Chapter 4     Identification, Integrity, and the Authentic Development of the Person

Introduction

The preceding chapters have tracked and developed a movement in luck-egalitarian debates over the basis for consequential responsibility from arguments about the importance of genuine choice towards a recognition of the relevance of an agent’s identification with a preference. One school of thought argues that where an agent identifies with a preference, he sees it as part of his personality, and regards it as his own responsibility. He ought therefore to be held responsible for its cost. The other draws the opposite conclusion. An agent’s strong identification with a preference provides a reason not to hold him responsible for its cost. Identification with a preference demonstrates the importance of that preference to its bearer. Moreover, since it is unlikely to be on account of the cost that the person identifies with a preference, it is a matter of compensable bad luck if the preference happens to be expensive.

Each of these approaches, though, runs into difficulties highlighted in Chapter 2. Judgements of consequential responsibility need a solid basis, but there are significant barriers to an agent’s identification with a preference providing that basis. Identification might lack the normative significance to ground consequential responsibility. I might identify with a preference that nevertheless has no great importance to me as a person. I might see as part of my personality my preference for the sound of strings to that of woodwind but have no interest in classical music. In these cases, it is difficult to see why the presence or absence of identification should determine judgements of responsibility for the cost of fulfilling the preference.

A parallel difficulty arises from the fact that identification does not necessarily involve approval. I might identify with a preference, in the sense of seeing it as part of my personality, for example, due to fatigue or depression. I might not welcome the preference, but not have the energy or will-power to do anything about it, and acquiescently accept that it is part of who I am. In these cases, as well, it is unconvincing for identification to determine consequential responsibility.

As revealed in Chapter 3, one response to these problems is to refine the notion of identification: ‘acquiescent’ identification can be excluded if the idea of approval is incorporated into the notion of identification; concerns that identification with a preference might not indicate that the preference has a significance to its bearer can be assuaged if the degree of approval is strengthened.
One route, pursued by Cohen, is for identification to reflect a person’s judgemental evaluation. Requiring identification to be based on judgement is intended to increase the strength of the approval that identification involves – the result of evaluation rather than simply liking. It also roots the approval in parts of the person that are normatively significant – evaluation is made in accordance with the agent’s other preferences and convictions.

However, even identification based on judgement does not guarantee that the preference is important to the person. My preference for strings over woodwind might be based on my judgement of the superior technical mastery required to play strings, but it is still the case that classical music doesn’t matter to me. Equal opportunity for integrity thus goes a step further by requiring that identification is based on judgement and that the identification indicates the preference plays a significant role in its bearer’s personality.

But even if identification is understood to incorporate these extra features, it remains open to distortion and manipulation. This was the lesson from the case of Robert Harris considered in chapter 2. Harris fully identified with his preferences for inflicting suffering on others. They reflected his judgement and mattered to him significantly as a person. We would not think that the presence of such strong identification settles the question of whether he is consequentially responsible, however. The distortion to his personality engendered by the extreme abuse and dysfunction he suffered at the hands of his family and the judicial system are also relevant. It is only plausible to hold that an agent’s identification with a preference ought to determine consequential responsibility if his personality is formed authentically.

What it means for a personality to be formed authentically is itself subject to variation and controversy. The core idea is that the process of formation is not distorted, not compromised by alien forces. How this is spelt out, though, is a matter of dispute. Dworkin’s account of equality of resources, which deliberately eschews metaphysical bases, makes authenticity a question of a person’s satisfaction with the formation of his personality. Equal opportunity for integrity argues that authenticity is a question of certain preferences – judgemental tastes - being formed in accordance with judgement rather than by processes that override it.

The first of these approaches suffers from the difficulties facing identification. Satisfaction with the formation of one’s personality can be manipulated or distorted. Indoctrinated cult members, for example, report satisfaction with the process by which their personality was formed even though it would not be considered authentic from any independent stand-point. The latter approach locates authenticity in the mechanism by which preferences are formed. This route
avoids the problems of the first person measure of satisfaction, but requires an account of the broadly metaphysical or factual matter of the authentic and inauthentic formation and development of the self.

Luck-egalitarians have tapped into the intuitive appeal of linking consequential responsibility to identification and authenticity without developing robust accounts of either notion. This undermines the plausibility of employing these notions and runs the risk that the notions, when better understood, do not in fact support the conclusions drawn from them. This chapter seeks to remedy this situation. Drawing ideas from contemporary work in the philosophy of mind and of action I will outline an account of the structure of the self which elucidates the notions of identification and the authentic formation of personality\(^1\). It will also clarify which parts of the person are integral, as required by equal opportunity for integrity, and the crucial question of the degree to which the process of preference-formation is a matter of compensable luck.

Although an exhaustive investigation is not here possible, the aim is to develop stand-alone accounts of identification, authenticity and integrity. The problems facing the conceptions of these notions as employed in the political debates over consequential responsibility will set the basic parameters of the investigation, but the conclusions reached will be independent of any political usage of the notions under discussion. This work thus prepares the way for the next chapter, which asks how robust accounts of identification, authenticity and integrity impact on the different accounts of consequential responsibility under consideration in this thesis.

Those parameters will narrow the focus of the investigation somewhat. Although identification can be considered simply a matter of recognition that a preference is part of one’s personality, I will stipulate on the basis of the discussion of previous chapters that it should incorporate a degree of approval\(^2\). I will also take the assumption from the political debates that the authenticity of a personality is not simply a matter of the elements of an agent’s personality being genuinely attributable to her as she now is. Rather it is also a matter of how her personality was formed, how she came to have the personality she now has. These assumptions, I hold, are justifiable independently of debates about consequential responsibility\(^3\). I will not,


however, attempt to justify them here, but rather use them as starting points to clarify the notions that form the focus of this chapter.

Section I will start with the question of what identification with a preference might mean, and how that identification gains its significance. This significance, I argue, comes from being rooted in other parts of the self, parts that play a more important role in an agent’s identity. Section II will examine two types of preference that might form the basis for the identification with others. Section III will look at how these ‘grounding’ preferences themselves attain the authority to represent their bearer. The explanation, I hold, lies in both the integral role they play in an agent’s identity, and the way in which they were formed. Section IV will then look at the role of luck in the authentic development of personality.

Note that the focus of this chapter will be on the internal rather than the external conditions of the authentic development of personality. That is, authenticity is considered in terms of the relations between different elements of personality and its internal mechanisms of change. What is not addressed is the impact of the external environment, the social conditions in which a personality is formed, for example. I will return to that issue in the next chapter, when considering how external conditions - in particular the demand to bear the costs of one’s preferences - impact on the authentic development of personality.

Section I Identification and Authenticity

Step one is to look more closely at what it means for someone to identify with a preference. The core idea under examination is that identification with a preference is, at base, a matter of being welcomed or approved of. Seminally, Harry Frankfurt has highlighted the importance of a person’s ability to step back from the desires that he feels pulling upon him at any one time to reflect upon their worth, forming ‘higher-order’ attitudes towards those ‘lower-order’ desires, and thereby taking a stance on whether he wants a lower-order desire to motivate his action. This process of reflective evaluation enables a person to form desires about desires. If he is then able to act on the desires that he wants to have and wants to be operative, then he achieves an important degree of control over his mental life and his actions in the external world⁴.


