UNIVERSITY OF CALIFORNIA
RIVERSIDE

Perpetual Openness: A View of Governing the Self

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Philosophy

by

Marie Grace Evanston

March 2023

Dissertation Committee:
Dr. Andrews Reath, Chairperson
Dr. John Martin Fischer
Dr. Agnieszka Jaworska
Dr. Coleen Macnamara

The Dissertation of Marie Grace Evanston is approved:

_____

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgments

Many thanks to my committee: to Andrews Reath for being a steadfast and patient presence throughout this anxiety-inducing process; to Agnieszka Jaworska for her incisive comments; to John Martin Fischer for his encouragement; and to Coleen Macnamara for believing in me and helping me to believe in myself.

Dedication

To my father, who has always pushed me to grow even when it was beyond my comfort zone. And to my mother, my favorite interlocutor.

ABSTRACT OF THE DISSERTATION

Perpetual Openness: A View of Governing the Self

by

Marie Grace Evanston

Doctor of Philosophy, Graduate Program in Philosophy
University of California, Riverside, March 2023
Dr. Andrews Reath, Chairperson

"Autonomy" is an equivocal term. The central problem of this dissertation is finding an adequate account of the most robust kind of autonomy, which I call "Philosophical Autonomy". I argue there are two common ways of interpreting "self-governance": governing *from* the self, and governing *the* self. These two notions are often confused in the literature, but they must be kept distinct.

Philosophical autonomy is properly understood as governance of the self, not governance from the self. Governance *from* the self holds presupposes that I already have a substantial self which has the requisite authority needed for autonomy, and that I self-govern when I act from this substantial self. In contrast, governance of the self requires that I decide what my substantial self will be; I cannot simply accept the substantial self I already have. As the fullest kind of autonomy, philosophical autonomy must be governance of the self; it must require me to decide on my very self.

After surveying current accounts of autonomy offered in the literature, I conclude that none of them provide a satisfactory account of governing the self. This is because they all fail to explain how we can be meta-active: that is, how we can decide what our substantial selves will be. I then proceed to diagnose why meta-activity has been impossible to account for, and propose a solution. The resulting theory is one which centers the need to be perpetually open to feedback about the ways I am/have been meta-passive. Perpetual openness requires that I take the question "Who will I be?" seriously such that I am committed to giving a good answer to this question. But the only way I can give a good answer is if I presuppose that there is some objective standard for what counts as a god answer. I must therefore take the world seriously. The fullest expression of autonomy, which started out as an ostensibly navel-gazing project, ends up requiring me to engage sincerely with the world and other people.

Table of Contents

Introduction

Fundamental in modern liberal societies is a deep concern for individual freedom. We believe that a person should be free to pursue the life they deem best without coercion or oppression from others: to practice the religion, pursue the projects, and abide by the values of their choosing, so long as they do not hurt or impede the freedom of others. At the core of liberalism is the idea that each person has this ability to chart her own life – an ability unique to humans, and closely connected to the special dignity that persons have. This individual freedom is also known as autonomy.

There are many problems with autonomy, not least of which is the multiple senses the term has. This dissertation will be primarily concerned with just one problem: how to differentiate a particularly exacting kind of autonomy, which I will call "philosophical autonomy". Philosophical autonomy is, I will argue, the fullest expression of individual autonomy. The main goal of the dissertation is to describe what philosophical autonomy involves and how it is possible.

Typically, it is assumed that all ordinary adults have the basic ability to make decisions for themselves, and so have the essential capacity of autonomy. Individual rights are supposed to ensure the space for individuals to exercise their freedom. For political and legal purposes this simple conception of autonomy not only works, but is preferable to more exacting conceptions. If we set the bar for autonomy too high, we could end up denying rights to large chunks of the adult population.

But many philosophers, though they do not want to deny the importance of this basic kind of autonomy plays, share a concern. They worry that just because a person has

the freedom to decide on particular actions does not yet mean that she is autonomous. This is because the very processes by which she deliberates – the values, worldviews, perspectives about what is proper and appropriate for certain kinds of people, and so on – might be importantly unfree. So although a person's actions might be free, and therefore be sufficiently protected by liberal rights, if we are genuinely concerned with individual freedom we need a different conception of autonomy.

This brings us to a fuller – but not yet the *fullest* – conception of autonomy. This fuller conception is meant to be continuous with the ordinary conception of autonomy. We believe basic autonomy is important because a person should have the right to decide what the kind of life she wants to lead. The philosopher points out that things are a bit more complicated, since in many cases it is far from obvious that a person has actually been free to decide for herself. Subtler influences – chief among them socialization – can undermine the person's freedom to pursue the life she wants. This means that simply addressing external forces like coercion, as rights are meant to do, is not enough for autonomy. In other words, the same basic concern for individual freedom which underlies political autonomy compels us to this fuller conception of autonomy. The goal is to clarify the conditions under which we can say a person really does want what she has chosen for herself. The key question becomes "How can we distinguish autonomous (authentic, genuine) desires from ones that are non-autonomous (conditioned into us, inauthentic)?" The trick is to get in touch with the core self, uncorrupted by nefarious influences, which

knows what it really wants. Once we do, we will be autonomous in a fuller sense[1]. Again, this kind of autonomy is connected to more basic notions of autonomy: a belief that every person has a special inner self is one of the underlying motivations for ensuring individual freedom.

But this fuller kind of autonomy poses its own problems. It is difficult to pinpoint the standards for authentic desires, not least because the influence of socialization and historical circumstance is so pervasive. No matter how deep into the self we go, it seems we may never disentangle ourselves from these influences. For some purposes, this may not be a problem. So long as the person is satisfied with the life he has, along with the freedom (and external support) to re-examine this life if he feels compelled, there may be no need for an outside observer to worry that he is not *really* autonomous in this fuller sense.

But the logic of self-governance itself presses us beyond this fuller kind of autonomy. For even if I find a way to "get in touch" with what I "really" want, it seems that I am still relying on something merely given to me. I simply *happen* to find compelling what I do: my interests do not come from *me*. If I want to be *fully* self-governing, I need to go beyond them. This is a concern which authenticity-centered autonomy cannot adequately address. It is a concern for self-governance *itself*, a desire to be fully in control of myself[2].

---

[1] In chapter one, I will introduce the term "governing from the self" to correspond to this kind of autonomy.

[2] Can authenticity and autonomy come apart? Surely authenticity is separable from some senses of "autonomy"; I argue below that it is separable from philosophical autonomy.

At first glance, this may seem identical to the concern for authenticity, which is about deciding for myself the kind of self and life I will have. But the two come apart. The underlying goal of authenticity is best described as *satisfaction*, *fulfillment*, or *self-expression*. The idea is that by getting in touch with one's "essence" and expressing this unique perspective, a person will be better able to lead a rewarding life and increase her chance at happiness, in addition to the value inherent in such self-expression[3]. But the underlying goal of philosophical autonomy is *control* or *self-responsibility*[4]. Philosophical autonomy is not concerned with outside influences simply because they might be hindering my ability to discover what it is I most desire, value, or care about. Rather, it is *directly* concerned about the fact that while I am inclined to believe I am in control of my self, it is extremely likely that I am not actually in control. The most basic and intimate thing about me – my very *identity* – is largely shaped and determined by external forces. These external forces include not only socialization, the main obstacle for authenticity, but also more "innate" traits. Interests and predispositions which I simply happen to have are just as much outside my control as the circumstances and society I was born in to.

But it also seems to be an important concept for some senses of autonomy, including the one I am sketching here. I will continue to occasionally speak of authenticity-centered autonomy because it is a useful contrast for the autonomy I am interested in.

[3] Once again, we see how authenticity-autonomy is connected to minimal/political autonomy. By guaranteeing external freedom via rights, political autonomy is meant to give each person the ability to choose the life they think will give them the best chance at happiness. It is also meant to respect the dignity of each person inherent in his ability to self-direct. Part of this dignity could be seen as the unique inner self that each person has, and should be valued for.

[4] Not to be confused with *moral* responsibility. I explain this further in chapter 2.

Philosophical autonomy is therefore the fullest kind of autonomy. The exacting standard it demands is that I be in control of, and therefore responsible for, my self[5]. While it is similar to both minimal autonomy and the fuller autonomy many philosophers are concerned with, it is identical to neither. The problem is articulating what this demanding kind of autonomy requires, and if it is possible for us to achieve it.

Chapter 1 will differentiate philosophical autonomy from other notions of autonomy, illuminating what it involves and why it is important. In particular, I will argue that we must keep distinct two ways of interpreting "self-governance": as governing *from* the self, or as governing *the self*. Although the difference between these two has sometimes been implicitly recognized in the literature, it has never been explicitly laid out and many have failed to keep the distinction in mind. It is the latter interpretation – *governing the self* – which lies at the heart of philosophical autonomy. As such, I will use "philosophical autonomy" and "governing the self" interchangeably throughout this dissertation.

Chapters 2 and 3 will survey the current literature on autonomy and allow us to conclude that none of the accounts presently on offer are sufficient for philosophical autonomy. Since the distinction between governing from the self and governing the self has not always been recognized, part of the job of these chapters will be to clarify which views are only of governing from the self, and which approach governing the self. Chapter 2 will argue that certain kinds of accounts can only be governing from the self, and therefore cannot be sufficient for philosophical autonomy. Chapter 3 will look at accounts

[5] The term I will use for this, also introduced and expounded in chapter 1, is governing the self.

that have elements of governing the self, but ultimately concludes that none of these accounts succeed. Chapter 4 will turn to my own account of philosophical autonomy, supplying what is missing in the current literature. Finally, Chapter 5 will discuss various problems of my account and explain how it works in practice.

But before we can disentangle philosophical autonomy from other conceptions of autonomy, we first must disentangle autonomy in general from several notions with which is it commonly intertwined: moral responsibility, identification, and internality. Making these clarifications will set the stage for the rest of the project.

Section 1: Moral Responsibility, Internality, Identification, and Autonomy

Identification is a core concept in contemporary philosophical debates surrounding both moral responsibility and autonomy. The term was first introduced by Harry Frankfurt in his seminal piece "Freedom of the Will and the Concept of a Person". In this paper, Frankfurt argues that a person is only morally responsible for actions which she identifies with, for it is only such actions that can be said to come from *her* as a *person*. With this article, Frankfurt set off a long debate. What is required for identification, and whether identification is indeed necessary for moral responsibility, has been hotly contested.

Identification has subsequently been used in the autonomy literature as well as the moral responsibility literature. Authors have invoked identification directly to ground autonomy – e.g., Gerald Dworkin's piece "The Concept of Autonomy"; and authors have understood *other* philosophers who have used the concept to be providing accounts of autonomy – e.g., Sarah Buss' piece "Autonomy Reconsidered". The latter move is problematic, as John Martin Fischer makes clear in his piece "Responsibility and

Autonomy: The Problem of Mission Creep", since in some cases the original author was intending to provide an account of moral responsibility and *not* autonomy. I'll discuss this further in a bit.

Originally, identification was meant to be the ground for determining what counted as *internal* to an agent. The idea was that if someone identifies with a desire, this desire thus becomes central to him *qua* person. It is no longer a whim or passing fancy: it is part of who he is, and therefore internal to him. However, as the discussion surrounding identification developed it was recognized that these two concepts can come apart. Something can be internal to a person in the sense that it is a core part of the kind person he is even if he does not consciously or actively *identify* with it. In such cases, he *is identified* with it[6].

We thus have four concepts which are commonly interwoven in philosophical discussion: moral responsibility, autonomy, internality, and identification. Because they are so commonly intertwined, we must clearly separate them: first moral responsibility from autonomy, and then internality from identification. Once we have done this, we can ask the core question: can either internality or identification provide an adequate ground for autonomy?

1.1: Moral Responsibility and Autonomy

Moral responsibility is the practice (or set of practices) of taking persons to be appropriate targets of blame and praise and self-regarding attitudes such as guilt and pride.

[6] I credit Agnieszka Jaworska for this helpful way of putting the distinction, from her paper "Caring and Internality".

This means that moral responsibility involves both *judging* persons to be responsible as a matter of fact and *holding* them accountable as a matter of interpersonal practice. The practice of moral responsibility goes beyond the simple fact that people cause things to happen. We only attribute moral responsibility to beings that have certain capacities. If I trip on a crack in the sidewalk and run into you, bruising your arm, this is not the sort of action that most would take me to be morally responsible for, even though I did cause it in some sense. It does not come from any of the capacities that we think are relevant for moral responsibility – for instance, my ability to reflect, to value, to decide what to do, or to form intentions. It is in virtue of having these capacities that we take people to be appropriate targets of blame and praise. Which capacities ground moral responsibility and how precisely they are linked to the practices of judging and holding responsible are core questions of the moral responsibility literature.

Autonomy is the ability to govern oneself (or the actual *state* of governing oneself). What it means to "govern oneself" and what this necessarily involves are the founding questions of any account of autonomy. As a very rough starting place, agents are thought to be autonomous when their desires, values, and projects are "genuinely their own" and not the result of illicit influences such as brainwashing, indoctrination, coercion, and oppressive forms of socialization. While such cases mark the most obvious boundaries of autonomy, what it means for something to be "genuinely your own" - and what is involved in determining what is "genuinely your own" – is elusive. The basic intuition, I take it, is that your actions are not decided for you by other people. They come from you as an independent person. "Independent" need not mean that you are *substantially* independent

from other people – that is, it doesn't require that you be self-sufficient and unattached – but it *does* mean that you are not simply an empty vessel for what others want you to be (or want from you). You are an active and substantial force in your actions (at the least), in the life you have and the kind of person you are (at the most). This is what I mean when I say you are not just an empty vessel or raw block of clay for others fill and shape how they want; you provide to a large degree your own content and shape.

Some of the classic questions posed in an account of autonomy include:

1. How should we understand the self involved in the relation of governing oneself – both the self doing the governing and the self being governed?
2. What makes something "genuinely one's own"? What does it mean to "really" want something? Involved in this question are 5 and 6:
3. What separates illicit influences from legitimate ones?
4. What do we mean when we talk of the "true self" or the "real self"?

These descriptions, brief as they are, should make clear that moral responsibility and autonomy are distinct concepts which come apart. There *is* a sense in which moral responsibility and autonomy might seem to go together. We only consider certain kinds of beings to be the appropriate object of moral responsibility: namely, all fully-functioning adult humans. In contrast, we do not hold animals, infants, or the severely mentally impaired to be responsible. This is largely because we do not take them to have the necessary kinds of capacities which ground moral responsibility. They lack, for instance, the ability to reflect on different options for action; to reason about the points in favor of and against each action; to perceive the different kinds of reasons there might be for or

against each action; and to form a deeper awareness of themselves or the world such that they can engage in this rich reasoning and decision process. (We might also want to include in this list the affective dispositions to be moved by a wide variety of reasons.) Whatever the details, the common theme is that (typical) adult humans have a form of control over their actions which animals, small children, and the severely mentally impaired do not, such that the latter group cannot be held to the same standards. There is a common understanding of autonomy which goes along with this sort of control. We typically think that all adults are autonomous in the sense that they have the inherent ability to direct their own lives and make their own choices. In this way, it is plausible to think that it is the same kind of control over one's life and decisions which make one both autonomous *and* morally responsible.

But this is a very minimal sense of autonomy. Does it fulfill the above description? This depends entirely on how robustly we interpret the requirements of "self-governance" and how loosely we draw the boundaries for being considered autonomous. If we define "self-governing" merely as having avoided the most extreme forms of brainwashing and coercion, then the vast majority of adults *will* fall into the category of "autonomous", and it becomes more likely autonomy is coextensive with moral responsibility (whether the two *are* co-extensive would require more investigation). However, under a stricter set of requirements for what it means to "govern oneself", this minimal sense of autonomy would not guarantee autonomy in the fullest sense. A person might very well be minimally autonomous in the sense that they are choosing actions of their own volition; but if the

desires and motivations which have led them to make these choices are not genuinely their own, they lack fuller kinds of autonomy.

Minimal autonomy is closely related to *political* autonomy, which we will discuss further in Chapter 1. As indicated above, political autonomy is (or should be) accorded to all full-functioning adults and is meant to provide a space within which persons are free to exercise their minimal autonomy without outside interference. Within certain parameters (most paradigmatically, that no harm is done to others), the choices a person makes are up to them and should be respected. Since a typical adult can reason and understand the different choices available to them, it is assumed *any* choices they make "come from them". In a sense, this is undeniably true, and for political purposes of respect for individual freedom it is important to keep this assumption. But full autonomy looks more critically at the motivations behind the choices. Are the underlying values, interests, etc., ones the person truly endorses or wants? Is the person in control of not only their particular actions and decisions, but of the deeper reasons *why* they make the choices they do? Thus, while minimal autonomy is important, it must be distinguished from full autonomy.

