the extent to which she is autonomous with respect to more peripheral or trivial areas of her life (e.g. how she likes her eggs), but this idea needs to be spelled out in some way.

did the conditions I have built upon, and that a theory of autonomy which includes those conditions alongside historical conditions on the development of autonomous pro-attitudes will for that reason better capture the “essence of self-government.”<sup>76\*</sup>

Steven Weimer  
Bowling Green State University  
Department of Philosophy  
[sweimer@bgsu.edu](mailto:sweimer@bgsu.edu)

---

<sup>76</sup> Mele, “History and Personal Autonomy,” 279.

\* For helpful comments on this paper, I thank an anonymous reviewer for the *Journal of Ethics and Social Philosophy*.