⁴ “Freedom of the Will and the Concept of the Person”. First published in Journal of Philosophy 68 (1971), 5-20, reprinted in The Importance of What We Care About. Frankfurt discusses identification with desires rather than preferences per se.
The classic example is of someone who has an unwelcome desire of some sort, such as a rogue drive to watch someone fall from a skyscraper, which, although in a straightforward sense is his, is yet a desire that is unbidden and unwelcome. Instead of simply acting to fulfil the desire, he reflectively evaluates it and forms a second order desire that he does not want the unbidden desire to determine what he does. If he then nevertheless acts upon the first-order desire he is not the master of his will. If he succeeds in blocking the desire, though, and his higher-level desire is operative, then he acts autonomously, as the author of his acts, rather than just as a receptacle driven by the forces of his desires.

Now, Frankfurt is discussing autonomous action, an agent’s ability to evaluate the first-order desires that operate upon him, and act in accordance with those that he approves of, rather than an agent’s approval of the elements of his personality. His focus is on approval of what a person does rather than who he is. Nevertheless, Frankfurt’s reflective evaluation is equally relevant to the latter. People are able to detach themselves from the pull of preferences, reflect upon them, and form higher-order attitudes that have as their object lower-order preferences. Should the agent approve of a lower-order preference, then it ceases to be a preference he just happens to have, and becomes a preference he identifies with.

This is the basic idea behind the notion of identification. But it is subject to controversy. An important objection, initially raised by Gary Watson, is that the mere fact that a preference is second-order rather than first does not guarantee that it is vested with agential authority. The influence of peer-pressure, or advertising, for example, can mean that second-order appraisals of first-order motives can reveal no less authentic aspects of the self than the first-order motive itself, and undermine identification rather than secure it. This is precisely what damages the political usage of identification by Dworkin and Cohen.

Another problem is that the requirement for approval at a higher level does not tell us anything about the nature that the approval must take. We do not yet know whether the approval can be weak, whimsical even, or whether it must be more strongly grounded, in judgement for example. In 'Freedom of the Will and the Concept of the Person’, Frankfurt argued, strikingly, that anything goes. “... a person may be capricious and irresponsible in forming his second-order volitions and

---

give no serious consideration to what is at stake... There is no essential restriction on the kind of basis, if any, upon which they are formed.\footnote{p. 19 note 6. ‘Volitions’ are operative second-order desires.}

The lack of restriction on the grounding of approval of desires captures an important aspect of autonomous action. We can sometimes confer approval on our desires on no more basis than that we feel like it, without this meaning that they are not acts that we author ourselves. As chapters 1 and 2 revealed, however, when it comes to ongoing aspects of personality, such as a preference with repeated relevance to a person’s motivation, such minimal grounding looks too flimsy for identification to be attributed with normative significance.

The problems with the grounding of identification are heightened because, as is familiar, Frankfurt’s hierarchical approach either leads to an infinite regress or requires an arbitrary cut-off point. What makes a first-level desire identified with is an agent’s positive approval of it, in the form of a second-order desire for that desire. But then approval of the second-order desire is also required to render it authentically the agent’s own, so there must be a third-order approving state, and in order for that third-level approving state to be the agent’s own it must be endorsed by a fourth-level approving state, and so on. Since we cannot simply equate desires of any particular level with the agent’s authentic self, we are forced into an unpalatable dilemma. Either the need for each level of approval to be itself endorsed leads into an infinite regress, or we attribute a particular level with authority to represent the agent, leading to accusations of arbitrariness\footnote{Watson, op cit.}.

These observations reveal two important points about models that use the ideas of reflective evaluation and hierarchies of attitudes to convey authority. Firstly, to plausibly ground meaningful identification, evaluations cannot simply be formed on any basis whatsoever. Rather the evaluation needs to be rooted in something more significant, something that represents more of the agent than passing whims about which desire they fancy endorsing. Second is the closely related problem that whatever it is that provides the basis on which a desire is approved of itself requires the authority to authentically represent the agent. What is needed is a basis on which assessment of desires or preferences must take place, and reasons to accept that that basis authoritatively represents the agent\footnote{Frankfurt in fact offers, over the years, a number of possible solutions to these problems. The concerns of this chapter differ from those of Frankfurt, however, so his solutions will only be considered where relevant to the establishment of authoritative approval.}. 

\footnote{7}
The Libertarian Response

One means to escape the problems of arbitrariness and regress is to follow the path pursued by Kane and locate agential authority in the act of deciding. By deciding to act in one way rather than another, the agent aligns himself with a vision of the person he wants to be, thereby conferring upon it the stamp of authentic agency.

Chapter One, though, established that decisions of this kind themselves suffered from a damaging arbitrariness. Metaphysical libertarian free choice requires independence from an agent’s inherited character. But if a decision is not sufficiently rooted in the agent’s ongoing desires, plans and projects, then it can be as arbitrary a representative of the agent’s ‘real’ self, as can any level of the hierarchy of desires.

The task then is to establish responses to the two problems highlighted – what kind of grounding approval requires to make identification authoritative, and how these grounds of approval are themselves authoritative – that do not appeal to an independent, libertarian, ‘free’ will.

Section II Caring, Love, and Self-Governing Policies

Frankfurt himself, in later work, provides a potential candidate for the element of personality that might play the ‘grounding’ role necessary to bolster identification: caring. Although Frankfurt does not himself suggest that caring might ground identification – he does not think that identification needs grounding – his exploration of the notion raises interesting possibilities. The idea would be that identification that is grounded in an agent’s cares ought to be considered authoritative. I identify with my preference for left-leaning government, for example, and the identification is genuine because it is rooted in my cares about social justice.

Cares could also be those parts of personality that are most central, that must be safeguarded if we are to respect integrity. Establishing this kind of centrality is what motivates Frankfurt’s discussion of caring. He holds that what a person cares about serves as a reference by which the agent guides himself in his life and conduct. To care about something is to be invested in it. It is more than to simply desire something, and different from valuing it.

---

10 See in particular his “The Importance of What We Care About”, in The Importance of What We Care About, op cit.
11 Ibid. p. 82.
[Caring] differs from having feelings of a certain type, such as those of powerful attraction or of intense desire or of compelling delight. It is also not equivalent to or entailed by any judgement of appreciation of the inherent value of its object. That a person cares about something has less to do with how things make him feel, or with his opinions about them, than with the more or less stable motivational structures that shape his preferences and that guide and limit his conduct. \(^{12}\)

Cares appear good candidates for preferences that are integral to a person because they are inherently persistent and forward-looking\(^{13}\): “the notion of caring, implies a certain consistency or steadiness of behaviour; and this presupposes some degree of persistence”\(^{14}\). Moreover, caring is paradigmatically personal\(^{15}\). Cares reflect a person’s specific concerns, distinguishing him as an individual.

It is also worth noting that cares are not simply abstract judgments of worth. It is possible to judge something valuable, yet not care about it. This combats a weakness of Cohen’s portrayal of judgemental tastes. The presence of judgement does not in itself render a taste significant to identity. A care, in contrast, might incorporate a judgement of value, but adds the crucial element of motivational effectiveness, so that it necessarily plays a role in a person’s motivational structures.

Not all cares, though, are crucial to integrity. We certainly have cares that define us as individuals, but we also have cares that are not central to our identity. I care about the survival of my local pub, which is threatened with closure, but this care has no great bearing on my shape as a person. Shedding it would not threaten my integrity. To combat such worries, Frankfurt appeals to a subset of cares, which he holds to most centrally constitute our identities. These are the cares that we cannot help caring about, which Frankfurt calls ‘volitional necessities’:

The essence of a person is a matter of the contingent volitional necessities by which the will of the person is constrained. These constraints... pertain to the purposes, the

\(^{12}\) Frankfurt, “Autonomy, Necessity and Love” in The Importance of What We Care About, p. 434.

\(^{13}\) The relevance of these characteristics to responsibility for a preference was discussed in chapter 1, section VI.