Setting aside minimalist notions of autonomy, the distinction of moral responsibility and full autonomy is probably clear to most of us. (For the rest of this section I will use "autonomy" to mean "fuller autonomy" for the sake of brevity.) Perhaps most basically, one can be morally responsible for actions which are nevertheless non-autonomous. Consider the case of Andrew Clark from The Breakfast Club, who tormented one of his team members because he wanted to impress his domineering and aggressive father, and subsequently felt horrendously guilty about his action. I take it to be clear that

Andrew was not acting autonomously: he did not condone his action, he regretted doing it immediately afterwards, and he only did it because of the influence of his father. Put another way, the action did not express his "true self" or what he "genuinely wanted". But it is also clear that he was still morally responsible for his action: he was appropriately subject to blame and censure from others, and his guilt was completely fitting.

Autonomy and moral responsibility can come apart, but the precise connection between them and how they differ remains murky. For example, John Martin Fischer suggests that one way moral responsibility and autonomy come apart is in their compatibility with weakness of will[7]. Moral responsibility is clearly compatible with weakness of will – that is, we can be morally responsible for weak-willed actions we performed against our better judgment. (Indeed, our better judgment is probably an additional demerit against us, since we clearly knew better!) In contrast, if one performs a weak-willed action, one is necessarily *not* autonomous, since by definition you acted against what you took to be best and therefore against what was most aligned with your "true self"[8]. This seems imminently plausible. However, Nomy Arpaly and Timothy Schroeder have argued that the "true self" should not be understood as what a person takes to be "the best thing to do", but with what is most thoroughly integrated into a person's

[7] From "Responsibility and Autonomy: The Problem of Mission Creep"

[8] In "Mission Creep", Fischer uses the term "true self" to refer to the self which grounds autonomy. He has subsequently informed me that he now thinks of autonomy in terms of the "ideal self", and would agree that the "real self" corresponds to what Arpaly and Schroeder are talking about.

character[9]. In such cases, a weak-willed action might be *more* reflective of a person's authentic self, and therefore more autonomous. (Interestingly, Fischer makes space for this possibility when he allows that the "true self" might not be the "rational self".) It is thus unclear whether autonomy is compatible with weakness of will.

I take it the issue here goes to the very heart of autonomy – what *counts* as "the true self"? Is it connected to one's reason? Is it one's self taken as a whole? Is it something else? It's also essential to note that Fischer explicitly connects the idea of a "true self" to autonomy, while Arpaly and Schroeder are explicitly talking about moral responsibility. The fact that the notion of a "true self" is readily used in both kinds of discussions indicates the intuitive plausibility that either internality or identification could ground moral responsibility *or* autonomy; an intuition which explains why discussions of moral responsibility and autonomy tend to dovetail and then be easily confused. (This is, of course, a version of Fischer's point in "Mission Creep".) Indeed, while Arpaly and Schroeder are talking about moral responsibility, the "whole self" they describe sounds remarkably like an *authentic* self, and as noted above authenticity seems to be closely connected to certain kinds of autonomy. For our current project, discussions of moral responsibility, while importantly distinct from discussions of autonomy, may nevertheless provide useful insights.

Two more ways to potentially differentiate moral responsibility and autonomy are (1) that autonomy sets a higher bar than moral responsibility and (2) that while moral responsibility does not entail autonomy, autonomy entails moral responsibility. Again,

[9] "Praise, Blame and the Whole Self"

both seem quite plausible. While (typical) adults will be morally responsible for most of their actions, it is not guaranteed that most adults will be autonomous most of the time, because this seems harder to achieve. The example of Andrew Clark above lends support to this idea. Given that one can be morally responsible without being autonomous, it seems clear that moral responsibility does not entail autonomy. But the converse – that autonomy entails moral responsibility – has intuitive support. It would seem that if we meet the higher demands of agential governance which are required for autonomy, we will almost certainly meet the threshold for moral responsibility. This is Fischer's view. However, not all agree. For example, Marina Oshana argues that moral responsibility and autonomy are grounded in different conceptions of rationality – one thick, one thin – and therefore that they come apart completely such that one could be autonomous and yet not be morally responsible[10]. Roughly, moral responsibility requires being sensitive to shared evaluative norms such that you see these norms as giving someone (including yourself) a potential reason to act. Oshana calls this requirement "normative competency". This need not mean that you act on these reasons, or even that you ultimately agree with these reasons. The idea seems to be that by seeing shared evaluative norms as even *potential* reasons, you demonstrate that you are a member of a moral community and are invested in living and interacting with others under a shared set of moral rules. Oshana argues that one can be autonomous *without* being normatively competent. Autonomy requires that your desires and motivations be authentically your own such that you genuinely endorse and stand behind them. It thus seems to require the ability to reflect on your motivations, to formulate plans based on the

[10] "The Misguided Marriage of Responsibility and Autonomy"

results, and to carry out these plans. But none of this requires being sensitive to normative considerations. On Oshana's view, one can be both autonomous and a "moral idiot".

Oshana's work suggests that autonomy need not be conceived as setting a higher bar than moral responsibility, but simply a different kind of bar. In turn, this suggests another question: while autonomy is clearly different from moral responsibility, is it nevertheless related in a significant way to the same sorts of capacities that ground moral responsibility? Or is it, as Oshana suggests, simply a different thing altogether? For present purposes, all we need to be certain of is that autonomy is not the same thing as moral responsibility.

1.2: Identification and Internality

While autonomy and morality are easily distinguished, historically the same has not been true of identification and internality. This is because originally identification was meant to provide the criteria for internality. *Internality* is the idea that for persons there are some desires and motivations which are more truly "our own" than others. (Notice that this sounds very similar to autonomy's requirement that desires be "genuinely our own".) A desire is *internal* to me when it cannot be separated from who I am as an agent or viewed as an alien force which is simply passing through me. It is supposed to be part of "where I stand" as a practical reasoner and actor. The basic idea is that while each of us have various desires and motivations, it seems that only some of these are defining of who I am as a person. Internality seems central to personhood. It has also been taken to be central to moral responsibility, as we will discuss in a bit. The key question is: *how can we determine when*

something is internal to a person? This question is at the heart of "Freedom of the Will and the Concept of the Person", where Frankfurt first introduced the term "identification".

Identification plays a central role in both Frankfurt's account of acting freely, meant to ground moral responsibility and his conception of personhood. Frankfurt famously differentiates first order desires, which are directly concerned with action, and second order desires, which are concerned with first order desires. Second order volitions are a particular kind of second order desires: namely, desires that the first order desire they are concerned with actually be what moves one to action. Since one's will is (by Frankfurt's definition) simply the effective desires which drive you to act, a second order volition is essentially the desire that you have a certain kind of will. Someone identifies with a first order desire when they have a second order volition that this desire move them to action. It is the fact that we identify with certain desires – which is really the fact that we care about what sort of will we have – which makes us *persons*[11]. And since my second order volitions show the particular will I care about having, they also reveal *the kind of person* I am. Thus, according to Frankfurt, identification seems to be the criteria which pinpoints the desires which are defining of me – in other words, identification shows what is genuinely *internal* to me.

Much of what Frankfurt says makes intuitive sense. While we all have a hodgepodge of desires simmering around inside us, persons seem to be more than just this

[11] To clarify: the way I understand Frankfurt, it is indeed the person's *actual state* of caring, and not just the capacity to care. A wanton is fundamentally someone who simply does not care about the will he has. If he were to start caring, then he would become a person; but he must *actually* care, and not simply have the capacity to care.

hodgepodge. Since we have the ability to reflect on the desires we find ourselves with, to evaluate and then decide which desires we actually want to act on, not all these desires are equal. Some are given more weight than others. Since it is the very ability to differentiate between the various ingredients of the hodgepodge, to pull some out and make them important in ways others are not, which makes us *persons* in the first place, it makes sense that this very same act would be what defines me as the *particular* person I am. In other words, it makes sense that the act of *identification* would be what determines a desire as either *internal* or external to me, even though all desires are obviously "mine" in some sense.

It's important to note that in this initial paper, Frankfurt strongly suggests that identification is typically (but not necessarily always) a conscious act[12]. He states that "the unwilling addict identifies himself, however, *through the formation of a second-order volition*, with one . . . of his conflicting first-order desires" (emphasis added). He goes on to argue that "When a person identifies himself *decisively* with one of his first-order desires, this commitment 'resounds' throughout the potentially endless array of higher

[12] I say "strongly suggests" because his view *is* a bit ambiguous. Here's a passage from the paper: "Examples such as the one concerning the unwilling addict may suggest that volitions of the second order, or of higher orders, must be formed deliberately and that a person characteristically struggles to ensure that they are satisfied. But the conformity of a person's will to his higher-order volitions may be far more thoughtless and spontaneous than this." This could be interpreted as Frankfurt saying that second-order volitions can be unconscious. However, the passages I quote in the above in the body of the paragraph suggest that second order volitions require something more robust from the agent; I don't simply *discover* them in myself, I *do* something to implement them. I am therefore taking Frankfurt to mean that there is conscious *endorsement* without there necessarily being conscious *deliberation* behind this endorsement. In later papers, Frankfurt does claim that second order volitions can be unconscious such that I "discover" them in myself in a more passive way.

orders" (Frankfurt's emphasis), such that there is no doubt about whether he genuinely identifies with the first order desire – that is, whether the desire is truly his. This need not involve a drawn-out deliberative process, but it seems clear that in this original paper "identification" meant *explicit, conscious* rejection or endorsement of first order desires. I will continue using "identification" with this meaning[13].

It has been convincingly argued by multiple people that identification and internality can, and often do, come apart. In a series of papers, Nomy Arpaly and Timothy Schroeder argue that something can be internal to an agent even if he has consciously rejected it and doesn't identify with it[14]. "Where the agent stands" is not determined by just one privileged aspect of the psyche, such as second order volitions or reason more generally, but by considering the agent's motivational structure as a whole[15]. Bernard Berofsky argues for the same point, and also articulates the reverse relation: simply identifying with a motivation is not enough to make it internal to you[16]. There needs to be confluence between our rational, reflective selves, and our larger patterns of behavior and emotion. Angela Smith argues that even in cases where I consciously reject a motivation,

---

[13] Some continue to use "identification" in a sense which seems simply synonymous with "internality". For example, David Shoemaker claims that identification is mainly a passive thing. For the sake of clarity, I am only going to use "identification" in the stricter sense that requires *active* or conscious identification.

[14] " Praise, Blame and the Whole Self", "Identification and Externality"

[15] While Arpaly and Schroeder do not use the term "internality", they speak in terms of the whether an action expresses an agent's self, so it is clear that they are nevertheless dealing with internality.

[16] "Identification, the Self, and Autonomy".

it can still embody my implicit evaluative judgements, and so still be embedded in where I stand as an agent[17]. Agnieszka Jaworska spells out the wedge perhaps most clearly:

> "the philosophical usage of the term 'identification' often does not discriminate between two senses: the ontological sense I have just stated - where the task is to pick out attitudes that properly belong to the agent from the sea of happenings in the agent's psychic life - and the psychological, subjective sense focused on how the agent perceives aspects of his psychology and whether he regards them as his own. . . Internality is meant to track attitudes that "speak for the agent," that are fully the agent's own, regardless of the agent's own view of the matter, which may, after all, be mistaken."[18]

Jaworska goes on to reserve the term "internality" exclusively for ontological identification – that is, what is *actually* internal to and defining of an agent, and not simply what he believes or wants to be internal to him. This is how I have been understanding the term. While Jaworska is the only one to talk explicitly in terms of internality, it is clear that all five authors mean to say that just because we do not consciously *identify* with a desire does not mean we are "off the hook", for the desire may still be imbedded in us such that it can truly be said to represent us.

The common theme of all these views is that where you stand *qua* agent, and what is most characteristic of you *qua* person, is revealed in the broader and more enduring patterns of your judgments, actions, and emotions, not simply what you consciously endorse or reject. Arpaly and Schroeder argue that what matters is how "integrated" an attitude is into the whole self, and sketch a series of examples which show that (1) the more a particular motivation plays a major role in our life, (2) the more satisfying we find it, and

[17] "Conflicting Attitudes, Moral Agency, and Conceptions of the Self"

[18] "Caring and Internality"

(3) the more it remains unopposed by similarly "deep" motivations, the more it seems to be a genuine characteristic of the self. Berofsky says that "fundamental desires that are deeply satisfying, that play a central role in the explanation of my verbal and nonverbal behavior and experience, and the dislodging of which would be deeply disruptive" are internal such that we cannot extract them from ourselves by mere reflective decree. Smith says it is the motivations "that actually seem to be structuring and motivating [a person's] patterns of thought and feeling" which are truly her own. And Jaworska speaks of one's "stance towards the world," and of attitudes "constituting one's standpoint towards the world". Such "stances" and "attitudes" are deeper and more pervasive than second order volitions which are focused on a single first order desire. A standpoint shapes how we are inclined to interpret, or what we tend to notice, about various situations; how we are disposed to emotionally react to certain things; what we will likely see as reasons for acting; and, of course, how we are likely to act. All five authors point to the importance of affective, emotional, interpretive, and volitional patterns – patterns which seem to involve the whole self (to borrow a phrase from Arpaly and Schroeder), and not simply our rational, reflective selves.

Internality was meant to express the idea that as persons and agents, we are not simply a "psychic stew" of desires. We have a more definite character than this, and it seems that some of our motivations are more expressive of who I am, *qua* person, than others. Identification is the conscious attitude of *viewing* something as internal to you, or perhaps deciding that you *want* something to be internal to you. But these two can come apart. Like so many other forms of self-knowledge, we can be mistaken about ourselves,

and we can misjudge what is "really" our own. Identification is not an adequate criterion for internality, and the two must be kept separate.

1.3: Can Either Internality or Identification Ground Autonomy?

It seems fairly clear that internality cannot be enough for autonomy. Something may be internal to a person in precisely the sense that it is embedded in her emotive, cognitive, and motivational patterns, but if she reflectively rejects this thing and resents the power that it has over her, we could not say she is autonomous with regards to it. To borrow an example from Agnieszka Jaworska, consider a woman who still cares deeply for her abusive ex-husband. She finds herself worrying about him, and even catches herself contemplating stopping by his place to make sure he's taking care of himself (eating healthy food, running laundry, and the sorts of things she used to do for him). She hates that she can still care so deeply for a person who treated her so wrongly, and she wishes she could surgically remove this part of her. In this case we would want to say both that her care for her ex-husband is internal to her, and that she is not autonomous in so far as she does still care. Two things to note: in such cases where internality and autonomy diverge, the fact that the motivation is not autonomous does not, of course, make it any less internal. And in such cases, the agent is still fully morally responsible for any actions which come from this non-autonomous motivation.

Could identification provide the relevant criterion for autonomy? Even if identification *is* necessary for autonomy, it is nevertheless not sufficient. Being autonomous means that you actually *act* in ways which are self-governing. If you identify with a motivation (say, exercising regularly) but you fail to act on this motivation, you not

autonomous. In asking whether identification is the relevant criteria for autonomy, what we are really inquiring is whether identification is what sets the relevant *standard* for autonomy: the measure by which we judge if each person is acting (or is in general) autonomous.

There are intuitive reasons to think that identification is the relevant standard for autonomy. We believe that autonomy requires a kind of self-governance and self-control which goes beyond simply acting on the motivations one happens to have. It requires that the person take a more substantial stand on their desires, cares, interests, and so on. In other words, autonomy seems to require our considered, reflective judgment about the motivations we want to be a defining part of us. Identification – at least in the sense that I have been using it, where it involves active, conscious reflection on oneself – seems to fit this bill.