\(^{14}\) Frankfurt, “The Importance of What We Care About” p. 84.

preferences, and the other personal characteristics that the individual cannot help having and that effectively determine the activities of his will. In other words, they are specified for any given person by what he loves, by what [he] cannot help caring about...[Such necessities] mark [his] volitional limits, and thus, they delineate [his] shape as [a] person.\textsuperscript{16}

Volitional necessities are things about which we cannot help but care, but which we also "cannot help being wholeheartedly behind".\textsuperscript{17} If a particular care is volitionally necessary for an agent, then it is not in his power to give it up at will. He cannot change this fact about himself. And yet he wholeheartedly favours this. The question of whether he approves of the volitional necessity does not arise.

Since cares are not always central to personality, nor always inescapable, the essence of volitional necessity is better captured by the idea of love than by caring. Love need not be for a person, though that is a central case. It can be for life itself, or an ideal. And although we sometimes fall out of love, it is something that we cannot give up right now, at will.

Love is a "powerful source of reasons". "Insofar as a person loves something, he necessarily counts its interests as giving him reasons to serve those interests".\textsuperscript{18} Love provides us with "final ends to which we cannot help being bound; and by virtue of having those ends, we acquire reasons for acting that we cannot help but regard as particularly compelling".\textsuperscript{19}

Loves look like being the parts of the person that cannot be given up without threatening integrity. They are resistant to change, define the self, and are the ultimate source of reasons for action for a person. That they are self-defining also incorporates the notion that they have the authority to represent the agent. They do not need to be authorised by the agent, as might a lower-level desire, because they themselves form the basis on which he assesses his other motivations and courses of action. They provide the foundations from which he can assess the worth of his lower level desires. As such, they stand in the way of an infinite regress.

In response to concerns that volitional necessities might form an arbitrary cut-off point, Frankfurt argues that although they do provide a cut-off, it is not arbitrary. Rather it is necessary for

\textsuperscript{16} Ibid. p. 443.  
\textsuperscript{17} Frankfurt, "Taking Ourselves Seriously, Getting it Right", p. 55.  
\textsuperscript{18} Ibid. p. 52.  
\textsuperscript{19} Ibid. p. 53.
autonomous agency: "With respect to a person whose will has no fixed determinate character, it seems the notion of... self-direction cannot find a grip."\textsuperscript{20} "I believe that the most genuine freedom is not only compatible with being necessitated;.. it actually requires it. In my opinion, actions [are] autonomous when they are performed out of [the necessity of] love."\textsuperscript{21}

Now, the concern here is not with establishing freedom or autonomy, but with finding preferences able to ground identification with other preferences, whilst themselves having authority to represent the agent without needing approval from a further 'higher' preference. There is, of course, the additional question of whether these 'grounding' preferences are those that are crucial to integrity. If we accept Frankfurt's line then loves can fulfil all three requirements. If love is a matter of necessity, such that the agent cannot shed it, then we have at once reached the elements of personality that are most crucial to identity and integrity and that play the 'grounding' role for identification. Concerns that these 'grounding' preferences are not approved of are removed because these preferences constitute the agent's central identity. They are the Archimedean point from which 'lower-level' preferences can be assessed, but do not themselves require assessment or independent approval to have authority to represent the person.

Although Frankfurt's love provides answers to the philosophical problems under investigation, there is a suspicion that these solutions come too cheaply. The element of his account that does the most work, perhaps, the claim that love enjoys a kind of necessity, is also the most controversial. The necessity at once removes concerns about the authority of love to represent the person, and also provides the grounding on which other elements of personality can be assessed. But should we be so willing to accept that love does not need to be approved of, that it \textit{always} has authority to represent the person? Can’t people have inauthentic loves instilled in them, like those who are indoctrinated to love the leader and values of their cult? Are we happy to accept that people are just stuck with the loves they happen to have, that because people are constituted by what they love, they are unable to move beyond the bounds their loves set? Can people not shed their loves and develop new ones, so that reconsideration of loves is not off-limits, as Frankfurt suggests?

Frankfurt's offer of an Archimedean point is useful, but that something needs to play this grounding role does not establish that any particular grounds is itself 'necessary', in Frankfurt's sense of being essential to the agent. As Michael Bratman has observed, finality of end is not the

\textsuperscript{20} Frankfurt, "Rationality and the Unthinkable", in \textit{The Importance of What We Care About}, p. 178.
same as necessity of end. One can care centrally about something, without ambivalence, whilst retaining the ability to give it up. One can be fully committed to a life of religious devotion, or artistic creation, think it good and morally permissible, find it rewarding, and have no intention of changing it, and yet still be able to change it if given compelling grounds to do so. That one *would* not change one’s commitments does not mean that one *could* not change them. Indeed, if one could not change them, even when given compelling reasons to do so, this would seem to undermine one’s authority over these parts of one’s personality, rather than indicate that they constitute one’s central personality.

One argument in Frankfurt’s thinking that might be employed to safeguard the putative necessity of love holds that the necessity is not grounded in the impossibility of change, but rather in the bounds of personal identity. The idea is that a change in one’s most basic commitments would be so fundamental as to result in a different person. As Bratman puts it, such a change would not be volitionally possible for the person because the very same person could not begin and end in that way.

The argument that change of central commitments necessarily amounts to losing one’s identity is, however, too strong. What matters is not simply whether one changes one’s most central commitments or loves, but *how* they are changed. It might sometimes be true that someone who has their deepest love torn away is not the same person afterwards. Those who unexpectedly lose the spouse they have built their world around, or are unable to pursue their driving concerns, like Galileo when forced by the religious powers to forsake his science, can seem genuinely to be broken as a person. Their identity does not survive the loss.

But in the vast majority of cases, people are able to alter the things they love without such dramatic effect. If Galileo had given up scientific inquiry because, on reasoned consideration he thought a life of religious devotion a better life for him to lead, then we would not say that his integrity is undermined and his identity destroyed. As J David Velleman has eloquently argued, changing one’s commitments does not amount to psychic death.

_Self-Governing Policies_

23 Ibid.
24 “Identification and Identity” in Buss and Overton, eds. Velleman attributes the appeal of Frankfurt’s argument to a “fetishistic obsession with the ideal of psychic unity, which is misplaced, unhealthy and unattainable”. I will consider the plausibility of a unified self in section III.
An alternative approach to understanding what might play the ‘grounding’ role in identification, and constitute an agent’s most central self, which avoids the problematic claims to psychic necessity is to look at the issue in functionalist terms. Drawing substantially on the key elements of Frankfurt’s thought - reflective evaluation and the bonds of identity - Bratman himself proposes an alternative candidate to Frankfurt’s love: self-governing policies.

The direction of thought prompted by Frankfurt’s highlighting of the importance of reflective evaluation and approval for a preference to be autonomous led to the idea that evaluations needed to be grounded for them to have the authority to genuinely represent the agent. Higher-order attitudes or desires were needed to assess lower-order ones. Bratman’s insight is to bring to the fore the fact that, due to the time and computational restraints we as agents face when deciding how to act, it is inefficient to have to go through a process of assessment of desires, and the formation of higher-order attitudes towards them, every time we need to make a decision. Instead, he posits, we form attitudes that have a recurring validity. These are policies. A policy is an intention-like commitment to treat certain considerations as justifying when it comes to deciding what to do. As such, it is forward looking, persistent, and provides the grounding for the assessment of options.

Policies are good candidates for being central elements of an agent’s personality because they encapsulate the agent’s standpoint. They are also forward-looking. When settling on a policy, the agent is settling not just on the person he wants to be right now, but on the kind of person he wants to be over time, by settling on the considerations he wants to provide him with reasons in his future deliberations.