The problem is that having reflectively endorsed a motivation does not seem to be enough for autonomy. The worry is that the process of reflection itself might be (mis)shaped by illicit forces, such that even if we identify with something, the process by which we come to identify with it might be tainted by pernicious influences. In other words, the very process which is supposed to guarantee autonomy may be un-autonomous. This is a criticism that many feminists give. The classic example is that of a woman living in a patriarchal and conformist society who reflectively and consciously decides that she wants to fulfill the traditional role of obedient housewife. Now, a general theory of autonomy should not automatically rule out certain life choices as unautonomous. The worry behind this example is not *what* the woman chooses, but the reasoning process *behind* her choosing

it. Let's say this woman is young and attending college, and is deciding whether to focus on her studies or on finding a husband. It is likely, given the society she has grown up in, that she has deeply internalized expectations and norms surrounding gender such that she cannot help but value herself in terms of whether she lives up to the role society has set out for women. She looks around and sees that men and women *are* clearly different; after all, they act different in many ways. She believes that simply because women are better suited for caring for families and men are better suited for the higher stakes world of public life does not mean that one is better than the other. She may look around and find herself genuinely pitying woman who have "chosen careers over family", who have "lost their unique feminine charm and become too much like men". She thus consciously and actively decides that she wants to get married, and will drop out of college as soon as she finds a suitable fiancée.

This example works so well because it is clear that the forces shaping this woman's reflective process are autonomy undermining. The idea that men and women are naturally and necessarily better suited for certain things means that she automatically discounts certain possibilities for herself; the idea that women are most valuable and most personally fulfilled when they are loving wives and mothers shapes how she values herself (and what she values for herself); and she is unaware of the ways in which these gender roles actually serve to keep her disempowered. When societal pressures close off certain options and values, they are already autonomy undermining, but this example has the additional factor of closing off certain options in values in a way that locks the woman into a subordinate position. Even if she identifies with her decision, her decision was not made autonomously.

Thus identification in its most active form of reflective endorsement is not enough for autonomy.

Note that this example also demonstrates that the combination of identification and internality cannot be enough for autonomy. If this woman's decision to become solely a wife and mother sets into motion a complex pattern of volitions and emotions centered around pursuing and then living out this role, it will be fair to say this is internal to her. She also identifies with it. Nevertheless, she will not be autonomous.

Notice also that none of this discounts the initial intuitions in favor of seeing identification as relevant for autonomy. It is still possible – perhaps even likely – that identification is a *necessary* ingredient for autonomy, if not a sufficient one. This is a substantial question which I will save for chapter 3. For now, we need only conclude that autonomy cannot be seen as bottoming out in either internality or identification, and so any satisfactory conception will need to go beyond either of these concepts.

Now that we have shown autonomy to be separate from and non-reducible to moral responsibility, identification, and internality, we can turn to delineating one particular kind of autonomy: philosophical autonomy. This is the goal of Chapter 1.

Chapter 1: Demarcating Philosophical Autonomy

Autonomy is typically seen as a form of activity which elevates humans above other animals. While animals act on their instincts, humans choose their actions. This is the distinctive form of freedom at the core of autonomy: *self-governance*.

The tricky part is making sense of "self-governance". For such an important concept, autonomy is remarkably equivocal. Joel Feinberg has identified four different meanings "autonomy" can have[19]; Nomy Arpaly has identified eight[20]. Gerald Dworkin has lamented that "[a]bout the only features held constant from one author to another are that autonomy is a feature of persons and that it is a desirable quality to have"[21] – and in fact even the latter has been contested!

In this chapter, I will demarcate what I take to be self-governance in its fullest form: philosophical autonomy. First, I will emphasize the importance of differentiating philosophical autonomy from political autonomy. Then I will argue that there are two ways of interpreting the idea of "self-governance", and that only one of these can ground an account of philosophical autonomy proper. Finally, I will consider relational accounts of autonomy, and argue that only some of this literature will be directly relevant to a notion of philosophical autonomy.

---

[19] "Autonomy"

[20] "Which Autonomy?"

[21] "The Concept of Autonomy"

Section 1: Political Autonomy vs. Philosophical Autonomy

Autonomy is not just a philosophical term of art: the concept plays a key role in modern liberal societies. The foundational belief of liberalism is that each individual should have the freedom to make their own decisions about the lives they lead, the values they have, and the interests they want to pursue. It is assumed that all full-functioning adults have the ability to make these decisions for themselves – to be *minimally* autonomous or self-governing – and therefore that they should be guaranteed the right to make these decisions. I will call this kind of autonomy *political autonomy*. Political autonomy is a standing people have based on their capacity for minimal autonomy, and which grounds claims to a set of autonomy-protecting rights – rights intended to grant individuals adequate social "space" to exercise their minimal autonomy. (I say "claims" because while all qualifying adults *should* be guaranteed these rights, not every society will actually grant all qualifying adults these rights).

Political autonomy is used in legal, medical, and (of course) political contexts. It is relevant for discussions about the kinds of rights necessary for ensuring people's freedom to exercise minimal autonomy (freedom of speech, freedom of association, freedom of religion, to name a few examples). It grounds the basic respect we owe to others such that we do not treat them paternalistically, and so grounds the authority that people have to make medical decisions for themselves. It is also assumed that only (minimally and thus politically) autonomous people should be prosecuted to the full extent of the law – people who are severely mentally disabled or insane are not held to be responsible in the same way.

Political autonomy is not what I will be concerned with in this dissertation. Political autonomy is importantly distinct from *philosophical* autonomy, which is the state of being self-governing such that your values, carings, interests, projects, and beliefs come from you. In common parlance, these things are "genuinely your own". (While this is not an *in*accurate description, we must keep in mind that for philosophical autonomy, the criterion for genuineness is that I am in *control* of it. "Genuinely one's own" may connote authenticity-centered accounts, but these are not what we are concerned with.) What the criteria are for "genuinely your own" is one of the key questions for an account of philosophical autonomy. But however we parse the requirements, it will be much more demanding than the minimal autonomy which grounds political autonomy. Political autonomy is in many ways a negative set of norms: it tells us the ways we cannot interfere with other people. It is meant to give all full-functioning adults space to decide their own lives. Philosophical autonomy probes beyond this and asks whether these very decisions are free such that the person can be said to be robustly self-governing. Whereas we can assume that every typical adult has minimal autonomy, and therefore political autonomy is granted (or should be granted) to every typical adult, it is entirely possible that the achievement of philosophical autonomy is rare.

Here's an example to demonstrate the difference between the two. Let's say Olawale is a young man with full mental capacities. He has been raised as a Jehovah's witness in a small, tight knit community of Jehovah's witnesses. One day he's in a car crash, and the doctors say he needs a blood transfusion. Based on his religious convictions, he refuses. Since he is autonomous in the sense required for political autonomy, it would

be a wrongful violation of his rights for us to ignore his explicit decision and give him the blood transfusion anyway, and we need to respect his choice. It is a separate question whether Olawale is philosophically autonomous. Since he has lived his (relatively short) whole life in a single, small community, where one set of beliefs and values has been consistently conveyed to him, it is unclear if he has had the resources to seriously consider alternatives for himself.

To be clear, political autonomy and philosophical autonomy have different purposes and different scopes, and they should not be viewed as in competition with each other. Even if it turns out Olawale is not philosophically autonomous, this does not mean he no longer has the right to make his own medical decisions. The purpose of political autonomy is to set the public norms which ensure that people are given the freedom to make their own decisions. This is valuable and necessary separate from whether someone meets the standards of philosophical autonomy.

Whereas political autonomy sets a standard of public norms, philosophical autonomy sets a personal ideal. This means that philosophical autonomy can be thought of as a more intimate thing; it concerns my relationship to myself. This intimacy is also shown in the fact that it is only appropriate for me to provide feedback on how philosophically autonomous people are if I am particularly close to them.

To articulate the difference between these two conceptions more fully, it may be useful to refer to Feinberg's piece "Autonomy". Here, Feinberg distinguished between four senses of autonomy: 1) the capacity to govern oneself, 2) the actual condition of governing oneself, 3) an ideal of virtue, and 4) the authority to govern oneself (which I understand to

be the right to make decisions for oneself). Philosophical autonomy is clearly an ideal, so it maps onto (3). Feinberg's characterization of (philosophical) autonomy as an ideal of virtue indicates what I suggested above: that it is a more intimate concept, connected to an individual's personal flourishing. Political autonomy refers to the standing of a full-functioning adult such that he is the proper bearer of rights. This standing means he is the ultimate authority in his own life. Thus political autonomy is best understood as referring to this standing of authority (4) which is protected by rights. The capacity to govern oneself, i.e., the capacity for minimal autonomy (1) is what grounds this authority and its resultant rights (i.e., political autonomy). In turn, the rights political autonomy guarantees are meant to support the actual condition (2) [22]of being self-governing.

It is essential that we not confuse political and philosophical autonomy. We've already noted that political autonomy should not be assigned on the basis of philosophical autonomy; relatedly, we must not critique philosophical autonomy for setting standards that are too high for political autonomy. The two kinds of autonomy have different scopes and purposes, and so if a theorist argues that a person fails to have philosophical autonomy, this does *not* mean that she believes this person should be denied the full accord of rights given to all full functioning adults. This is a mistake I believe John Christman makes in his piece "Relational Autonomy, Liberal Individualism, and the Social Constitution of

[22] The broadest of Feinberg's four distinctions is (2), the actual condition of autonomy. As Feinberg notes, while having the standing of authority (4) is all-or-nothing – if you have it, you have the same rights as everyone else – to what extent you have achieved the condition of autonomy is a matter of degree. This is simply a reiteration of a previous point: we must differentiate between the (minimal) condition of autonomy relevant for political and legal rights and the (much harder) condition of autonomy in the sense of philosophical autonomy.

Selves". Here, Christman criticizes Marina Oshana's view that autonomy requires one not to be involved in social relation of subservience[23]. He argues that this sets too high a standard, such that large groups of people – including those in oppressed groups – would be excluded from the status and rights of *political* autonomy. Christman is right to worry that setting the standard for political autonomy too high would strip the concept of its "usefulness as a marker of the (equal) moral and political status" of persons, but the only reason he has this worry is because he conflates political and philosophical autonomy. We must not hold political autonomy to the standards of philosophical autonomy, but we also must not hold philosophical autonomy to the status required for political autonomy. To do this would strip philosophical autonomy of its usefulness as an aspirational ideal.

Although we need to differentiate between political autonomy and philosophical autonomy, the two are interrelated. The condition of being autonomous admits of degrees, and we can conceive of the minimal autonomy which grounds political autonomy as being at one end of this continuum and philosophical autonomy at the other. While philosophical autonomy is not the same as minimal autonomy and is much harder to achieve, we can think of it as the fullest realization of the capacities involved in minimal autonomy. Thus I take it that political autonomy, which allows for the full exercise of minimal autonomy, is conducive to (and perhaps necessary for) philosophical autonomy.

[23] Oshana, "Personal Autonomy and Society"

For early liberal thinkers, it might have been thought that political autonomy would guarantee fuller autonomy[24] (but perhaps not philosophical autonomy, since I suspect that few theorists had this fullest kind of autonomy in mind). Give each person the freedom to do as he wants, and his choices will automatically reflect his "genuine" self, meaning he will be self-governing in the way required for any kind of autonomy we might be concerned with. Since we live post-Freud and in the wake of critical theory, we know that simply being guaranteed outward freedom does not mean we have inner freedom such that the desires we freely act on our genuinely our own.

But even now, there's a reason why people might conflate basic autonomy with philosophical autonomy: the typical experience of ordinary people. A normally functioning adult just going about her life, engaging in mundane actions, typically takes herself to be the one who is in control of her actions. Whether it is choosing a banana over an apple, or choosing a career as a publicist over a journalist, she takes it she is in charge. Most of the time, other people also assume this about her. This assumption that she is in charge is what gives her actions authority in the sense that others need to respect her decisions and choices; they come "from her". This idea that actions "come from" the person is frustratingly vague, but it the core idea for any notion of autonomy. We believe that political autonomy is

[24] For example, Locke said: "The natural liberty of man [i.e., living in a state of nature] is to be free from any superior power on earth, and not to be under the will or legislative authority of man . . . The liberty of man in society is to be under no other legislative power but that established by consent in the commonwealth . . . [this is] A liberty to follow my own will in all things where that rule prescribes not, not to be subject to the inconstant uncertain, unknown, arbitrary will of another man." Locke indicates here that the will of a man found in the state of nature carries over to a state of society, such that so long as he is free from the "arbitrary will of another", he is unproblematically following his own will. (From "Two Treatises of Government", second essay.)

undermined when a person is coerced, manipulated, or brainwashed, *because* such things undermine her ability to be in control of her actions; the actions no longer seem to "come from her". As we have seen, the idea that actions "come from" the person in some uniquely robust way is closely connected to what it means to *be* a person (and the particular person you are) in the first place. This idea of an action "coming from the person" can easily be thought to ground political autonomy *or* philosophical autonomy.

The gap between the politically autonomous self and the philosophically autonomous self is explained by the fact that my actions can indeed "come from *me*" (in either the minimum sense of coming from my agency or in the thicker sense of coming from traits internal to me), while *who "I" am* is itself not under my control. The difference between these two can be forgotten because of an ambiguity in the very term "self-governance". This is the topic of the next section.

Section 2: Two Notions of "Self-Governance"

"Self-governance" means that *I give myself the law on which I act*. This is a frustratingly vague idea: what does it mean to give myself the law? I believe that there are two common ways of parsing this idea which we must be careful to distinguish.

"Self-governance" might mean that I govern *from* the self. On this view, I already have a substantial/contentful[25] self – perhaps a semi-inconsistent one, but nevertheless a robust enough self to serve as a starting place – and I self-govern when my actions, choices, and overall life are directed from this self. Much depends on how exactly to characterize

[25] By "substantial" I do not mean a metaphysical substance (e.g., a soul). Rather, I mean that the self already has substance to it – i.e., already has certain contents. In this dissertation I will use these two terms, "substantial" and "contentful", interchangeably.

this already (semi)substantial self, but the but the core idea is that I am autonomous when I govern my actions from the set of motivations (desires, values, interests, etc.) which are essential to and defining of who I am. In this case, "giving myself my own law" means that I act in ways that are congruous with my substantial/contentful self. This is why it is self-governance – the law comes from *me* because it comes from what is *defining of* me, and therefor has the proper authority.

Multiple people working on autonomy have given accounts of governing from the self. One clear example is found in Frankfurt's more recent paper "Autonomy, Necessity, and Love". This is where Frankfurt introduces his concept of "volitional necessities", and argues that the things that we cannot help but will express the deepest, most integral parts of us. As he says, "The essence of a person . . . is a matter of the contingent volitional necessities by which the will of the person is as a matter of fact constrained. . . Our essential natures as individuals are constituted, accordingly, by what we cannot help caring about." In other words, the self I have is essentially defined by my carings – in particular, the carings which are so deeply rooted, I cannot help but have them and express them. Frankfurt goes on to argue that "A person acts autonomously only when his volitions derive from the essential character of his will". When I express what is most deeply ingrained in my self – the affective, cognitive, and volitional matrix which is bound up with caring – then I act autonomously. Frankfurt's idea of autonomy as grounded in volitional necessities is a paradigm case of governing from the self. He does not go into how I acquired the self I do, or delve into the origins of these carings. They could be the result of socialization, or indoctrination, or pure coincidence. Indeed, he takes care to qualify them as "contingent"

volitional necessities. For Frankfurt, it does not matter where the self came from, or if we had absolutely no role in shaping it (though presumably most of us had at least some role). What matters is that I *have* a self, and that certain things are so integral to this self as to be defining of it. I am autonomous when I act from this self.

Another example, which builds off Frankfurt's work, is found in David Shoemaker's "Caring, Identification, and Agency". (Though "autonomy" is not in the title of this piece, Shoemaker makes clear that it is autonomy he is concerned with.) Shoemaker's addition to Frankfurt is to greatly expand the scope of volitional necessities. According to him, all of my cares are at any given moment beyond my control: I cannot help but care about what I do right now. I can, of course, endeavor to change a caring I wish to rid myself of, but this will take time and is only something I can bring about indirectly: I cannot stop caring simply because I decide not to care. But this means that any time I act on the basis of a caring, I am volitionally necessitated, and I act freely (i.e., autonomously). Shoemaker's justification for this account of autonomous action is exactly the same as Frankfurt's: who I am is based on my "nexus of cares", so when I act based on these cares, this action is grounded in *me* and thus seems to be a clear case of self-governance. As he puts it, "if who I am as a developed agent is both made possible by, and is a function of, my emotional commitments—then what else could provide the authority for self-determination than my cares?" Shoemaker's account is therefore a clear example of self-governance as governing from the self.