Just as with cares, though, not all policies are central. Some, such as a policy to always sit at the back of the bus, might have an ongoing impact on our behaviour, but do not usually represent core elements of personality. To locate those parts of personality that play a central role over time, Bratman takes the kind of hierarchical approach prominent in Frankfurt’s early work. We have higher and lower order policies, the higher having a more overarching role. These might be such things as a policy of developing a strong concern with honesty, or of trying to be more playful with others. Bratman calls these ‘self-governing policies’, and holds that they play a central role in governing an agent’s behaviour over time, because they provide a reference point.

26 M. Bratman, “Reflection, Planning and Temporally Extended Agency”, p. 47.
from which agents are able to assess both which lower-order policies to have, and what considerations they want to bring to bear on their decisions on how to act.

In parallel with Frankfurt’s love, such policies are able to play a deeper role than merely representing where an agent stands. For Bratman, they constitute and support cross-temporal continuity and connections that are central to personal identity over time. They provide one portion of the backward- and forward-looking overlapping psychological connectedness that constitutes and maintains identity\textsuperscript{27}. Their stability arises not from necessity, but from the fact that each person has an interest in maintaining a cross-temporal coherence and unity to their life. Change to what one centrally cares about tends to interfere with one’s plans, and frustrate the achievement of the things one cares about, including one’s concern to maintain integrity to one’s life over time. Self-governing policies are open to change, but there is a strong presumption against it. They support a kind of conservatism: what one cares about most centrally will function as a kind of ‘defeasible default’\textsuperscript{28}.

But Bratman requires an extra element that plays no role in Frankfurt’s account. The authority of self-governing policies needs to be established. This move is necessitated by the switch to seeing a person’s central commitments, now identified as self-governing policies, as being open, though strongly resistant, to change. For if self-governing policies can be reconsidered and altered, rather than simply providing the foundational bedrocks Frankfurt attributes to loves, then the question arises as to why those particular self-governing policies have the authority to stand for the agent.

\textit{Should} a self-governing policy be authoritative, then it may be able to play the stable, persisting, co-ordinating role outlined by Bratman. It will be able to represent where someone stands, to speak for him. As such, self-governing policies seem plausible candidates for rooting identification – for providing the grounds on which one assesses one’s other preferences. Moreover, since they represent an agent’s ongoing vision of the life he wants to lead, they look like the elements of personality that are most central to an agent’s identity, that are crucial to integrity.

\textsuperscript{27} Ibid. pp. 43-46. I will not here defend or fully investigate this neo-Lockean account of identity.

\textsuperscript{28} “A Thoughtful and Reasonable Stability” pp. 88-89. See also his “Identification, Decision and Treating as a Reason”, \textit{Philosophical Topics} 24, no. 2 (1996), and “Reflection, Planning and Temporally Extended Agency”, \textit{The Philosophical Review} 109, no. 1 (2000).
However, for self-governing policies to play these roles, a plausible account is needed of how they gain their authority, and the conditions under which changes to self-governing policies carry authenticity.

**Section III  Centrality, Authenticity and Change**

As mentioned above, self-governing policies are the products of reflective evaluation, and occupy the upper levels of the hierarchy of policies and preferences that make up personality. Yet the authority of such high-order attitudes itself needs establishing – and without falling into the regress of appeal to yet higher-order attitudes. At the same time, simply attributing a particular level of attitude with agential authority risks arbitrariness. The approach I will consider in this section argues that the threats of regress and arbitrariness can be blocked by an appeal to structural features of an agent’s personality. In particular, I will examine the idea that it is the relation between a self-governing policy and other self-governing policies that determines its authority.

The first reason for taking this route relates to an important feature of self-governing policies, that they are intention-like rather than desire-like. The reason they are intention-like is that they provide considerations that the agent intends to play a role in his future deliberations. He wants the consideration to be recurring (where relevant). What is important about intentions is that they are more definite than mere desires. We can have desires for all sorts of things, but have no intention to act on them. When we form an intention, however, we are committed to following it through – though of course the commitment is normally not irrevocable, and new information can make it imperative to reconsider or abandon it. However, once we have formed an intention there is, normally, a rational pressure not to reconsider or abandon it. Intentions, though, are subject to rational demands of consistency that desires are not. I can desire to spend next year in Poland, and also desire to spend next year trying to get promoted in my London-based job without being open to accusations of criticisable irrationality. But I cannot intend to both spend next year in Poland and to spend it in London climbing the slippery pole (not Pole).

Once central preferences are seen as policy-like, and hence intention-like, stronger norms of consistency come into the equation than for more desire-like states. This is relevant to the establishment of authority for a self-governing policy. It rules out the possibility that two central preferences can both authoritatively represent the agent’s stance whilst also conflicting.
Consistency with other self-governing policies looks like a first prerequisite for a self-governing policy to have authority to represent an agent.

But the consistency requirement amongst self-governing policies does not mean that all conflict, even amongst self-governing policies, necessarily undermines authority. It should be possible for a self-governing policy to be authoritative even when one experiences some sorts of conflict, and even violates the policy in a particular case. Policies also need to be stable, yet open to revision. It is important therefore to clarify which conflicts are compatible with a policy being authoritative, and which not.

Whilst policies that undermine each other cannot both authoritatively represent an agent, there will also be clashes between self-governing policies that do not prevent either being authoritative. Take for example, a devout Christian who has a strong commitment to attending mass every Sunday, but who also is committed to supporting his son’s quest to become a professional athlete. One Sunday morning, his son has a crucial athletics trial, and requires support. There is obviously a clash amongst the father’s commitments. But if the father opts to attend the athletics instead of church that does not mean that his religious commitments do not authoritatively represent his agency, or that they no longer qualify as being central to his personality. The fact that he is forced into prioritising need not undermine his commitment to the ‘loser’ in the prioritisation.

The undermining inconsistency is not revealed in conflicts over which commitment an agent decides to act on at any one moment, but rather in whether the commitment provides a reason to be factored into the deliberation. So whilst attending mass and supporting his son are both able to consistently act as considerations in his decision, this would not be the case if he was having difficulty deciding what to do because, for example, he thought that his son ought to devote less attention to sport and more to his religion. In this case, he would not be sure whether he wanted his commitment to support his son’s athletics to figure in his reasoning. The ambivalence would make him conflicted over the question of whether supporting his son provides a valid reason to feed into his deliberations about how to act – even if he does in a more straightforward sense desire it.

29 See Bratman, “Reflection, Planning and Temporally Extended Agency”, p. 49.
The idea, then, is that, to have authority as a self-governing policy, a commitment must be reason-giving in a person’s deliberation on how to act and must not conflict with other self-governing policies concerning his treating that commitment as reason-giving\textsuperscript{30}.

But a consistency requirement might not be sufficient to eliminate self-governing policies that are not invested with authority. A consistency requirement amongst self-governing policies allows that a self-governing policy might be isolated, yet still consistent with others. Someone raised as a catholic who has lost their religious belief might still have a policy of deferring to the exhortation of priests, which, whilst not inconsistent with the rest of his self-governing policies, is nevertheless not supported by them, and a hangover from his upbringing. It does not have the authority here sought – authority to ground identification with lower-order preferences and to be an authentic, non-alien, central element of a person’s identity.

What is needed for such authority is a means to bolster the relationship between a self-governing policy and the rest of an agent’s personality. The remainder of this section will argue that this can be provided by coherence with other key elements of personality.

**Coherence**

The first step towards understanding how a coherence requirement for authority might work is to establish which aspects of personality are relevant - amongst which elements of personality the coherence must hold. It is the relation amongst self-governing policies that has so far been examined, but it may be reasonable to extend the relevant category to include certain kinds of beliefs.

The significance of beliefs to policies can be illustrated by consideration of an alternative source of inauthenticity for self-governing policies. This need not occur by manipulation or distortion of preferences or desires. It may also occur by the imparting of false or misleading beliefs. An example comes from the book ‘The Coram Boy’, which centres on well-to-do women in the 19\textsuperscript{th} century who bore illegitimate children, and gave them away – with endowments of jewels – to middle-men who convinced them that they would place the child at the Coram Hospital for deserted children, thereby securing the child a decent life. Unbeknown to the mothers, though, the middle men were conmen who killed the babies and kept the endowments for themselves.

\textsuperscript{30} Ibid.
They nevertheless won the trust of the mothers by feeding on their desperation and imparting in them false beliefs which grounded their desire to hand over their children\textsuperscript{31}.

In this macabre scenario, were the mothers to learn of the real motives and intentions of the middle men, so that they no longer believed handing their child over to be a means of safeguarding it, then their desire to hand over their children ought, if they are rational, to disappear. Desires and, by extension, preferences, are sometimes in an important sense (to be elucidated below) dependent upon relevant beliefs, and can be rendered inauthentic if those beliefs are inauthentic.

This belief-dependency can, however, also be a source of authority. Beliefs can play a grounding role in evaluations of preferences that helps render the outcome a genuine expression of a person’s agency. This claim finds backing in the common-sense idea that a person’s character is partially defined by his convictions regarding the truth. I might believe Labour candidates are more honest than Conservatives. You might hold that Oxford is a superior university to Cambridge. Someone else might hold that partisan rivalry of this sort is futile and childish. Such attitudes play a part in characterising each of us\textsuperscript{32}. That an evaluation of a self-governing policy, for example, is carried out in light of such beliefs, then, might serve to anchor it in a person’s central self. In response to concerns that self-governing policies, even following evaluation, might not have authority to represent agency, an appeal to those beliefs that represent the agent’s convictions can provide a means of bolstering and increasing the authority of the evaluation and the self-governing policy.

But as the Coram Boy example illustrates, further conditions are necessary to establish that the grounding of a policy in belief is a source of, rather than an obstacle to, authority. Exactly parallel concerns arise with beliefs as did with desires and preferences. The fact that beliefs are themselves open to manipulation means that being supported by a belief may not increase the authenticity of a policy or preference.

We seem to be going round in circles. But that can sometimes, surprisingly, be helpful. For, looking back at the problems and solutions associated with preferences being assessed on the basis of other preferences, parallel options are open to bolster the use of beliefs in supporting preferences. Specifically, we can borrow the idea that some beliefs are more authentic than


\textsuperscript{32} Attributing a role of this kind to (some) beliefs is suggested by Laura Waddell Ekstrom, “Alienation, Autonomy and the Self”, op cit.
others. Those beliefs that are haphazardly adopted, formed perhaps out of unevaluated whim, fear, guilt, or convenience, will be less authentic representations of the agent than those formed by critical reflection with the aim of assenting to what is true. Although it is not here possible to fully develop an account of the process, it seems plausible to argue that only those beliefs formed in an appropriate way by one’s evaluative capacities ought to provide authentic grounding for self-governing policies.

The parallel between beliefs that are able to authentically ground preferences and desires that do so might even extend further. Recall Bratman’s conviction that certain of our desires play a recurring role in our lives, providing a consideration in our deliberation whenever relevant circumstances arise. These policies provide reference points from which both courses of action, and other preferences, can be assessed. Bratman also suggested that they were intention-like, and subject to strong norms of consistency. Parallel considerations are applicable regarding beliefs. We might regard certain beliefs as having recurrent relevance, such that they form future-oriented and procedural principles. A principle of avoidance-of-regret might be an example. In such cases, it seems reasonable to suggest that these principles can play an ongoing anchoring role in the same way that a policy might, and moreover that they are also subject to stronger norms of consistency than other beliefs.

The idea here is that a self-governing policy must, in order to be authoritative and authentic, be consistent both with an agent’s other self-governing policies and with his future-oriented and procedural principles. This broadens and strengthens the necessary support. It also provides us with the second category of elements of personality which need to cohere in order for a self-governing policy to be considered both authoritative and an authentic representation of agency. The task now is to elucidate how this coherence requirement might be understood, and why it is attractive.

Recall that attempts to ground authority in Frankfurt’s original hierarchical account led into problems of infinite regress or arbitrary cut offs. Equally, a consistency requirement amongst self-governing policies seemed insufficiently strong to guarantee authority. Moving from a consistency requirement to a coherence requirement avoids both these problems.

---


34 This idea is hinted at by Waddell Ekstrom, ibid. note 32.
Take, as a working outline, coherence to be a matter of mutual defence in face of external challenge or internal doubt. And let us posit, as with the consistency requirement, that the coherence need not be across all the elements of an agent's personality, but rather within the smaller category of central beliefs and policies, those that play an ongoing co-ordinating role in a person’s identity.

Vis-à-vis the regress problem, when faced with the question of whether some policy is authentically one that we want to have, we need not ascend to ever higher levels of approval. Rather we can appeal to our central preferences and beliefs to assess whether they can be used in its defence, whether they back it up and provide a case for its adoption or maintenance. If those states together form a network of support, a coherent and interwoven structure, then we can say that the policy or preference is authoritative and authentic.

In terms of the danger that consistency amongst those central policies or beliefs that ground evaluation being insufficient to guarantee authenticity, a coherence requirement provides a much stronger means of preventing inauthenticity. To be consistent a policy or belief need not be supported by other policies or beliefs. It simply needs not to contradict them. But coherence does not allow isolation. It insists that a preference is integrated and supported by other preferences and beliefs, making it far more plausible that it is an authentic representation of agency.

Requiring central preferences to cohere represents a stringent demand. A preference that has been evaluated and endorsed, and is consistent with one’s other central preferences, will not qualify as an authentic self-governing policy if it is not supported and intertwined with one’s other central preferences and beliefs. As such, the coherence requirement will be resisted by those who are wary of overly unified pictures of the self. Frankfurt, recall, did not see the need for reflective evaluation to be rooted, because he was unwilling to restrict the ‘approving’ self. In recognition of the diversity of the self, he wanted to allow for a wide range of psychic elements to ground higher-order attitudes. And although in later work Frankfurt strengthens the basis on which evaluations are made, demanding that the agent's endorsement must be wholehearted, even that requirement does not go as far as coherence. It demands only that the wholehearted

---

35 Lehrer, op cit.
36 This is the central argument of Waddell Ekstrom in “Alienation, Autonomy and the Self”, ibid., section V provides a useful account of how the subtleties of a coherence account might be understood.
38 See section I, above.
agent be free from ambivalence, such that at least one of a pair of conflicting psychic elements is rejected “as an outlaw”\(^{39}\).

The concerns with the stringency and unmerited unity of the self inherent to a coherence requirement are mitigated on two accounts. The first is that coherence is not an all or nothing affair. Rather it is a matter of degree. Those preferences that are the most coherent are those that are most integrated and supported by the agent’s other central beliefs and preferences. They do not have to reach an absolute status of complete coherence. Rather they are taken to be the most authentic representatives of the agent’s central self, because they encapsulate the standpoint from which he guides his life more than anything else.

The second mitigation comes from the fact that a coherence requirement does leave room for outlying clusters of preferences that are largely unrelated by content to many of the other elements of one’s character, but which are, amongst themselves, inter-related and mutually supporting, to count as elements of the agent’s central self. Given that such preferences may lack competitors they need not be incoherent\(^{40}\).