This is further seen in the role he allows for reflection. While he does think it can play some role, he envisions this mainly as a clarificatory one: reflection allows me to see

what it is I *already* care about, and, by providing such clear-sightedness, it prevents me from being self-defeating. As my self-knowledge grows, "as [I] come to know more and more about what it is [I] care about and to what degree [I] care about it", I will be better able to consistently will what is most aligned with my cares, and thus act autonomously. In short, reflection thus does not play an *active* role in determining *what* I care about, but simply a *passive* one in uncovering what I already care about. I do not determine the self I have, but discover the self I *already* have.

This aspect of Shoemaker's account was worth going over in some detail because it exemplifies a common aspect of similar governing-from-the-self accounts. While such views take the self as pre-given in some sense, often they discuss how "discovering" this self (i.e., what the person really wants or values) will likely take some work. Further, it may take some work to develop a cohesive, well-ordered self out of the various desires and so forth the self already has. Thus "governing from the self" may still require doing work *on* the self. Nevertheless, the point remains the same: the core components of the self are already there, even if we have to discover and organize them.

We have discussed how "self-governance" can be thought of as governing from an already existing self. But "self-governance" could be thought of a different way: as governing *the self*. This means that I don't just *discover* my "true", substantial self and then determine my actions in accordance with this self; instead, I determine my self such that I control (in some to-be-specified sense) *what* I want and care about. I decide what content is going to essential to me. When autonomy is thought of this way, "giving myself my own

law" means that I determine which ways of acting will be congruous with myself – I decide what will be defining of me. I decide, in short, what my self is going to be.

The key difference between these two understandings of self-governance is whether they see the self as (more or less) simply given or pre-determined in an unproblematic way. Governing-from-the-self accounts assume that there already is a self. Perhaps this self is somewhat inconsistent, or maybe parts of it are repressed, but it is nonetheless *there*, even if we have to do some work to get in touch with it or learn about it. On such view, this pre-formed, pre-given self serves as a solid ground for authorizing motivations, and thus as a solid ground for governing from the self. In contrast, governing-the-self accounts hold that the self I already have cannot simply be assumed to have authority. It must create this authority for itself by deciding what self it wants to be. The difference between the two accounts is *not* whether I have already have a fairly solid self, but whether this self can serve as adequate grounds for the authority self-governance requires.

Unsurprisingly, these two interpretations can be interrelated: governing the self might be conceived as a subset of governing from the self. Why? Because in order to decide what kind of "self" to have, it seems likely that I need to already have some self from which to decide. It is a familiar critique that there is no completely neutral point from which to survey my options. Perhaps even more relevant to the case at hand is that if there *was* such a neutral point, this neutrality would seem to hinder rather than help me make a decision. How would I decide what self to have? What criteria or relevant considerations would be available to me?

Even if governing the self does come down to a version of governing from the self, the distinction remains relevant. In fact, it is essential. Accounts which only emphasize governing *from* the self cannot work as an account of philosophical autonomy; governing *the self* must play the key role. This is because the self a person (originally) has will always be largely a product of socialization, familial upbringing, and raw psychological dispositions. While it seems impossible to strip a person of these things entirely (and likely there wouldn't be a person left if we did) the fact remains that the more the person's self is simply decided for her by such forces, the less philosophically autonomous she is. Philosophical autonomy asks us to be as in control of our own selves as possible. An account which takes a pre-given self as an unproblematic ground is therefore untenable for this kind of fullest autonomy. Governing from the self accounts will likely work for what I have called "fuller", authenticity-centered autonomy, but they will not work for the ideal of philosophical autonomy.

To be fair, I am under-describing the process of socialization and parenting, and it therefore might seem that I am underselling accounts of governing from the self. Nonetheless, my above point still stands: let me explain in a more nuanced and accurate way. Raising a child is not simply a matter of shaping them like clay; it is essentially interactive. It calls on active responses from the child. In particular, a parent guides their child in the development of self-directed abilities: things like self-control when they are young, and as they get older, the ability to critically reflect on their own desires and motives. One common and essential way this happens is as a natural development of the parent holding their child to normative standards. When a parent communicates to their

child “Well-behaved children don’t do this, they do that”, or “A good person does x and believes y”, the child learns to evaluate himself according to these standards. (Presumably the more sophisticated skill of “evaluate” comes later; initially, the child might simply *see* himself in these normative terms.) He thus learns (or at least starts down the path of learning) the ability for self-reflection. This is not simply a matter of the parent forming the child however they wish. It *presupposes* a complex form of agency in the child, and it nurtures their development into a mature self-directing, self-reflective agent. In other words, socialization and upbringing may shape the person, but they do not shape him brutely; they shape him in a way that requires and promotes his complex agency.

The problem is that the contours of one’s agency – the values and beliefs that guide the ways it is exercised, the norms and standards a person holds himself to and uses to reflect critically on himself – all these factors are often simply given to the person. We learn the values and morality of our parents and culture; we are instilled with an understanding of what is “common sense”, what is natural and therefore to be taken for granted; we learn what roles and dispositions are suitable for which kinds of people. Our agency may be complex and self-directed, and it may even be self-reflective; but unless we are able to reflect on the very norms and worldviews which shape our agency – which determine the basic limits of our self-direction and self-reflection – our very self, complex as it is, is largely decided for us. It therefore cannot serve as adequate grounds for philosophical autonomy.

To further explicate the difference between governing-from-the-self and governing-the-self, and why any satisfactory account of philosophical autonomy must

involve the latter, it may be useful to refer to the work of Diana T. Meyers. In her book *Self, Society, and Personal Choice*, she argues that autonomy must be thought of as having 3 aspects: self-discovery, or the ability to differentiate authentic desires which are "genuinely yours" from alien or inserted desires; self-direction, or the ability to act on authentic desires[26]; and self-definition, or the ability to create and re-create an authentic self. Meyers argues that most contemporary accounts of autonomy emphasize self-discovery and self-direction, but forget the crucial aspect of self-definition. Accounts like Frankfurt's and Shoemaker's emphasize the need to reflect on what it is you really want (self-discovery), and they thus embody the intuition that oftentimes what we really want has been suppressed by factors like socialization. But by thinking that all we have to do to get to the authentic self is clear away the impacts of socialization, these accounts assume that each person already has a pre-given authentic self[27]. Meyers believes such a pre-given self cannot be an adequate basis for autonomy. Saying that one's "true self" is the self stripped of all socializing influences would put you at the mercy of whatever traits happened to be innate to you – even if you might find that some of these traits are undesirable and ones you would wish to change. And this doesn't seem to be autonomous

[26] I suspect that self-direction is mainly a question of agency – that is, how we actually move ourselves to act. This isn't to say that *without* self-direction we can be autonomous, but I suspect this means self-direction is not the real heart of autonomy.

[27] Stoljar and Mackenzie suggest that Meyers was misreading Frankfurt, and we should understand second order volitions in terms of self-constitution, and not of self-discovery. I actually think that Frankfurt's earlier work is plausible read this way. However, Meyers' point is still a potently relevant one, since it seems clear to me that in his later work Frankfurt moved closer to self-discovery accounts, and away from (at least potentially) self-constitution accounts. Thus even if Meyer's missed her initial target, it still landed on a different important target.

at all! As such, a conception of the self as already given, just waiting to be discovered, doesn't seem to be a helpful conception for autonomy. This leads to Meyers adding the aspect of self-definition, or the ability to create (and re-create) what the authentic self is in the first place.

As the discussion of Meyers already makes clear, I am certainly not the first person to recognize the difference between governing-from-the-self and governing-the-self (although the explicit label for the difference is mine). The implicit recognition of the importance of governing the self shows up in various forms throughout the literature. I would like to (semi-briefly) outline these various forms. Let's discuss what I take to be a "light" form first: much of the emphasis on reflective endorsement seems to be motivated by the intuition that simply having a motive which is undeniably a part of you is not enough to guarantee that whenever you act on this motive you thereby act autonomously. This intuition pushes us part of the way from conceptions of autonomy as governance-from-the-self towards conceptions autonomy as governance-of-the-self. The idea of reflective endorsement remains powerful precisely because it is difficult to see how one could be genuinely autonomous in acting on an internal motive if one has never even stopped to reflect on whether one endorses this motive. This seems like a nascent form of governing the self. "Reflective endorsement" could also mean a version of self-discovering – realizing what it is I *really* want and then endorsing that. Nonetheless, some accounts use "reflective endorsement" as actively *deciding* what I want to give priority to, and have therefore recognized at least implicitly the importance of governing the self. (We will look at some of these accounts in Chapter 3.)

As mentioned previously, many theorists have pointed out that simply reflectively endorsing a motive is not enough for autonomy, since the process of reflection might itself be tainted. This is a critique given by many feminists, who point out that one's reflective endorsement can be substantially pushed in one direction by societal expectations, norms, and values. (For example: Paul Benson's "Autonomy and Oppressive Socialization" and Ann Cudd's *Analyzing Oppression*.) The example of the housewife given above is meant to demonstrate this. The worry is that the self I currently have, as embodied by my current patterns of valuing, desiring, and reasoning, is itself problematically decided by forces which are not me, and so any reflection this self undertakes is likely to be tainted by these autonomy-undermining forces. The need for the process of reflection to itself be subject to stricter norms thus pushes us further into the territory of governing the self.

A good example of an account which provides stricter norms of reflection is Gerald Dworkin's view in early papers such as "Autonomy and Behavior Control" (1976) and "The Concept of Autonomy" (1981). Dworkin argues that autonomy is identification (which he calls "authenticity") plus procedural independence. Dworkin understands identification as active reflection and endorsement of one's own motives. Thus, his view already has some rudimentary elements of governing the self (at least if we assuming reflective endorsement means actively *deciding* what it is I most want). However, Dworkin's addition of procedural independence gives his view more robust elements of governing the self. A person has procedural independence when her process of reflection and deliberation is not the result of manipulation, indoctrination, withholding of crucial information or options, or the suppression of her abilities for critical reflection. Now, the

addition of procedural independence does not require some additional form or "level" of reflection – that is, it does not require my reflecting on deeper parts of myself, such as why I might be motivated to endorse a particular motive. It only requires that the reflection occurs under the right sort of conditions. In this sense, it is not explicitly adding an additional layer of governing the self. Nonetheless, if a person were to take Dworkin's notion of procedural independence seriously, it could cause him to start reflecting on the deeper reasons why he is inclined to value certain things or give preference to certain kinds of desires. Once this process has been set into motion, there is a deeper form of self-governance going on; the person is no longer simply accepting things about himself but questioning them and gaining more insight into them. Dworkin's formula of "autonomy = authenticity + [procedural] independence" thus pushes us towards a fuller conception of governing the self. But as indicated, this seems to be only a potential result of Dworkin's view, and not a necessary part of it. Furthermore, even once a person does start reflecting on the deeper factors why he is disposed to reflect and decide in certain ways, this still does not give us insight into how he can positively go about governing the self. In short, we are not yet all the way to a robust form of governing the self.

Let me briefly mention two more authors who have implicitly recognized the need to differentiate between autonomy as governing from the self and autonomy as governing the self. First, in "Volitional Necessities" Gary Watson argues that Frankfurt's new account of autonomy based on caring simply cannot give us a necessary or sufficient condition of autonomy. He points out that what one "identifies with" in the sense of reflectively endorsing can come apart from what one "identifies with" in the sense of deeply caring

about. In cases where a person is subject to a genuine volitional necessity such that he cannot but act a certain way, and yet this volitional necessity goes against what he reflectively endorses, this indicates that there is a deep divide within the person's very agency. Watson's idea of identification as reflective endorsement appears to be a version of governing the self insofar as we *decide* on the self we want to be and the desires, values, and motives we want to act on. If we combine Watson's point with my above characterization of Frankfurt's volitional necessity as "governance-from-the-self", we can understand Watson's critique as saying that governing from the self simply cannot be the end-all of autonomy. Simply acting in ways which express deeply internal carings – simply governing *from* the self – does not mean we act autonomously, precisely because we might be acting against the decisions we have made for the selves we want to be. We might be going against the ways we have tried to govern *the self*.

Finally, I want to mention a point made by Marilyn Friedman in her paper "Autonomy and Social Relationships: Rethinking the Feminist Critique". Although Friedman poses this point as a critique of the usefulness of autonomy, I believe we should take it instead as a point in favor of an account of autonomy as governing the self. She argues: "Before encouraging people simply to be more fully and coherently what they already are, we should first think about what it is that they already are. At this historical juncture, rather than promoting autonomy, we might be better off urging that some of us change what we "really" are. . ." In this passage, Friedman is thinking of the ways in which societies with warped systems of value and justice lead to the development of warped selves. (She speaks in this passage specifically of men and how they are subject to "patterns

of socialization that lead [them] to focus obsessively on asserting themselves apart from or against others", so her point applies to all people living in such societies, and not just those in oppressed or devalued groups.) Friedman characterizes autonomy as the condition of people who are "fully and coherently what they already are"; that is, she straightforwardly characterizes it as governing from the self. She then gives the potent critique that some of us might have selves which are actually quite problematic; presumably in a moral sense, but also in the sense that these selves are being cut off from certain forms of human value. Friedman is writing from a feminist standpoint, and would want to claim that traditionally "non-masculine" traits such as emotions and relationships are a valuable and essential part of human life. Now, such critiques of the selves "we already are" may come from a standpoint unconcerned with autonomy, as Friedman suggests here. But ideally, the ability to critique and refine the self should be an essential *part* of autonomy. Certainly the ability to critique the socially constructed self must be a part of autonomy. Friedman's point once again shows the difference between governing-from-the-self and governing-the-self, and why the latter adds a form of value which the former cannot provide on its own.

I have argued that there are two ways of understanding autonomy as self-governance. We may think of it as governing from the self, in which case the authentic self is already largely given: we simply need to get in touch with, perhaps slightly re-arrange, and then act in accordance with this self. We "give ourselves the law" because the law comes from or expresses our essential selves, but we do not really decide on this self-law;

we simply discover it[28]. Alternatively, we may think of autonomy as governing the self, in which case the self we already have is not simply taken to be the essential self and thus as grounding authority: we therefore have to define and re-create the contentful self. We "give ourselves the law" because we actually *decide* that this particular law (or set of laws) will be defining of our substantial selves. I have argued that philosophical autonomy cannot be founded solely on governing from the self; it must mainly rely on governing the self. (What role governing from the self *can* play in philosophical autonomy is a remaining question.) Governing from the self may be an adequate and useful conception for other forms of autonomy, but it simply won't suffice for the full ideal of philosophical autonomy.

Of course, we are left with a big, looming question: What does it mean to govern the self? What does this involve? Since the self is always "pre-given" to a large extent, and we cannot create ourselves ex nihilo, what does it mean to define, create, or recreate ourselves? This is the key question for an account of philosophical autonomy. We have thus come upon the question, and the problem, which is at the heart of this dissertation. Before we begin to address this question, there is one more approach to autonomy we must consider which will help clarify the concept of philosophical autonomy.

Section 3: Relational Autonomy

Feminism has a complex history with the concept of autonomy. In the 1970s, it was regarded as an obvious conceptual tool for the project of liberation. Then, beginning in the

[28] Given governing-from-the-self's assumption that the pre-given self serves as adequate grounds for authority and its subsequent focus on *action*, I suspect that this is better thought of as an account of self-determination. Self-determination and self-governance are clearly closely linked, and sometimes seem to be used interchangeably.

1980s and through the 1990s, it came to be viewed with suspicion as feminists speculated it was too tied up with masculine-centric ideals inimical to feminist concerns. The criticisms which came out of this period led some feminist philosophers to try to rehabilitate autonomy. The new forms, with all their variances, have been dubbed "relational autonomy". In this section, I will briefly outline the core criticisms of feminists and assess their bearing on an account of philosophical autonomy. Then I will differentiate between two kinds of relational accounts, only one of which will have direct bearing on a conception of philosophical autonomy.

3.1: Problems with Traditional Notions of Autonomy

Following Maeve Cooke, we can fruitfully understand the feminist criticisms of autonomy as falling into two main camps[29]. The first set is centered around the theme that traditional conceptions of autonomy misrepresent what the self is and thus only value certain kinds of selves, discounting others. The second set is centered around the theme that autonomy problematically presupposes a unified "self" or subject to begin with. We'll start with the second, which we can call the postmodern critique, and which can be set aside quickly.