This is a potentially attractive feature of a coherence account of authenticity, since it allows a degree of idiosyncrasy. As such it offers a deferential nod towards to those who have concerns about the notion of a unified self. We all, after all, have preferences, of which we approve, that are a little eccentric, and perhaps not often expressed, but we nevertheless see as an important part of ourselves. In fact, that they are a bit quirky may lie behind their attraction and their importance to us. They allow us a break from being our ‘normal’ selves.

Note, though, that the coherence requirement favours those idiosyncratic preferences that are supported within a small, yet related and mutually supporting, cluster rather than those that are one-off. And this seems intuitively right when we come to think of the central parts of personality. We all have one-off quirky preferences that we endorse. I’d like to throw myself out of an aeroplane – ideally with a parachute – one day. But we wouldn’t feel that their fulfilment was particularly important. If the quirky preferences form a cluster, though – say, although generally speaking a sensible family man, I begin to see a taste for occasional extreme sports as part of who I am - idiosyncratic, perhaps, but a not insignificant part of my person – then their relevance to my identity, and their claim to be seen as a central part of my self, increases.

\(^{39}\) Frankfurt, “Identification and Wholeheartedness”.

\(^{40}\) Ekstrom, “Alienation, Autonomy and the Self”, section 5.
Isolation of preference does seem to undermine authenticity and authority, but idiosyncrasy need not.

A coherence requirement amongst central beliefs and preferences is able to resolve the problems with which this section started. It is able to restrict concerns that an agent’s identification with a preference might be distorted and inauthentic. The claim reached is that identification with a lower-order preference is genuine if it is grounded in an agent’s self-governing policies. These policies are authentic and authoritative if they cohere with other central policies and beliefs. Cohering self-governing policies constitute an agent’s central identity.

*The Importance of Origins*

The appeal to coherence to establish centrality and authority may be seen to render considerations of the origin of a self-governing policy redundant. Coherent self-governing policies and beliefs, after all, constitute an agent’s central self, ground identification, discriminate against inauthentic preferences in a way that consistency requirements do not, and avoid the regress/arbitrary cut-off problems inherent to hierarchical accounts.

But there are reasons why coherence alone is not sufficient to guarantee authenticity amongst central preferences. It is possible that a person could have preferences that enjoy a high level of coherence, and thereby constitute a person’s most central identity, but are not authentic. A person’s entire life can be coherently built around an inauthentic self-governing policy. Unhappy examples of this kind sometimes occur when a person suffers a significant trauma of some kind. The case of Tony Martin might be an example. Martin lived alone in an isolated farmhouse. Traumatised by a number of burglaries on his property, Martin became entirely preoccupied with defending his home. Reports suggest that he placed man traps, obtained an unlicensed shotgun, and made his house more like an arsenal than a home. When two teenagers tried to burgle him Martin was ready, shooting one dead, injuring the other. The grizzly conclusion is less important here than that Martin's actions were entirely coherent. The fear and anxiety instilled in him by the initial burglaries became the master of him, driving his behaviour and preferences so that his pre-existing preferences, policies and beliefs were forced out and revised to a new set that centred around – cohered around – his all-consuming policy of protecting himself and his property.

---

41 Alleged burglaries, in fact. The police later cast doubt on their occurrence.
This case illustrates that the coherence of central preferences and beliefs is insufficient for authenticity. Origins matter, too. An alien preference that grounds change in someone’s personality to such an extent that the rest of his preferences and beliefs ultimately cohere around that alien policy does not, on account of that coherence, become authentic. We would not think that Martin’s personality, coherent though it might have been, was authentic, because it was built around preferences whose origins are alien, driven as they were by fear and paranoia.42

Continuity in a coherent set of beliefs and preferences is vital for their authenticity. The coherence cannot be punctured by an alien intruder, be rebuilt again on the basis of that alien intruder, and regain authenticity. In terms of identity, the same person would not survive the process. Rather the change must be the other way round. The authentic, cohering self, must settle which changes occur to it, rather than the changes determining the future shape of the self. Understanding how change can be coherence-preserving is crucial to reaching a full account of the authenticity of self-governing policies. This is the challenge for the final section of this Chapter.

Section IV Judgemental vs Deliberate Change

We can start to understand how change to central preferences preserves coherence, and how not, by considering a second example that will shed light on the distinction, that of Andromache in Euripides’ Women of Troy. With Troy sacked, Hecuba, the Trojan Queen, urges Andromache, a fellow member of the Trojan royal family, to approach her fate with defiant acquiescence. Troy has gone, her family is torn asunder, and she will soon be the concubine of a Greek king whom she hates. Yet Hecuba tells Andromache to recognise the inevitability of her fate, to leave behind her Trojan identity, her allegiances, her culture, and values. Instead she must construct herself again, making the best of her new situation so that she might, instead of being condemned to endless suffering, find purpose anew. In response, Andromache reluctantly accepts her fate, and travels, enslaved, to Greece. There she becomes the mistress of Achilles’ son, reinventing herself, and bearing him a child.

The case of Andromache is in some senses parallel to that of Martin but illustrates a different aspect of the importance of the origins of preferences. Andromache deliberately, and knowingly, 

42 Martin, in fact was ultimately deemed not fully responsible on account of an inability to govern his preferences and behaviour.
recreates herself, developing new tastes and commitments to suit her new situation. If we assume, with a little artistic licence, that the range of options open to her is not so narrow as to mean that her choice is compelled, then her new preferences might be thought authentic. She identifies with them as the only set of preferences likely to provide her with any satisfaction at all. Now, so long as we assume that the new preferences Andromache develops form a coherent set, an account of authenticity that does not include a procedural element would judge her new set of preferences authentic.

However, a procedural, coherence account of authenticity would reach the opposite conclusion. This is because the change in Andromache’s preferences is not made in accordance with her central preferences and beliefs. It is not because her judgement moves her to think that the values and tastes of Athenian life are worthy of adhering to that she develops them. It is simply because she needs to be integrated into her new society, which happens to subscribe to a particular set of values. If she had been taken to Carthage instead, she would have adopted the values and commitments she takes on are not adopted because of their accord with her central beliefs and preferences. Rather she forces herself to take them on out of hopelessness.

What the cases of Andromache and Martin share is that a new set of preferences is developed which itself coheres, but which is not formed on the basis of its coherence with the pre-existing personality. In Martin’s case, the change is driven by a powerful interference in his motivations that pervades the rest of his personality. In Andromache’s case, the change is driven by her own deliberate manipulation of her personality, but it is nevertheless enforced against her judgement, rather than in accordance with it. The argument I will develop in this section holds that neither change in personality is authentic.

A third scenario can demonstrate the contrast with authentic change to personality. Let us alter the circumstances of Andromache’s self-transformation, and assume that rather than being enslaved by the Greeks after her city had fallen, she spent time in Greece on a royal visit in peaceful times prior to the war. Although at first suspicious of their foreign ways, Andromache comes to feel that Greek culture is in fact preferable to her own. The tough, unflinching attitude that she had previously been appalled by seems to make sense when she sees the social stability and military success that accompanies it. The pursuit of knowledge, which at first she saw as
lofty, begins to fascinate her, and so on. Before long she comes to change her policies and beliefs about the right way to approach life. She develops new self-governing policies.

In this scenario what is different (apart from the complete disappearance of the dramatic interest of the story), is the relationship between the changes made and the preferences and beliefs that make up Andromache’s central personality. The changes are driven by their accord with her personality, rather than enforced upon it. In both of the posited scenarios, Andromache has reasons to change, and decides in their favour. Yet in only one of the two cases are the changes to her personality authentic. The remainder of this section will explain why.