Postmodernists worry that there is ultimately no self which can be serve as the basis for autonomy. Various traditions dovetail into this view. Freud emphasized that the self is opaque and often self-deluding; Foucault believed that the subject is always the site of practices and discourses of power, and is perhaps wholly constituted by these practices; and Nietzsche is typically read as arguing that there is no unified "self" other than the

[29] "Questioning Autonomy: The Feminist Challenge and the Challenge for Feminism".

myriad drives the psyche contains. Many postmodernists go beyond mere skepticism that such a coherent self can be found; they worry that the very idea of a central, unified self is oppressive, since it requires that one cut off or suppress parts of oneself to achieve a recognized "self". They further worry that by prioritizing the coherent and rational self, traditional liberal projects have in fact been oppressive by denying selfhood to those who were deemed incoherent and irrational.

While postmodernists have many fascinating ideas (I am particularly sympathetic to the idea of being open to different and potentially contradictory parts of ourselves and others), insofar as it is an *ethical* project postmodernism seems self-defeating. Cooke lays this out succinctly in her piece "Questioning Autonomy". To summarize her point, insofar as we are committed to acting in ways that promote social justice, it seems we need to be able to act. We need to be able to take up a (semi-)coherent reflective standpoint from which we can say "X is wrong, and this why"; and we need to have some agency which is (a) not simply the result of the interplay of societal forms of power – that is, which is our own agency from which we can take a stand against these forms of power – and which is (b) unified enough such that we can intentionally act in productive ways. We thus have powerful pragmatic and ethical reasons for assuming there is *some* coherent self. It is undeniably true that historical liberal projects have been oppressive by using the very notion of rational self-hood to deny rights to those deemed not to meet the threshold. However, the obvious solution is not to throw out the idea of selfhood altogether, but to recognize the selfhood of those who have traditionally been denied it. Indeed, this seems a key part of any social justice project. Therefore, the postmodern criticism that the very

subject or self at the heart of autonomy does not exist in any substantial way is one we can set aside.

The first, less radical group of objections allow that there is a "self" but worry that autonomy has traditionally presupposed an inaccurate picture of this self. This inaccurate picture leads us to misconstruing what is required for autonomy and leads us to value only certain kinds of selves. The criticisms here are more concerning, and I will go over them more carefully.

The common theme of the following critiques is that traditional conceptions of autonomy are rooted in a problematically male-centric perspective; specifically, a perspective that emphasizes substantive independence and reason at the expense of interpersonal relationships and caring. The core target of these critiques is that traditional conceptions of autonomy view the individual atomistically; he is seen as "disembedded" from any social, cultural, and historical context, as essentially independent from others, and ideally substantially independent from others. One of the points of feminist theorists is that this is obviously false: all persons are in some degree a product of the particular context they were raised and live in, and all persons are intimately dependent on one another for their development and survival. This false view of the individual is worrying for its problematic consequences. There are six key criticisms.

Criticism #1: The atomistic view of the person naively assumes that an individual is automatically "free" to pursue whatever projects and values he wants so long as no external obstacles are impeding him. Such a naïve view is what leads one to believe that political autonomy will by itself ensure all more robust forms of autonomy; all we need to

do is guarantee the individual certain rights to ensure others do not interfere with him. Feminism brings into view that there are often subtler, more pervasive and insidious forms of oppression which need to be addressed before the individual can be "free" in the way idealized by autonomy (a central theme of this dissertation). This means talk of "rights" is woefully inadequate.

This is an essential point, and it is completely congruous with contemporary accounts of autonomy in general, and a philosophical understanding of autonomy in particular. In fact, the need for a philosophical account of autonomy largely comes out of such insights. The question of what makes my self genuinely "mine" only gets raised if we assume that the self I currently have might *not* be genuinely mine – that it may have been largely constructed for me in problematic ways. Thus, this criticism does not speak against philosophical autonomy but supports it. The work feminists have done in this area will almost certainly be valuable for developing philosophical autonomy.

Criticism #2: The next three criticisms are tightly interwoven. The first (second overall) is that traditional accounts of autonomy tend to emphasize self-sufficiency and independence from other people, and by doing so devalue relationships, cooperation, and intimacy. This is a problem because it diminishes an aspect of human experience which seems essential to a good life. By doing so, it leads to warped individuals. Relationships with others are a source of rich meaning, and the ability to form such relationships and express care and intimacy are essential to a well-adjusted and fully realized person.

These are excellent points, but once again they are not inherently inimical to an account of philosophical autonomy, nor to contemporary accounts of fuller autonomy.

Philosophical autonomy does not emphasize substantive independence from others; rather, it emphasizes a particular relationship to oneself such that one is in control of how one interacts with the world and with other people. The target of this criticism – the substantially independent and self-sufficient person – seems to be more of a cultural ideal than an ideal of autonomy proper. Philosophical autonomy would likely have its own criticism to cast at this target – namely, that insofar as someone adopts such ideals simply because they have been socialized into accepting them, they are in fact not fully autonomous. What has been passed off as "autonomy" within the cultural mythos is not actually autonomy.

However, this second criticism, and the feminist concern with relationships and caring it captures, do pose an interesting question for autonomy. Relationships are undoubtably a valuable part of the human good life, but does an autonomous life necessarily need to have these things? If it does, then we have ventured into a substantive account of autonomy, which faces its own problems (see chapter 3). But if it doesn't, then the belief that humans should prioritize relationships is separate from questions of autonomy, since one can be autonomous with or without these relationships. As soon as we realize that autonomy is not tied to substantive independence, the point that relationships are important misfires as a criticism.

Criticism #3: The third criticism is in some ways a version of the second, but its implications are different. In extreme versions of autonomy, relationships are viewed as a *threat* because such relationships are at odds with independence. This leads to isolated individuals and a warped view of the world, since it blinds us to the ways we are dependent

on other people. In opposition to this, feminists emphasize that relationships are essential – not just for individual well-being, but for the development and support of one's very autonomy! This criticism more directly relates to autonomy since it emphasizes not just that autonomy should be open-ended to the value of relationships, but that autonomy *requires* relationships. It seems well within the realm of philosophical autonomy to accommodate this insight; however, different versions of it will have different impacts on autonomy. I'll discuss this point further in section 3.2.

Criticism #4: The fourth criticism is once again closely related to the second: by assuming that the standard individual is unattached and self-sufficient, traditional ideas of autonomy marginalize and thus devalue the experience and the identities of many (if not most) women, who have not viewed themselves in these ways (precisely because their social context would not let them view themselves this way). The idea here is that autonomy has been closely linked to notions of personhood (traditionally, "manhood") and to what gives humans their unique kind of dignity and value. But if the standard for "capable of autonomy and therefore uniquely human" is "unattached and self-sufficient", this means that people who have identities which are primarily relational and other-focused are seen as *less* than fully human. It also indicates that these people were simply not autonomous at all.

This criticism both gets something right and gets something wrong. While it is true that it is biased and inaccurate to say that only self-sufficient individuals are fully human and (potentially) autonomous, the fact is that in societies where women and other marginalized persons were given a supportive and other-focused role, this role *was* in fact

a problem and *did* make them less autonomous. It could even be argued that it denied them the opportunity to express the full extent of their humanity. How can we critique these identities while avoiding disrespecting and devaluing them as they have traditionally been? Two key but potentially contrary moves are required here: we want to value identities which were primarily relational and recognize them as potentially autonomous, but we also want to recognize that these relational identities *were* partly harmful. This latter move is the flip side of criticism three. Relationships can both support *and* be detrimental to individual autonomy. The question becomes how we can differentiate autonomy-supporting relationships from autonomy-harming ones.

For present purposes, we can conclude that none of this speaks against philosophical autonomy. Again, this kind of autonomy does not automatically preclude identities which are bound up in close relationships with others. Philosophical autonomy *does* emphasize the importance of one's relationship to oneself, and this suggests a potential route to answering this predicament: one's relationships with others can be fundamental to oneself so long as they do not interfere with having the right sort of relationship with oneself. This is line with some theorists, such as Joel Anderson and Axel Honneth[30], who have suggested that self-regarding attitudes such as self-respect and self-esteem are essential for autonomy. This is a suggestion we will take up later in chapters 3.

Criticism #5: Traditional accounts of autonomy tend to emphasize reason over emotion, desire, and embodiment, and they therefore discount the essential and enriching role such things play in a human life. While this is a valid criticism of some traditional,

[30] "Autonomy, Vulnerability, Recognition, and Justice"

masculine-centered notions of autonomy, it's harder to see how it maps onto philosophical autonomy. Philosophical autonomy does not intrinsically downplay the important role that emotions, desires, and physical existence play in our lives. It *does* emphasize the importance of reflection and developing self-knowledge, but this is not necessarily a problem. The deeper question becomes what *role* do desires, emotions, and embodiment play in autonomy? More specifically: How do they interact with cognitive attitudes like beliefs and reasons to help or hinder the development of autonomy? Does autonomy necessarily give priority to reason? If it does, is this a problem? If this *is* a problem, does this mean the conception of autonomy should be revised or rejected, or that we must accept that autonomy is inimical to certain forms of the good life? I am inclined to believe that autonomy does give a prime role to reason, but that this need not be incompatible with giving adequate weight to emotions and their ilk. If we adopt this optimistic position, the key question is the second one: how do things like emotions and desires interact with reason to contribute to autonomy?

Criticism #6: The final criticism which feminists raise is similar to post-modern worries surrounding the subject. Traditionally, autonomy has seemed to assume a transparent, rational, self-aware agent. But the more seriously we take psychoanalytic theories which pose the subconscious as playing a major role in our behavior, the more implausible this view becomes. Once again, this is not incompatible with philosophical autonomy, and in fact describes one of the key motivations for developing a conception of philosophical autonomy. We often *aren't* aware of the deeper motivations that operate within us and cause us to think, feel, and act the way we do, and this is precisely why we

need a deeper notion of autonomy. Now, the worry might be that because we are so psychologically opaque to ourselves, philosophical autonomy poses an ideal which we could never reach, and since "ought" implies "can", we should abandon the demanding ideal of philosophical autonomy. But I think this is unnecessarily fatalistic. Even if self-knowledge is hard to come by – which it certainly seems to be – this does not mean we cannot continue to make progress. Indeed, it seems we should continue to try even *if* all we can hope for is continued progress.

Of the six criticisms lodged against traditional accounts of autonomy, none undermine the plausibility or desirability of philosophical autonomy. As we have seen, some in fact support it. Some of these criticisms do make points which an account of philosophical autonomy will need to keep in mind. Of particular interest are the questions about what role relationships with others can play in individual autonomy: How do relationships support autonomy? How do they harm autonomy? How can we parse the difference? Such questions are at the heart of relational accounts of autonomy. The different ways we can answer them allow us to distinguish two kinds of relational accounts.

3.2: Causally vs. Constitutively Relational Accounts

Relational accounts can be distinguished by the role they believe relationships play in autonomy. They fall into two general types. The first type emphasizes that relationships with others – and the social circumstances one was born and raised in more generally – contribute to the *development* of autonomy. These are known as *causally* relational accounts. On such views, relationships can help autonomy by supporting the development of capacities involved in autonomy, or they can harm by stifling or discouraging the

development of these capacities. I suspect that the vast majority of accounts of autonomy are implicitly causally relational; the position that we could develop the complex abilities required for autonomy even if we were not raised in a human society seems untenable. But causally relational accounts of autonomy are distinguished because they bring these causal conditions into the spotlight and analyze what exactly is required for the development of autonomy.

The second type of relational account goes beyond this; it emphasizes that certain relationships with others are not just necessary conditions for the development of autonomy but are a *constitutive part* of autonomy. These are known as *constitutively* relational accounts. The idea of constitutively relational accounts is that while autonomy is no doubt a condition that has to do with states internal to the agent – whether you are governing your self in the proper way – there are also conditions external to the agent which are necessary for exercising autonomy at any particular time. For example, you might need to not be involved in any relationships which effectively put you in a position where you do not have de facto control over your life and actions – that is, a position of obedience and subservience. This is Marina Oshana's view. Or you might need to be in a situation where enough options are available to you such that your choices are not unduly constrained, as Susan J Brison argues.

There are, unsurprisingly, accounts which seem to blur the line between "causal" and "constitutive". For example, Joel Anderson and Axel Honneth present a sensitive account of how close intimate relationships with friends and families can shape our self-regarding attitudes of self-respect, self-trust, and self-esteem – attitudes which are

necessary for autonomy. They convincingly argue that being raised by a family and community can support the development of these self-regarding attitudes, but they also suggest that without the *ongoing support* of these attitudes, it will be much harder for the individual to maintain them, and thus to maintain her ability to be autonomous. Much rides on this "much harder": if the lack of support made it *impossible* to maintain these self-regarding attitudes, then this would clearly be a constitutively relational account.

For the purposes of this dissertation, we can put aside causally relational accounts. While looking into the necessary conditions under which autonomy is developed is important work, especially for questions of social justice, such accounts do not offer an analysis of what autonomy itself involves. The external conditions they identify and describe may be necessary for setting the individual up for her best chance at becoming autonomous, but they are only background conditions (even if necessary ones). They do not take the agent all the way to autonomy: that, she must do herself.

Feminist philosophers have illuminated some important questions any account of autonomy will need to address. Are external factors necessary for autonomy, as constitutively relational accounts suggest? What is the role of desire, emotions, and the body in autonomy? How do such things interact with reason to contribute to (or hinder) a person's being autonomous? What sorts of relationships are compatible with autonomy? Which are incompatible with it? How are we to differentiate oppressive, autonomy-undermining forms of socialization from forms which not only helpful but necessary? None of these questions speak against a need for philosophical autonomy, and perhaps an account

of philosophical autonomy will not need to answer all of them, but they are certainly questions to remember.

Conclusion

We have laid out the basic contours of the concept "philosophical autonomy". Philosophical autonomy is a demanding aspirational ideal which may be hard to achieve, and which we cannot simply assume all people have. It must not be confused with political autonomy. It is primarily based on governing the self (deciding on the substance/content of the self), and not just governing from the self (determining actions in accordance with the contentful self you mostly already have). The motivating concern of philosophical autonomy is that the desires, motives, values, and so on which are defining of us, and which lead us to make certain choices and take certain actions, may nevertheless not be rooted in us. The decision processes by which we chose particular actions, the very *self* we have – including what we value, what we want for ourselves, what we take to be true – may be importantly unfree. Only an account of governing the self can address this concern. In the chapters to come, it will be this idea of governing the self which serves as the touchstone.

Chapter 2: Surveying the Literature, Part One

Structural Accounts

Introduction

The next two chapters will provide an overview of the self-governance literature. The goal is twofold: 1) to classify the current views of self-governance in terms of my distinction of "governing from the self" and "governing the self"; and 2) to argue that none of the accounts currently on offer yields an adequate standard for governing the self. In the course of undertaking 1) and 2), we will also be able to 3) clarify what precisely an adequate account of governing the self requires. This chapter will focus on those accounts which are properly understood as governing *from* the self. Since this categorization is new, I must argue for classifying them thus. Although governing from the self cannot adequately ground philosophical autonomy, this inadequacy will illustrate what is necessary for an account of governing the self (and what is *un*necessary).

To remind the reader, the core distinction which will guide our investigation is between two different ways of interpreting the idea of "self-governance". On a "governing *from* the self" view, I already have a contentful self – certain core interests, values, desires, cares, and so on which are essential to and defining of the particular person I am – and I count as self-governing when my actions and choices are directed from this essential self. This contentful self might be semi-inconsistent; I may need to excavate it from the internalized expectations and values of others, and then refine it. Nevertheless, philosophers who have governing from the self views hold that I *have* an pre-given contentful self; I already have particular contents which are defining of who I am. To be

autonomous is to act from what is defining of me. The "law" I follow comes from *me* because it comes from what is *defining of* me. This is what gives it the authority necessary for self-governance.

But on a "governing *the* self" view, this contentful self cannot provide the authority necessary for autonomy in the strictest sense. If this contentful self is determined for me, then I did not authorize the content of this self in the first place. Although the content may define who I am, the problem is precisely that I did not define who I am. On a governing-the-self view, I decide what my substantive self will be. When autonomy is thought of this way, "giving myself my own law" means that I determine which ways of acting will be congruous with myself – I decide what will be essential to me.

I do not want to deny that governing from the self pinpoints a valuable form of autonomy. When people are fighting against oppressive forms of socialization which marginalize particular identities, I believe governing-from-the-self is the concept which they are implicitly invoking. For example, queer people fighting for the right to be themselves are asserting that something they take to be an essential aspect of themselves – and the free expression of which is a core component of their well-being – should be recognized as legitimate. They are asserting that they should have the right to lead their lives in accordance with a contentful self which they take to be essential to them. Governing from the self accounts assuredly have a role to play, and they capture something which can legitimately be called *autonomy*[31]. It's for this reason that I have called the autonomy a

[31] It also seems clear that governing from the self is importantly linked to political autonomy. My hypothesis is that political autonomy defines the external space within which all adults living in a liberal society are (ideally) guaranteed the right to choose

governing-the-self view attempts to capture *philosophical* autonomy. In what follows, I will use these two terms synonymously.