We can understand the difference between authentic and inauthentic self-transformation by returning to the relationship between a desire and the beliefs or reasons that support it. As discussed, self-governing policies depend both on other desires, and certain relevant beliefs. Two aspects of the dependence are relevant to the current investigation. A self-governing policy, such as a policy to pursue a new career, depends for its justification on various relevant beliefs - beliefs, for example, about one’s employability, about the availability of alternative options, etc. Were a person to lose these beliefs, he would lose justification for the policy. The second dependence hinges on the fact that losing justification for the policy ought to make a difference to whether the person still has the policy. The actual existence of the policy might not survive the discovery that the alternative jobs involve longer hours. Of course, the disappearance of the policy won’t always occur, but this opens up criticisms of irrationality. Ordinarily a self-governing policy owes its existence (as an empirical psychological fact) to its bearer’s own deliberations and overall assessment of his situation.

But this dependency does not apply to self-governing policies that we deliberately bring about in ourselves. We can, as the original case of Andromache demonstrates, produce new central preferences, beliefs or policies in ourselves by means that bypass judgemental evaluation. We can force them upon ourselves by means of mental discipline, for example, or if short of that, by hypnosis, or boot camps, etc. But because the policy does not have the right kind of relationship

---

43 Scholars of Ancient Greece will have to forgive my unresearched caricature of Greek life.
44 The following account of the authentic development of tastes draws substantially from Richard Moran’s development of the distinction between being passive and active towards elements of one’s mental life as being the essence of identification. “Frankfurt on Identification: Ambiguities of Activity in Mental Life”, in Buss and Overton, 2002, pp. 189-217.
45 This kind of dependency, of course, does not apply to brute tastes. This is the basis of Cohen’s brute/judgemental taste distinction.
46 The expensive judgemental taste objection hinges on this fact.
with our beliefs and preferences in these scenarios there is an important sense in which the new policy is inflicted on me, even if I am the one inflicting it\textsuperscript{47}.

The difference emerges from the coherence account of centrality and authenticity developed in the previous section. The policy I inflict on myself does not enjoy the same kind of relationship with the rest of my preferences and beliefs as a policy I arrive at via deliberation. In the case of a policy that results from deliberation I am not only prepared to justify the policy (as I might be with a deliberately acquired, inauthentic preference), but the presence or absence of justification makes a difference to the presence or absence of the policy itself, and the direction of the policy is in fact guided by the direction of my background thoughts about what is desirable. As Moran notes, an agent authentically arrives at a policy or preference "not because he has \textit{produced} it in himself, but because he takes the general question of what he wants here to be the expression of his sense of what he has best reason to pursue in this context. Were those considerations to be undermined, the [preference] itself would be undermined." That means there is a sense of control he doesn't have in these cases. He won't be in position to retain the preference, or to reproduce it at will, if he abandons the considerations that supported it, so that the object of the preference now seems worthless to him. Of course, he could force himself to keep the preference, and even drive his other beliefs and preferences to support and cohere around it. But this is external control. He would be forcing the preference to dominate his reason, and his network of relevant beliefs and preferences, rather than vice versa. This represents alienation from his cohering central self rather than the expression of it\textsuperscript{48}.

Moran uses an illuminating distinction to highlight the different ways in which a person can develop central preferences. The inauthentic 'deliberate' route, is akin to taking responsibility for the preferences of a second person, say a child under one’s care. One can exert influence on those preferences, but they are not the expression of one’s will. One can manipulate and even enforce change, but the process is one based on opposition. The sense of change that is relevant to authentic, or what I shall henceforth term ‘judgemental’, development of preferences is not one that reduces to the ability to exert influence over one’s preferences. The idiom of control is misleading in this context. At the beginning of our deliberation, we do not aim to produce a particular preference in ourselves, as we might with respect to another person, but we rather hold open our preference to how the balance of reasons fall out. Deliberation must be able to make a difference, so that our preferences emerge in ways consistent with our character and reasoning. Our other attitudes, our policies, preferences and beliefs, must have some voice

\textsuperscript{47} Moran, op cit. p. 198.
\textsuperscript{48} Ibid. pp. 198-99.
in determining the preference we arrive at. And this requires that we are in fact passive with regard to the outcome of our reasoning. We must be open to adopting it. Should this passivity be absent, so that the outcome was actively enforced, the outcome would be alien from our self, inauthentic, not the result of our judgement\textsuperscript{49}.

Judgemental change does not require approval with the resultant preferences as an extra step. The means by which the preference emerges ensures it. The new preference is formed in accordance with the norms of the agent’s practical reasoning, is grounded in his other, authentic, beliefs and preferences, and therefore coheres strongly with the rest of his central self. It is, by means of its coherence and origin, authentic, a genuine expression of agency. He is not faced with the question of whether the new preference represents what he ‘really’ wants, because the mode of generation is in fact the means of discovery (and development) of what he does in fact ‘really’ want.

The openness to revision of central preferences and the lack of agential control over the direction of judgemental change might be thought to conflict with the long-lasting stability commonly associated with centrality. But in fact it grounds it. The difference is that the persistence is not grounded in an absolute inability to shed a preference, or the ‘death’ of the person’s identity if shorn of it, but in the fact that the preference emerges out of and coheres strongly with the rest of the agent’s personality. It occupies a central role in his person at the moment, but also represents how he feels he should approach the future. Changing it risks undermining the continuing policies and commitments that most represent who he is. The result, as Bratman has observed, is that the preference is open to change, but the agent sees no reason to change it, and plenty of reasons not to\textsuperscript{50}.

Of course, very few of our preferences come into existence as a result of this kind of practical reasoning. Indeed, very few beliefs arrive as the conclusion of theoretical reasoning\textsuperscript{51}. But then not all beliefs or preferences are central. What is being established here is how to understand authentic ‘judgemental’ development of central preferences as opposed to inauthentic ‘deliberate’ development of central preferences. And the picture proposed is one by which an agent might develop and revise his central preferences. That it may, but very frequently will not, apply to other preferences does not undermine the validity of this picture. In fact, it seems to enhance it.

\textsuperscript{49} Ibid. pp. 198-203.
\textsuperscript{50} M. Bratman, “A Thoughtful and Reasonable Stability”.
\textsuperscript{51} Moran, op cit, p. 198.
It seems intuitively right that our judgemental preferences are significantly less numerous than our brute tastes.

The claim then, is that a self-governing policy must be formed ‘judgementally’ in order to be authentic. Before accepting this understanding of authenticity-preserving self-transformation, though, an important objection needs to be addressed.

The objection holds that the type of self-transformation deemed authentic is too cognitive, too considered, too rational. People often change organically, so that their tastes and beliefs gradually morph and alter, sometimes without their even realising. And at the other extreme, people sometimes experience epiphanies of some sort, where a dramatic change in personality can occur, typically as the result of a single, seemingly earth-shattering, incident. That changes to personality occur in these less cognitive ways ought not to mean that they are necessarily inauthentic.

Consider a woman who has decided to prioritise her career and put aside any thoughts of having children. She has a self-governing policy of not letting reproductive considerations stand in the way of career progression. If she unexpectedly gets pregnant she will have to decide what to do. Her relevant self-governing policy tells her she should give the child up for adoption. However, driven perhaps by maternal instinct, she begins to feel an unexpectedly potent attachment to the life growing inside of her and a corresponding desire to keep it.

It is likely that her unheralded desires to keep the child will precipitate a re-evaluation of her policy – a key feature of self-governing policies is that they are open to rational revision. She might, then, assess it in light of its coherence with her other self-governing policies and central beliefs. Perhaps after a thorough evaluation she decides to stick with her career-first policy and give up the child. However, when the moment comes to give up the child, she finds she is unable, her maternal instincts are too strong, her desire to keep hold of her child irresistible.