The above considerations illustrate an important point: the concern which motivates the need for an account of philosophical autonomy is rather unique. Unless we are clear about this concern, we are likely to lose track of what distinguishes a view of governing the self. Governing-*from*-the-self views aim to pinpoint the circumstances under which we can say we are in touch with what we "really", "genuinely" or "most" want. The key concern/concept here is *authenticity*; the ideal is that I direct my own life (my own self) based not on what *others* have told me I should want or value, but on what *I* most want and value. The relevant debate here surrounds 1) what qualifies as our "authentic" or "deepest" self, and 2) how we can go about ensuring that we have made contact with it. Depending on how seriously one worries about the impacts of socialization and the like, this authentic self can be closer or farther to our current contentful self, and (correspondingly) more or less difficult to get in touch with. However, the idea is that I already *have* an authentic core: a set of deeply rooted proclivities, interests, values, and so on. Because they are deep-seated, these things largely define the unique self I am. Furthermore, because they are deep-seated, I do not typically choose them – the task is for me to get in touch with them. As suggested above, authenticity is an important idea. It may seem to provide the heart of all more robust forms of autonomy.

---

their own lives and values. Governing from the self accounts attempt to define the conditions which will ensure that a person has not been internally blocked from pursuing their own authentic values and lives.

But authenticity is not the concern of governing-*the*-self views. To be sure, governing the self does involves the ideal that I will not direct my life based on the dictates of others. However, whereas authenticity involves me getting in touch with deep aspects of myself and being guided by these aspects – hence why it is governing from the self – governing the self does not accept even these seemingly deep aspects of myself as having the necessary authority. The worry here is *not* that I might not have the self I most want, find most fulfilling, or which expresses some unique inner essence of mine. The worry is that I might not have control over who I am: that in "my" actions and "my" decisions I am actually being determined by forces outside of and unknown to me. The concern which lies at the heart of governing the self is a concern for *responsibility*[32]. It is a desire to ensure that I exercise control at the deepest level of my self, where this control is *not* simply seen as the necessary means to having the self I most want[33]. In short, governing-the-self cannot be reduced to a concern with authenticity[34] or related notions of self-expression.

---

[32] As someone who is non-religious, I still see a deeper meaning in "Forgive them, Father, they know not what they do." How much of what we do is ultimately unknown to us? How much damage do we cause because we fail to understand and take control of ourselves?

[33] We could perhaps say that this means the concern at the heart of governing the self is formal instead of contentful; it is concerned with the form of activity and control instead of aiming at the ultimate content of fulfillment or authenticity.

[34] There is an ambiguity in the language here which can cause confusion. We typically say I am authentic if my deepest values (and so on) come not from others, but from myself. However, they can "come from me" in the sense that I have these values not because others told me, but because they stem from the core of my being; or in the sense that I have actively chosen them. The former is an instance of governing from the self; the latter is an instance of governing the self.

To be clear, when I say that governing the self is based in a concern for responsibility, I am not referring to *moral* responsibility. As discussed in the introduction, one can be morally responsible without being autonomous. I am using the term "responsible" similarly to how it is used as a character trait. When I am concerned about being responsible for myself, this means that I want to be aware of the influences acting on me, of the reasons why I do what I do, and ultimately be in control of myself and my actions. (This can certainly be related to a concern for moral responsibility: if I am concerned about doing the right thing, I may want to be in greater control of myself.)

There's one final way to conceive the difference between these two interpretations of self-governance: governing from the self is largely concerned with actions, while governing the self is largely concerned with the self. Governing from the self is perhaps more accurately described as *self-direction*[35]. Although self-direction clearly seems to be an important part of autonomy, philosophical autonomy is not in the first instance concerned with this executive function. Arguments which center around self-direction are therefore less relevant to our current project.

Section 1: Preliminaries

1.1: Organization and Scope

Let me begin with a few notes on how the next two chapters are organized. Self-governance is based on my having a particular authority with regard to my own decisions.

[35] Agnieszka Jaworska makes a very similar distinction when she points out there is a difference the governance which lies at the heart of autonomy – the decision of which values to hold, projects to pursue, desires to cultivate, etc. – and the execution of the results of this governance (from "Respecting the Margins of Agency").

We "self-govern*"* when our decisions have been properly authorized by us; that is, when they have the proper relation to, or grounding in, the self. When I speak of something (a desire, an action, etc.) having the proper "authority" or "being authorized", I simply mean this as shorthand for having the proper grounding in the self such that it counts as an instance of self-governance. What differentiates the various accounts of self-governance is how they explain this authorization and the self it is based on. In the literature, this authorization is not always an active, conscious decision; I will therefore sometimes speak of a desire's being "grounded in authority" or "having the proper authority". I will categorize the various accounts based on the story they tell about this authority.

We have already discussed the key categorization we are concerned with – governing from the self and governing the self – and seen how these provide different stories of authority. The goal of the next two chapters to determine of each account whether it is properly understood as concerning governing from the self or governing the self, and, if the account is attempting to be one of governing the self, whether it succeeds. In this chapter, we will start with those accounts that are most clearly of governing from the self; in the next chapter we will go into those which include some elements of governing the self, and end with those that are ostensibly aimed at providing an account of governing the self.

The literature relevant to our discussion will include some of the literature on moral responsibility. This may be surprising, since the autonomy required for moral responsibility is looser than philosophical autonomy. But as discussed in the Introduction, moral responsibility, like autonomy, often involves a notion of self-governance. (See

Appendix A for a fuller discussion.) These accounts can be illuminating for our purposes. I will therefore refer to the central concept being discussed as "self-governance", as opposed to autonomy. Per Fisher, it will be important to remember that any negative conclusion about an account's failure to capture philosophical autonomy should *not* be taken as an argument against the adequacy of the account for its original purpose.

1.2: Major Themes

As we survey the literature, we will find several themes showing up repeatedly. Briefly discussing two of these will allow for a more fine-tuned analysis.

Self-governance in general clearly involves a special kind of *activity*. However, it is possible to conceive of governing from the self as a largely automatic process: I chose my actions (or courses of actions, or larger projects) based on my core motivations (desires, interests, values), and if this self is stable and coherent enough, I will not need to reflect or question myself. In contrast, governing the self inherently requires this reflection and shaping of myself, since it requires me to act on myself. In this way, governing the self requires a higher-order activity than governing from the self. As we look at the different accounts, we will pay close attention to the forms of activity each involves.

Another element which shows up repeatedly is *reason*. Reason is taken to be connected to the uniquely human form of activity which underlies self-governance. The essential intuition behind according reason a central role in autonomy, I take it, is that *reason seems to allow for new degrees and forms of control*. First, human agents can "step back" from immediate inclinations and consider which actions they actually want to take, and which desires and goals they want to guide their decision process – a procedure which

requires reasoning. Reason also enhances our control by opening up our options: when we "step back", we allow ourselves to become aware of a wider variety of possibilites, as well as a wider variety of considerations for and against each option. Our actions are therefore not simply a result of external environments triggering instincts, but a result of considerations we take to be *justifications* for our actions. We can articulate and explain these justifications, and thus tell a story about our own actions which demonstrates a high-level of self-understanding (or self-interpretation). In other words, we have a richer *relation to ourselves*. Based on these considerations, our actions seem to be guided by us in a way they aren't for animals which lack reason. Reason plays a substantial, if not the key role in opening up the very possibility of self-governance.

I have mainly been describing the ideal case of careful and conscious deliberation. But I take it that even in less deliberate cases our reason is usually implicitly engaged. When I respond seemingly automatically to a situation which calls on me to act, often this is because it is simply clear to me what the "best" course of action is, and I don't have to expend mental energy explicitly weighing different reasons.

Another consideration for the significance of reason is that it *involves an expanded and enriched understanding of the world*. By allowing me to engage with and respond to multiple layers of meaning in a given situation, reason gives my actions a depth which makes them more robustly my own. Furthermore, our more sophisticated understanding opens whole new realms of meaning, and hence new kinds of reason for action: values are a paradigmatic case. When I am sad and my close friend tries to console me, this action means more than when my dog comforts me, because my friend can understand why I am

sad, appreciate all the aspects of my sadness, and appreciate the value of solidarity. (In this way, part of our own depth is intrinsically related to our appreciating the depth and richness of the world; alternatively, the richness of the world we move in corresponds to our own capacity for depth.)

Because reason is seen as central to distinctive human agency, many of the views we will be discussing include it as an element. We will look carefully at the role reason plays in accounts which do *not* work for governing the self, since this will elucidate what conception of reason might work.

1.3: Key Terms and Classifications

The majority of proposed accounts of self-governance are *formal*, i.e., *substance-neutral*. Such accounts hold that what matters for self-governance is that one's will have the right form, where this means that the agent has *the right kind of relation to her desires.* The correct form ensures her desires are "authorized", and thus that she is self-governing when she acts on them. Often, formal accounts will give a theory of which psychological states or mechanisms constitute the person herself, and claim these states are where the authority of self-governance resides. In these cases, "having the right form" means more precisely that *the elements of the agent's psychic economy have the correct relations to each other*. Michael Bratman calls these *psychological* accounts because they pick out one element (or set of elements) within the agent's psyche which properly plays the role of the agent. An advantage of such psychological accounts is that they avoid wat Bratman calls "the homuncular view of agency": the idea that there is a "little agent" in my head acting on the various parts of my psyche. As the above description suggests, psychological views

are well-positioned to provide an account of governing *from* the self: we simply figure out where the self is in the "psychic stew", and whatever comes from this "self" is authorized.

On a formal account the content or substance of what one wills is irrelevant. One can be autonomous while also deciding to be in a subservient position, or while willing evil things; so long as one's will has "the right form", one will count as autonomous. This is often seen as a strength (sometimes a must) for an account of self-governance. Intuitively, it seems illicit to put restrictions on what counts as self-governance "from the outside". Even when we think the life choices a person makes are objectively wrong, since self-governance is by definition *self*-governance, all that seems to matter is that the person has the proper relation to himself. Our approval or disapproval appears besides the point.

Some formal accounts can be sub-categorized as *procedural*. While formal views in general stress that the person's will has the proper internal organization, procedural views add on the requirement that the person undergo (or be capable of undergoing) a certain procedure to arrive at this form. Almost always, this procedure is reflective endorsement of some kind. On such views, having the right relation to one's desires includes undergoing this procedure.

Formal accounts are often referred to as structural accounts, "structure" being another word for form. However, I am going to reserve the term "structural" to contrast with the subcategory procedural (i.e., structural and procedural are both subcategories of formal accounts); an account is *structural* when it *only* requires that a person's will has the right sort of form, and the person does not need to actively undergo some procedure to arrive at this form. Structural accounts are therefore generally more passive than procedural

accounts. The devil is in the details here: an account may have ostensible elements of activity while still being structural. The essential point to remember is that the categorization of any account depends on where it locates the *authority* of self-governance. Since structural accounts hold that this authority is bestowed passively – I *only* need to have the correct form – they are unlikely to provide an adequate account of governing the self. However, since many of these structural accounts also contain aspects of activity, looking at them can clarify what *kinds* of activity are necessary for governing the self – and which will not work.

One final note: many of the authors working on self-governance speak in terms of "identification" or "endorsement". To remind the reader, the basic idea is that a desire has agential authority when a person "identifies with" or "endorses" this desire. These terms are therefore ways of getting at the concept of authorization, and are clearly related to self-governance. The question becomes what identification or endorsement consists in. On some accounts, "identification" denotes an active decision. On other accounts, identification becomes mostly passive: something which is bestowed more or less automatically in virtue of certain psychological states. While similar, "identification" and "endorsement" are not completely synonymous. "Identification" can be used in either the active or the passive sense, but "endorsement" indicates the active sense. Once again, active accounts seem more promising for the purposes of governing the self. Since many authors use these two terms, I will also use them while discussing their views.

In keeping with the order proposed above, we will begin with the most passive accounts and move to the more active: we will begin with structural before moving to procedural accounts.

Section 2: Coherence Accounts

In one sense, *all* structural accounts are "coherentist": they hold that an agent is self-governing if the person's will coheres with *the element(s) in her psyche which properly represents her*. This is how Sarah Buss uses the term in her SEP article on "Personal Autonomy". But I will reserve the term "coherence" for a smaller subset of structural accounts. On a coherence view, the self which forms the basis for self-governed action is coherent, meaning that it is free of internal conflicts and that its elements are mutually supporting. It is this internal lack of conflict and harmony – this *coherence* – which provides the authority necessary for self-governance. An agent is self-governing with respect to a particular action *A* when *A* fits with this coherent self. Because *A* fits into the coherency *defining* of the self, it must come from the self. Other views of autonomy may emphasize the importance of an integrated and unified self, but if they do not claim that this unified self is what *grounds* the authority of self-governance, they are not coherence *views*. For example, Harry Frankfurt's notion of coherence (which he calls "wholeheartedness") evolved over time: initially it played only a supplemental role to his idea of self-governance, and it was only later that it became the ground of self-governance.

Since the story coherence views tell about the authority of self-governance is essentially based on an already-unified self, they can only ever be governing-*from*-the-self views. But coherence views are not intrinsically incompatible with governing the self. The

problem is simply that the account they give of authority does not address the central move governing the self is concerned with: it does not explain how we can go about *shaping* and *changing* this coherent self. Coherence views simply assume that once we have this self, it will provide the necessary authority.

The above description, while accurate, is a bit of an oversimplification. Looking briefly at two coherence accounts will confirm that, even with added sophistications, coherence views are still only accounts of governing from the self. We will look at Laura Waddell Ekstrom's view of coherent character and Harry Frankfurt's view of wholeheartedness.

2.1: Ekstrom on Coherent Character

A paradigmatic coherence account is developed by Laura Waddell Ekstrom in her (helpfully titled) paper "A Coherence Account of Autonomy". Ekstrom claims we are autonomous when we act in accordance with our "most true or central" self, where this central self just *is* the *coherent* self. Ekstrom defines autonomous action not in terms of how we *acquire* the core aspects of the self, but in terms of how well our actions *accord* with the self when it is already formed: in short, her account is properly classified as "coherentist" in my sense, and is primarily about governing from the self.

However, the way Ekstrom characterizes this central self suggests fertile ground for a view of governing the self. Firstly, the beliefs and desires which are relevant to the "self" are not just any beliefs or desires: they are ones which are "endorsed by one's evaluative faculty". Such endorsement means that the agent takes these desires and

beliefs[36] to meet some standard of goodness. Ekstrom leaves open ended *which* standard of goodness, suggesting that it is up to the agent to decide on the standard. While all such goodness-meeting attitudes are a part of one's *character*, one's *most true or central* character is not all of these goodness-meeting beliefs and desires; it is the subset of these which cohere together. Ekstrom claims that the *self* is one's character *plus* the ability to refashion it, so presumably one's *most true/central self* is the relevant subset of one's character plus the ability to refashion it.

This idea of "endorsement by one's evaluative faculty" indicates that we play some active role in deciding what will be a part of the self. Furthermore, by basing this endorsement in reason, as opposed to a whim or a desire, the role we play seems to be more than just minimally active. Unfortunately, Ekstrom does not devote much time to explaining how we should understand this evaluation or its corollary standard of goodness. How would I go about making such an evaluation, or choosing the standard of goodness it is based on, in a way that counts as genuinely active? Ekstrom simply takes it for granted that I *already have* a central set of evaluative judgments, and that I can simply look to this to set the standard of self-governance. The possibility for revision which Ekstrom is careful to include in her definition of the self also seems to open up room for governing the self. But surprisingly, she says almost nothing about what this revision might consist in. Ekstrom thus leaves untilled the very aspects of her account which could ground a view of governing the self.

[36] By "beliefs", Ekstrom presumably means practical as opposed to theoretical beliefs.

The assessment of Ekstrom's view as one of governing from the self is cemented when we look more closely at the story she gives of the authority of self-governance. One's core self, which sets the standard of autonomy, is chiefly characterized by its coherence. This places a premium on consistency: it emphasizes ensuring the self is as coherent as possible, and thus downplays the importance of deciding what this coherent self should look like in the first place. This is especially clear in the argument Ekstrom gives for why we should take the coherent set of beliefs and desires to be the most central self. Her three reasons are 1) they are *long lasting*, in part because they are *mutually supporting*; 2) we will be most willing to *defend* them, since presumably they will be our *deepest convictions*; and 3) because they cohere together, we will *not be internally conflicted* when we act on them (i.e., we will be "wholehearted" in Frankfurt's sense). These three reasons emphasize the way the reasons we have fit together to form a stable and substantive character. They do not emphasize the active role we took in adopting these reasons, i.e., the role we took in forming this character. Interestingly, the aspects of Ekstrom's account which could ground a view of governing the self are characteristics of the more general, less central "self"; when it comes to the *true* self, these potential elements of governing the self are downplayed in favor of coherency.

Ekstrom's view might seem to be that authorization of a desire is actively granted. She argues that "[w]hen I act on an authorized preference, I act in a way that is autonomous *because I can give many reasons for* my act, *reasons that support each other in a coherent structure*." (emphasis added). The giving of reasons indicates that I endorse a desire, as opposed to finding that I am "satisfied" with certain desires: it indicates activity. However,

we must be careful. While the giving of reasons seems to be something I *do*, the fact that I give *these* reasons is because they are based in the evaluative judgements which constitute my central self. Once again, we are left with the question of how I acquired *these* evaluations, and how I acquired *this coherent set* of evaluations. While Ekstrom's account has elements of activity, this activity is not on the level – it does not address the central questions – necessary for an account of governing the self.

2.2: Frankfurt on Wholeheartedness

Frankfurt is famous for his hierarchical picture of the will. We "endorse" first order desires about the world with second order volitions about these first order desires, thus becoming "identified" with the first order desire. Frankfurt's earlier views held that the authority of endorsement was based in this hierarchy of desires. But in his more recent paper "The Faintest Passion", he switches to a coherence account, with the paper "Identification and Wholeheartedness" serving as a transition from his earlier to his later views.

Wholeheartedness means that I know unequivocally what I want. Frankfurt uses this term in two related ways. First, when we are wholehearted with respect to a particular desire, we are unequivocally "behind" the desire: we have no ambivalence about whether we "really want" it[37]. But this first use of the term is grounded in the second: we are fully

[37] Here's an example of Frankfurt using the term in this way (in F.P.):
"[Wholeheartedness] does not require that a person be altogether untroubled by inner opposition to his will. It just requires that, with respect to any such conflict, he himself be fully resolved. This means that he must be resolutely on the side of one of the forces struggling within him and not on the side of any other."

behind this desire precisely because it is *harmoniously integrated* into the rest of our volitional structure[38]. Our "whole heart" – our whole *self* – is behind this desire[39]. Frankfurt first introduced this concept in "Identification and Wholeheartedness". With this concept, Frankfurt introduced the *element* of coherence into his view. However, in "I&W" Frankfurt does not yet have a coherence *account*. In this paper, the primary move was still one of decision: I *decide* to either exclude or include a desire into myself. If I incorporate a desire, I need to then integrate it with the other desires I have endorsed. This process therefore hopefully *resulted* in a wholehearted self; but what *grounded the authority* of each of these desires was the decision itself.

In "The Faintest Passion", Frankfurt flips this relation. What gives authority to individual desires is just that they are harmoniously integrated into the rest of our will, i.e., that we are *wholehearted* with regards to them[40]. Instead of a decision to endorse individual desires, second order volitions take the form of passive state Frankfurt calls *satisfaction*. To be sure, these are still *second order* volitions: they are about my first order desires, and

[38] Here's an example of Frankfurt using the term in this way (in F.P.): "the health of the will is to be unified and in this sense wholehearted. A person is volitionally robust when he is wholehearted in his higher-order attitudes and inclinations, in his preferences and decisions, and in other movements of his will. This unity entails no particular level of excitement or warmth . . . What is at issue is the organization of the will, not its temperature."

[39] Frankfurt is careful to clarify that this need not mean we experience no internal psychological conflicts: for example, we might still be drug addicts struggling to overcome our addiction, and yet be wholehearted in this effort. Wholeheartedness means simply I am not conflicted with regards to what I want my will to be.

[40] Frankfurt uses the term "satisfaction", which I take to describe the relationship that this holistic state of the psyche stands to any particular desire that is thus coherent with it.

therefore about my will and not the world. The relevant change here is the lack of an active decision. Frankfurt claims that ambivalence (the opposite of wholeheartedness) "cannot be overcome voluntaristically . . . In other words, [a person] cannot make himself wholehearted just by a psychic movement that is fully under his immediate voluntary control." What this means is that I cannot choose my second order volitions; and if I cannot choose them, what gives them authority? The answer is: when I am *satisfied* with them.

Frankfurt takes pains to emphasize that satisfaction itself does not require any kind of active decision: it only requires an *absence* of desire to make a change in myself. (As such, it is not only passive but negative.) I am satisfied when I am not discontented or restless with regards to the self I have. This lack of discontent is itself grounded in the harmonious unity of the person's will. We can say that satisfaction is the relation of an individual desire to the whole psyche when this individual desire fits into the coherent whole of the person's psyche[41]. When it fits in this way, we have "endorsed" it[42].

---

[41] In fact, while we can say one is "satisfied" with a particular lower order desire, it seems that satisfaction properly speaking is about the state of psychic system as a whole. Frankfurt first introduces this concept as *self*-satisfaction. It thus includes the lower order desire in its scope, but it is not actively about this particular desire. It is simply a lack of desire to make any sort of change in the system. From the text: "Satisfaction is a state of the entire psychic system – a state constituted just by the absence of any tendency or inclination to alter its condition."

[42] This negative element in fact essential to Frankfurt's motivation for adopting this coherence view: it is his solution to the problem of the infinite regress, first pointed out by Gary Watson. If a positive element was required to endorse each lower-order desire – such as an active decision – then it seems we could always be alienated from this higher-order element, and so we would need a further higher-order element to endorse this one. By explicitly saying that no positive element is required, Frankfurt cuts off this infinite regress.

Frankfurt's coherence account is essentially a governing-from-the-self view. Insofar as I have a coherent self – *any* coherent self, of any origin – I self-govern whenever I act in accordance with this self. Authority is passively bestowed by the holistic state of my psyche. As is typical of coherence accounts, the central questions of how I can go about *shaping* this self, and whether the process by which I acquired *this particular* coherent self was legitimate, are left aside[43]. In short, questions of governing the self are ignored. In fact, Frankfurt is adamant that I cannot shape or control this self, not even to make myself more coherent – something which is more minimal than true governance of the self – since making myself *more* coherent would require a basis for this coherence, i.e., would still be governing from the self. Frankfurt argues for this by analogy: "The concept of reality is fundamentally the concept of something which is independent of our wishes . . . Now this must hold as well for the reality of the will itself. A person's will is real only if its character is not absolutely up to him."

[43] Actually, this is slightly inaccurate, for Frankfurt does address how I come by a coherent self: namely, there is nothing I can do to impact this process. He is adamant that I cannot shape or control this self, and thus precludes the very possibility of governing the self. We cannot even take action to make ourselves more coherent, something which is more minimal that true governance of the self . In other words, the passivity of satisfaction cuts both ways. There is nothing we need to do in order to be satisfied; but if we aren't satisfied, there is nothing we can do. This means it cannot work as an account of philosophical autonomy.

Here's an extended passage demonstrating these points: "We do not control, by our voluntary command, the spirits within our own vasty deeps. We cannot have, simply for the asking, whatever will we want. . . We can be only what nature and life make us, and that is not so readily up to us. This may appear to conflict with the notion that our wills are ultimately free. . . The dilemma can be avoided if we construe the freedom of someone's will as requiring, not that he originate or control what he wills, but that he be wholehearted [i.e., satisfied] in it."

There is one potentially active element in this account: Frankfurt claims that satisfaction must be based in *reflective awareness* of our current state, in the person's "understanding and evaluation of how things are with him". How exactly this works without taking an active stance of reflection and evaluation is unclear, for Frankfurt says little else about this element. It might be possible to have a general awareness of my holistic psychological state which is only implicitly operating in the background; perhaps this is what Frankfurt is requiring. In any case, since there is no active role of reflection to take – no action of endorsement – such reason in such a minimal capacity plays no role in shaping the coherent self[44].

2.3: Reflections on Coherence Accounts

Both coherence views we have discussed have hierarchical elements. However, these views are not properly speaking hierarchical because of where they place the authority which grounds self-governance. Hierarchical views hold that it is the higher-order nature of some specified mental state, attitude, or act which grounds authority. Coherence views hold that it is *the contentful self considered as a whole* which grounds authority[45]. It is the "contentful" which is decisive for our purposes. I can only have a coherent *self* if I

[44] Once again, Frankfurt does away with the need for active higher-order endorsements by claiming that the ordered unity of the person's whole will can fill in this role, and by doing so avoids the infinite regress.

[45] Since this is perhaps less clear in Frankfurt's case, here's one more extended quote: "Satisfaction is a state of the entire psychic system – a state constituted just by the absence of any tendency or inclination to alter its condition. . . A person is actually satisfied only when the equilibrium is not contrived or imposed but is integral to his psychic condition – that is, when that condition is settled and unreserved apart from any effort by him to make it so." We self-govern if we are satisfied with the will we have, and this is based on "the entire psychic system".

have a coherent set of *content*. Since coherence views hold that it is this coherent set of content which gives the authority, it is my contentful self which gives the authority. In short, coherence views are paradigmatic cases of governing from the self.

Coherence views are not intrinsically incompatible with governing the self. However, as soon as we prioritize governing the self, the value of coherence is no longer given a central place. When I hope to be able to actively shape myself, looking to the contentful self I already have is a non-starter, since it is precisely this content which I want to bring into question. Governing the self, centered on this active shaping, requires an explanation for what counts as an authorized modification of the self. Coherence views are limited to saying that what is authorized is what fits with my most coherent self. When it comes to cases of transformation, coherence views can only be silent.

I take it the intuition behind coherence views is the conviction that if I am internally ambiguous or divided, I will not truly have a functional self at all. Frankfurt, for instance, contrasts wholeheartedness with ambivalence, which is when a person is indecisive in his second order-volitions: that is, he does not know whether to be for or against a particular first order desire. Whereas wholeheartedness constitutes health, ambivalence constitutes sickness, for it leads to "self-betrayal and self-defeat". This makes much sense: how can one be an effective agent, let alone fulfill the stricter standard of autonomy, if one is internally divided and conflicted?

However, I think there is some reason to be suspicious of over-valuing a coherent self. In the case of governing the self, a self which is conflicted may actually serve a useful purpose. Take again the well-used example of the subservient housewife. She was taught

to believe in conservative values: for instance, that men and women are fundamentally different in terms of temperament and abilities, which means that they are naturally suited to find fulfillment in different roles. Let's assume that her life has gone smoothly according to plan and she is in fact happy with her situation. This woman might be thoroughly unified in terms of her beliefs and values; in fact, her personal happiness supports her belief system. The fact of that she is coherent in both the sense of being internally unified and Frankfurt's sense of being satisfied with her current state is not *by itself* enough to make her philosophically autonomous. Remember, the issue for philosophical autonomy is having thoroughgoing control over my substantial self. If this woman has simply accepted the self she was taught to have, then her very identity is the result of forces beyond her. She is not governing herself. In this way, her philosophical autonomy might be promoted if her self was less integrated. If there was a conflict between some of her beliefs, values, and desires, this could make her uneasy enough to start questioning what she was taught. It may be that a coherent, consistent, and stable self is a worthy goal of governing the self. Nonetheless, coherence itself cannot be what provides the authority of governing the self.

Section 3: Bratman on Lockean Cohesion

Over a series of papers, Bratman has developed sophisticated accounts of both self-governance in general and autonomy specifically. Bratman's views have strong elements of hierarchy and coherence, and both give central place to reasoning; however, the account he gives of the authority of self-governance is unique. He argues that the psychological states or mechanisms which shape us into a unified agent are what ground the authority of self-governance. Since these psychological states *make* us an agent in the first place –

which they do by "gluing" all the other psychological states together in a unified, organized structure – they must have the authority to speak *for* us[46].

Bratman's view is similar to coherence ones in that he emphasizes having a unified self, but this is a rather different kind of unity and it gives a different story about the authority of self-governance. While coherence views say that it is the unified, integrated *state of the psyche* and its various mental contents which grounds authority, Bratman argues it is those *particular* states which are *doing the work* of unification which ground authority: more specifically, those particular states which make possible unified temporally extended agency. On coherence views, the self just *is* the unified self; on Bratman's view, the self is those particular psychological states which *make* us unified. We will begin by looking at Bratman's accounts of self-governance, and then turn to his more specific views on autonomy.

The foundation of Bratman's view in his 2000 paper "Reflection, Planning, and Temporally Extended Agency"[47] is a conception of human agency: we are fundamentally *temporally extended*. Bratman proposes we make sense of this in broadly Lockean terms –

[46] I find Jaworska's way of putting it helpful: she says "the very attitudes which help [] fashion the agent out of what would otherwise be a mere collection of mental events occurring at one time or evolving over time [] have a plausible claim to speak for the agent."

[47] Although Bratman does not speak in terms of self-governance in this paper – he reserves this term for later discussions – I take it he is nevertheless talking about a form of self-governance for two reasons. 1) He is talking in general about things which constitute where the agent stands and thus have the authority to represent the agent; 2) He is particularly interested in making sense of the idea of endorsement/identification (which he calls "strong reflectivity").

that is, in terms of *psychological ties* between our past, present, and future selves (hence why I call it a "Lockean Cohesion" view). These psychological ties are what allow us to exist as temporally extended beings; therefore, whatever *supports* these psychological ties brings me into existence. Since it constitutes me, it has the authority to speak for me.

Bratman argues that there are two key kinds of psychological ties which glue together my agency over time. Continuities are simply the continued presence of desires, interests, and so on. Connections are when two psychological states at different times cross-reference each other[48]. Connections are a more robust kind of psychological tie than simple continuities because of their mutual cross reference. These two psychological ties – in particular, complex networks of both continuities and connections – form a stable identity for the agent over time.

Bratman claims many of these psychological ties do not simply happen to the agent; "she" has active role in creating and supporting them over time (the scare quotes will be clear momentarily). "She" thus plays a key role in forming her own identity. However, Bratman does not want to say the agent is some peculiar metaphysical entity which stands separate from and above these various psychological states. What we need is to find the special psychological state(s) or mechanism(s) which can fill the role of the agent: in particular, we want to find those which can constitute "her" *endorsement* of particular desires, actions and plans.

[48] Bratman's key example is forming an intention. I form an intention to go to the plant store at T1 which inherently refers to my later action of actually going to the plant store at T2. Correspondingly, my action of going to the plant store at T2 inherently refers back to my intention to go which I formed at T1.

Bratman initially proposes two elements which can support the functioning of such psychological ties: plans and policies. A *plan* is a specific course of action for a particular situation unfolding over a particular time. A *policy* is a more general plan for how to act in a recurring situation. As such, both support psychological connections and continuities[49]. While these plans and policies can create the temporal ties which bind me into an agent, Bratman wants to reserve the special role of *endorsement* for only a subset of these. Only those *policies* which have as their *content* a desire of mine can constitute my endorsement of this desire. Such policies are thus higher-order in Frankfurt's sense. Presumably it is because they are higher-order that they are fit to play the role of the agent in constituting and supporting her own identity. Bratman calls these *self-governing policies* (SGP).

A self-governing policy has three key aspects: it is a policy 1) to *treat* a particular desire as setting an end which can justify action; 2) to support this desire's *functioning as a motive* which actually drives us to action; and 3) to support its functioning *because* we see the desire as justifying action[50]. (In other words, 1 and 2 are essentially connected.) To

[49] Furthermore, we typically have a whole network of plans and policies which mutually support one another, thus playing an enormous role in supporting a coherent identity over time. As Bratman emphasizes, this cohesion is part of their purpose; I adopt plans and policies so that I am better able to take coordinated action, and so that I do not act in self-undermining and self-defeating ways.

[50] The idea behind (1) is that (2) is not enough; I could endorse "a desire's functioning as an effective motive" simply in order to alleviate the irritation of having the desire, or simply because I need to act in some way, and there's I might as well act on this desire. As Bratman points out, the intuitive idea of "endorsing" a desire is that such endorsement places the desire in a context of practical reasoning. I wish he had said more, but I take it he means that I see the desire, and the goal it points me towards, as good in some way, and thus as an adequate grounds for reasons – reasons I can use to justify my actions to others, and to myself. But (1) by itself is not enough; I can't really claim to have endorsed a desire if I theoretically believe it points me towards a justifying goal, but do

clarify, by "providing a justifying end" Bratman does not mean that we *always act* on the desire when it comes into play. Rather, since we see it as setting a justifying end, and thus as giving us a *reason* to act, we take it into account when we are deliberating. Furthermore, since different ends will have different strengths of justification, a self-governing policy will also help us determine what weight to give to each reason in our deliberative decision making[51].