The account of authenticity advocated in this chapter would hold that the preference she acts upon is inauthentic. It is not endorsed by her authentic self-governing policies. But this conclusion is regarded by some as embodying an overly intellectualised picture of mental life. That a person forms a coherent judgement about a policy does not in itself mean that the judgement is authentic. As section I discussed, higher-order policies, even when part of a

---

coherent set formed in accordance with judgment can be unduly influenced by the expectations of others or the social mores of society. That the woman judges it best that she should give up her child rather than follow her natural impulse to keep it could have been the result of any such illegitimate influence. The presence of judgement and the coherence between higher-level preferences and beliefs does not seem enough to guarantee that it is not in fact her natural desire to nurture the child that actually represents what she authentically wants, and what is most in tune with her most central self\textsuperscript{53}. Giving up the child might be consistent with her career-dominated life plan, but that life plan might be less authentic than her motherly desires.

This is an important challenge. But rather than precipitate an abandonment of the procedural, coherence account of authenticity, it suggests an extension of the mental elements that are involved in the authentic formation of preferences. As discussed, judgemental tastes depend on other beliefs and desires for their justification, and the justification roots their continued existence. But the influence of other 'lower' (on hierarchical grounds) elements of the mental stew may well extend beyond beliefs and desires. It seems plausible to assume that our central preferences and policies are also informed by less cognitive mental elements, such as emotions and affective responses. Charles Taylor, for example, in his thought-provoking "Responsibility for Self", discusses the idea that we develop our most central values and commitments by attempting to articulate our "largely inarticulated sense of what is of decisive importance"\textsuperscript{54}, which is an amalgam of pressures from various elements of one's psychic stew, both cognitive and non-cognitive\textsuperscript{55}.

Introducing an element of dependence of self-governing policies on non-cognitive mental elements means that authentic changes to personality can be brought about by changes in non-cognitive parts of our personality. This might happen gradually, or more dramatically. If I start to feel a creeping disillusionment with my commitment to equality, for example, or that there's something untrustworthy about undergraduates' excuses about the quality of their essays, then the policies and commitments they inform – to campaign for equality, or to be charitable to undergraduates – will be correspondingly weakened. The weakness may not precipitate a change in a central preference, but it will reduce the support that preference rests on, and render it less coherent, and ultimately more open to change.


\textsuperscript{54} Reprinted in Gary Watson, Free Will, New York: OUP, 1982, at 122-123.

\textsuperscript{55} The term 'psychic stew' comes from M. Bratman, “Reflection, Planning and Temporally Extended Agency”, op cit, p. 51.
The dependency of central preferences on so-called ‘lower’ mental elements introduces an attractive bottom-up element into the process of self-development, rendering the picture here presented less open to accusations of being overly rationalistic and insufficiently sensitive to the complexity and multi-faceted nature of the self. The ‘lower’ might crystallise into the higher or more central, which will then ground the approval of other ‘lower’ desires, but the process has an important circularity to it, a loop of inter-dependency, whereby the complex of lower, non-central, desires and other mental elements has the ability to either sustain or undermine the ‘higher’ or ‘most central’ preferences.

The question arises, though, of how far the bottom-up dimension can go. What if, for example, the woman’s powerful drive towards keeping her child drove her to a radical self-transformation, in which she abandons her career and revises her self-governing policies to form a coherent set revolving around her desire to nurture and support her child? Could this self-transformation be authentic? And if so, how would it differ from the Martin case? The transformation of Martin’s personality, after all, was precipitated by a powerful affective drive that resulted in a coherent set of self-governing policies.

The authenticity of the self-transformation in these cases must depend on the direction of fit between the self-governing policies and the new, powerful desire. What is crucial is that the self-governing-policy is settled on by consideration of what is desirable, rather than merely what is desired. In the case of Martin, it is the desires that drive the desirable. His overpowering concern with protecting his property drives the formation of a set of self-governing policies that cohere around it. The change of direction in the woman’s life to centre around childrearing can be authentic if the new self-governing policies are settled upon in light of their relation with her other policies, judgements, and convictions, so that the role she attributes to her maternal desires is dictated by their coherence with the rest of her personality, rather than the rest of her personality being driven to cohere with her forceful maternal desires.

In authentic self-transformation, evaluative judgement ultimately holds sway. This judgement is not the unbounded, ‘free’, choice of metaphysical libertarianism considered in chapter 1. The judgement itself is not made independently of the agent’s cohering, settled, personality – were it to be so it would be inauthentic – rather it is embedded in and emerges from the agent’s character. As such, there is an important way in which he is passive with regards to the outcomes of self-forming evaluations. He does not choose to adopt the outcome of his evaluations, because they are the means of discovering what he authentically wants and where
he stands. Once arrived at, they authoritatively represent him.

**Conclusion**

Debates around luck-egalitarian responsibility have moved from disputes over the relevance of genuine choice to a recognition of the relevance to responsibility of the importance a preference has to its bearer. This shift has not been accompanied by any real analysis of the concepts and constituent elements that have been employed as the basis for ascriptions of responsibility: the nature of central preferences; of identification with a preference; and of the authentic development of the person. This lack of attention to the grounds for consequential responsibility has undermined the plausibility of their political usage. This chapter has set out to do some of the broadly metaphysical groundwork that the political accounts require but conspicuously lack.

Section I argued that, in order to carry weight, identification must be understood as approval grounded in other aspects of an agent’s personality. The basis on which the approval is made, though, cannot be whimsical. Identification must be rooted in parts of personality that have a greater claim to represent the agent than the preference being identified with.

Section II examined several parts of personality that might perform the grounding role for identification. The first was a person’s cares, and particularly the subset of cares that is volitionally necessary, loves. Loves, it was posited, could provide the reference point from which people assess their preferences and behaviour. Moreover, loves can be considered those parts of personality that are most integral to identity. Loves cannot be changed at will. People are constituted by their loves so that the loss of a love amounts to a loss of identity. On this account, loves have authority to represent a person’s most central self.

This account was undermined by the possibility that loves might be inauthentic, brought about by illegitimate means. If so, they cannot simply be assumed to authentically represent the agent, but must be authorised to do so. Self-governing policies were proposed as an alternative. Self-governing policies are intention-like, persistent, and forward-looking. They set the considerations by which a person wishes to structure her person and guide her actions. Unlike loves, they are open to rational revision. People can change them, if given good reason, without violating their integrity or undermining their identity.

The openness to change raises the question of how self-governing policies attain their authority to genuinely represent an agent. Sections III and IV addressed this question. Section III argued
for an understanding of authority as a matter of the coherence between a self-governing policy and other self-governing policies, as well as with relevant beliefs, particularly those that are themselves formed authentically and have an ongoing role in co-ordinating an agent’s behaviour. Section IV argued that a procedural element must be added to the coherence account of authority. Changes in a person’s self-governing policies are authentic when made in accordance with her evaluative judgement. This is a process in which the agent is importantly passive, holding herself open to the conclusion of the process rather than directly controlling it.

The limitations of space have permitted only a brief sketch of the structure of the self and its mechanisms of self-regulation and change, rather than a full defence. Moreover, the relations within the self here presented are themselves contingent rather than a priori. Nevertheless I hope to have presented an independently plausible picture of the structure and functioning of the person which fleshes out a coherent understanding of the notions that have been used, without foundation or full explication, in the luck-egalitarian debate. The conclusions reached take us a long way from what might be called the ‘naïve’ first person accounts employed by Dworkin and Cohen, and by equal opportunity for integrity. The task of the next chapter will be to feed this picture back into the luck-egalitarian accounts, and assess the implications for consequential responsibility.