We must be clear: the agent does *not* endorse self-governing policies. Rather, self-governing policies *constitute* the agent's endorsement. While this nicely allows Bratman to avoid the homuncular view of agency, it leads to the question of how these self-governing policies themselves have the authority to endorse motivations. This is where Bratman's above points about temporal agency come in. Those things which help bring me into existence as a temporal agent must have the authority to represent me. Self-governing policies help to form my identity across time, and so they have this authority.

This is an essential point for two reasons. Firstly, in other accounts endorsement (or identification) is taken to be important because it can supposedly *explain* the authority which some motivations have. These endorsed motivations then set the standard for self-governance. But on Bratman's view, authority does not come from endorsement/self-

---

not actually incorporate this desire into my motivational schema. Finally, (1) and (2) must be related; the self-governing policy must support the desire's functioning as a motive because I see it as providing a justifying end.

[51] There is one final element of Bratman's view: in order for self-governing policies to constitute the agent's endorsement of a desire, the agent needs to be satisfied with them, where "satisfaction" means that the SGP is not undermined by other self-governing policies. This is, of course, satisfaction is the Frankfurtian sense: a lack of conflict, in this case specifically a lack of conflict within one's set of self-governing policies.

governing policies; the authority SGP have is not in any way unique to them. It is something they share with plans and policies more generally, since all three help constitute the agent's identity over time. Since Bratman continues to talk about endorsement, he must see it as relevant for a different reason. For Bratman, endorsement is relevant because it (and the SGPs which constitute it) is uniquely focused on the self. Plans and policies are about how I respond to (and act in) the world, but SGP are about how I respond to the various factors within myself[52]. SGP's are how I *act on myself*; they are thus a form of *self*-governance in a way plans and policies are not. Furthermore, *because* they are a form of self-governance, they are a particularly potent way to glue the self together. This "gluing" is part of their very purpose. Nevertheless, we must remember that the authority of self-governing policies is not unique to them. This authority is applied in a special *way* in self-governing policies because they are directed at the self, but this self-directedness is not what *gives* them authority.

But the fact that SGP constitute my authority (my endorsement) means that Bratman's view fails to capture what I take to be the main appeal of endorsement. Endorsement is a powerful notion because of the special activity that it attempts to capture: the activity of acting on myself. Now, Bratman is careful to include the aspect of acting on the self – this is precisely what SGP are supposed to do. But the process behind endorsement is supposed to allow me to distance myself from any particular aspect of myself, and ask "do I really want this to be a part of me? Do I really want this to guide my

[52] Agnieszka Jaworska asks if the same is true of Frankfurt's second order volitions. Yes: SGP are second order and reflexive in the same way second order volitions are. I take it Bratman was largely inspired by Frankfurt.

actions, my understanding, my self?" In a word, endorsement is supposed to be connected to self-constitution.

Bratman's view seems to do away with this aspect of critiquing and acting on the self. According to Bratman, my SGPs are supposed to *tell* me the answer to what I "really" want in each particular instance: for every case I am unsure of, I can step back and question what I "really" want, and my SGP will be at the ready to tell me what I really want and where I stand. SGP constitute, in other words, my "strong reflective endorsement". Again, it is essential to remember that for Bratman, I do not endorse these SGPs; rather, SGPs constitute my endorsement. This negates the need for an additional element of involvement from me and thus avoids the homuncular view. But this means the higher-level question of which central motivations I want to guide me across circumstances – that is, *which self-governing policies* I want to have – is left untouched. The key activity which is supposed to allow me to impact what is defining of myself is unaddressed.

In short, despite its complexities Bratman's account is essentially one of governing from the self. He explains that what forms a unified self out of the "psychic stew" of mental ephemera has the authority to speak for that self. SGP are a special instance of unifying mental components. They tell us what weight to give to which desires in practical reasoning across various situations, and so they unify our agency across time. We are therefore self-governing when we act in accordance with the actions these various SGP suggest to us. In short, we self-govern when we act in accordance with what is defining of our unified, temporal agency. But where do these SGP come from? What grounds and legitimizes the fact that my SGP support the functioning of these motivations, and not others? How can I

reflect on the fact that these motivations are central to me – and maybe even go about changing them? These are the critical questions which need to be addressed in order to exercise any kind of more substantial control over the self that I happen to have.

Bratman only addresses these issues very briefly at the end of his paper, and his remarks are telling. He allows that it is a "perfectly coherent thought" that the agent might want to make changes in her current polices, but claims that such a change "might involve criticizable instability". Indeed, on his view it seems it would *necessarily* involve "criticizable instability"; if stability across time is what grounds authority – indeed one's very agency – then any threat to this threatens the entire enterprise. This means that Bratman's view not only *cannot* explain governing the self; it seems to exclude the very *possibility* of governing the self.

So far we have talked only about Bratman's view of self-governance. In a set of three interconnected papers, Bratman argues that *reflexive* self-governing policies constitute a form of autonomy[53].

In "Autonomy and Hierarchy", Bratman argues that we should understand autonomous action as involving both agential direction and agential governance. Agential direction involves "sufficient unity and organization of the motives of action for their functioning to constitute direction by the agent". For Bratman, the agent just is the

[53] Bratman is humble in his aims: in each of these papers, he is careful to state that he does not intend to give the singular and defining account of autonomy. Multiple psychological attitudes and functions will likely be able to play the founding role for autonomy; he simply wants to argue that (reflexive) SGP are one such psychological attitude.

coherent, temporally extended being which results from the connections and continuities of psychic elements. An action which comes from this unity thus comes from the agent[54]. Agential governance is "a particular form of such agential direction: [namely,] agential direction that appropriately involves the agent's treatment of certain considerations as *justifying reasons* for action" (Emphasis added). When we act for reasons which we take to be justifying, we are acting in accordance with a (at least subjectively) normative standard. It is this the idea of a standard against which things can be justified which grounds the uniquely potent notion of agential governance.

Bratman argues that self-governing policies are a prime candidate for filling the role of autonomous action. SGP fill the role of agential direction since they are part of what forms the agent into an agent in the first place. They also fill the more specific role of agential governance because they set policies for what will count as a justifying reason, thus setting a normative standard for the agent. Based on this, self-governing policies as such are enough to constitute autonomous agency[55]. But if we add the requirement of *reflexivity* to SGP, Bratman claims SGP will certainly be able to play this role. Bratman is sensitive to the worry that in order to count as self-governing, it seems that the agent herself

[54] This makes more sense when we remember that what grounds authority for Bratman in general is the cohesion which make an agent temporally extended. SGPs are only one instance of such a "gluing" attitude.

[55] In "Autonomy and Hierarchy", Bratman argues that since self-governing policies are policies about deliberation, they are higher-order and they involve rational guidance in a way that constitutes self-governance. In other words, simply in virtue of what self-governing policies are – policies about how to guide my rational deliberation – they can play the role of self-governance.

needs to play a key role in the functioning of these attitudes. Since SGP play the role of the agent, and what we want in the particular case of autonomous agency is endorsement of SGP by the agent, a clear answer suggests itself: SGP can endorse themselves. They can do this by including as part of their policy *taking this policy itself* to be a justifying reason in deliberation[56]. More simply: the fact that I *have* this policy is *itself a reason* to follow the policy.

Again, we must keep in mind that the agential authority which reflexive self-governing policies have is *not* grounded in their reflexivity, but in their supporting Lockean cohesion over time. Their reflexivity simply shapes or directs this authority such that it shows up not simply as agential direction, but as agential governance.

We have already discussed why SGP will not work for governing the self. The addition of reflexivity does not endow them with this capability. We can still ask on what basis the agent comes to have *these* particular SGPs and whether this basis has the proper authority. On Bratman's view, SGPs have authority in virtue of what they *do* – unifying the agent – and not in virtue of what they are *based* on. My above comments about why coherence accounts cannot work for explaining the authority of governing the self apply equally to Bratman. While Bratman conceives of the relationship between authority and unity differently – it is not the unity itself, but what *causes* the unity which has authority for him – Bratman still gives primacy to unity in a way that precludes questioning the

[56] Bratman explains it this way: "The self-governing policies that are central to the model of autonomy that we are constructing will be in part about their own functioning. Such a policy will favor treating certain desires as reason-providing as a matter of this very policy" (emphasis added).

content of this unity. Indeed, my above comments on Bratman's view of self-governance still apply here: because he defines the agent's own activity in terms of the functioning of these SGP, there is no room left for the agent to question these SGP. Since authority for Bratman comes from what makes us unified agents across time, there is no space in which to decide on *which* policies we want to be the ones unifying us.

One might wonder if perhaps the person could use one SGP to question other SGP (Neurath's boat-style). But in order for this to work, we would need an account of how the person "switches" from the perspective of one SGP to the other. It would not do to say that one could only question a SGP from the perspective of a "higher order" SGP such that the "agent" is not really the one doing to moving (thus escaping the homuncular view of agency Bratman is so concerned to avoid). On this picture there is still a taken-for granted content in the form of the higher order SGP, and it would still be a view of governing from the self. While it makes sense to think that we must take *some* content for granted as we question other parts of ourselves, we still must be able to question *any* content. In other words, I need to be able to move in between the content of my self. This means 1) that my activity cannot be reduced to any one content (to any one SGP), and 2) we need an account of how I can "switch perspectives" between contents. In other words, despite Bratman's heroic efforts to avoid the homuncular view of agency, we still need there to be an agent which can operate on any of its contents, and thus an agent which is not essentially defined in terms of any one of them.

3.1: Reflections on Lockean Cohesion

There are two lessons to take away from our discussion of Bratman. The first is that any attempt to define not just the self but *agency* in contentful terms which "lock" the agent into a specific set of internal motivations cannot work if we want an account of agency to allow for the possibility of governing the self[57]. To clarify: I am making a distinction between the *self*, which we can think of as my current practical identity, and *agency*, which we can think of as the more abstract ability to act, including on myself. Other accounts have proposed that we understand self-governance – a particular *form* of agency – in terms of a self with substantial characteristics, such that we self-govern when our actions are grounded in those particular characteristics defining of or essential to our self. But Bratman defines *agency itself* in terms of a stable set of characteristics across time. However, as soon as we make agency itself reducible to specific contents, this means that I will never be able to act against these contents, on pain of not being able to be an *agent* at all (and therefore not being able to *act* at all). In sum, we must allow agency to be at least *separable* from the particular self, lest we foreclose the possibility of acting on and making changes in ourselves.

[57] Velleman makes a very similar point in "What Happens When Someone Acts". The "standard story of action" is that a combination of desires and beliefs give rise to reasons, reasons cause an intention, and the intention causes an action. Velleman points out that there is no room for the agent in this story. Proponents of the standard story might respond that the causal process just *is* the agent – a stance like Bratman's, since it avoids the agent being something above and beyond his mental contents. Velleman, like me, is dissatisfied with this. The role of the agent is to intervene between reasons and intentions, and between intentions and actions.

Secondly, Bratman obviously sees the unique form of activity inherent in using reason, which is why he gives it a central place in his view. SGP tell us what we should count as a reason for action. When we act based on what we take to be reasons – and especially when we deliberate, thus acting *on ourselves* on the basis of reasons – we engage in a characteristically human form of activity. Despite giving reason this prevalent role, Bratman's account still fails to accommodate governing the self. This elucidates a common theme: not every activity which involves reason will be sufficient for governing the self. We need to be able to shape what we count as a reason in the first place and influence this distinctively human capacity for deliberating and acting on the basis of reasons. As Bratman (and many others) illustrate, we can participate in these human forms of activity without governing the self: governing the self requires that we take responsibility for how we exercise these human forms, and the precise shape they take.

Section 4: Caring Accounts

Caring views argue that the agent is constituted by her set of cares, and so she self-governs when she acts from her cares. Since the authority of self-governance is based on what *substantively* constitutes the person, caring accounts are paradigmatic governing-from-the-self views. However, since this description is (as always) a bit of a simplification, we will look briefly at several caring accounts.

4.1: Seidman on Seeing as a Reason

Jeffery Seidman's conception of caring is based on Agnieszka Jaworska's explication of the concept[58]. Jaworska argues that cares are inevitably internal to the agent,

[58] "Valuing and Caring"

but denies that they could ground the authority of self-governance[59]. Seidman argues that with the addition of a rational element, cares *can* play such a role.

According to Jaworska, caring is a *complex network of secondary emotions*[60]. An "emotion" is not simply the way one feels at a particular time, but the underlying disposition to feel a certain way in particular circumstances (as Jaworska more elegantly puts it, an "ongoing psychic orientation"). "Primary" emotions are more or less instinctual responses to certain situations, which are triggered by immediate sense data. In contrast, "secondary" emotions are based in our *understanding* of a situation. They therefore involve higher-level cognitive processing (even if they do not require rational reflection). Furthermore, secondary emotions can ground non-emotional mental states such as intentions and plans. In fact, secondary emotions seem *necessary* to such things, since without them to direct my attention and interest to certain possibilities for action, I might be at a loss for what to do with myself. Their ability to play such a role is based on the higher-level cognitive processing they involve[61].

---

[59] "Caring and Internality"

[60] Jaworska was inspired by Bratman's idea that the cross-temporal psychological states which help to form a unified agent over time intuitively must be internal to the agent. She argues that caring is such an attitude which supports these connections and continuities, thereby unifying various psychological states into a single agent. Jaworska was not intending to provide an account of self-governance, but rather an account of the internality central to selfhood and full moral status. In a footnote, she remarks that it is such a self which could become the subject of self-governance.

[61] I think that Jaworska might be considered a "Lockean cohesionist" if it was not for the fact that she denies caring has the authority necessary for self-governance. However, it might also be that Jaworska is working with a different conception of self-governance than Bratman.

Caring is a *network* of multiple secondary emotions. If I care about someone, I am disposed to feel a whole range of emotions in various circumstances: joy, worry, hope, frustration, sadness, grief. But caring isn't just a collection of various secondary emotions: these emotions are unified into a coherent, expansive psychological structure because they are all centered around the same object of care. If we genuinely care about the object, some of these emotions will necessarily imply the others. If I care about my friend, I will feel happy for him when he is doing well and worried about him when he isn't; in fact, I am *committed* to feeling both, or else it seems I do not *actually* care. Secondary emotions are based on our understanding of the situation; analogously, caring requires the ability to understand how the object of one's care is faring. Importantly, this involves tracking how the object is faring independently of how the object impacts oneself: rather, the focus is on the object *itself*. This concern with the object "for its own sake" means that the agent "imbue[s] the object with *importance*"[62].

Seidman argues that Jaworska's understanding of caring is correct but incomplete. However, she gives us the key to the missing piece with her idea of *importance*. When we see something as important, we see it as a source of reasons for us. This idea of "seeing" is crucial. Seeing is *not* simply believing. To believe something is important means simply that I judge it is important. If something is genuinely important to me, I will *see* it as important – it will *shape* my perceptions, interpretations, and desires. In particular, it will

---

[62] Jaworska does not expand much on what "importance" means, but I take it the core idea is precisely that we think the object is "worth" something in itself, independent of us. This is why we track the object's own well-being.

shape the *reasons* I take myself to have. I might become aware of my cares such that I come to *believe* that something is important to me, but it is the *seeing as* important which is primary. Seeing something as a reason is essentially connected to seeing something as a reason *for me*; it exerts a certain pull on me. To see something as providing a *reason* means that I am oriented to it: it is embedded in my subjective experience of the world such that I am compelled to give weight to the considerations it directs my attention towards. Caring therefore does not just have an emotional aspect which can subsequently ground cognitive aspects (in the form of guiding plans and intentions); it has a cognitive aspect *in itself*.

Seidman argues that carings can provide grounds for self-governance because they are 1) invariably internal to the agent (i.e., she cannot be alienated from them) and 2) because they "constitute at least a part of the 'standpoint' from which she acts when she acts for reasons"[63]. This second requirement is meant to capture the importance of practical reason when it comes to self-governance. Since caring involves seeing something as a reason, this means we are very likely to actually treat this as a reason in our deliberations, and so carings constitute part of our standpoint as practical reasoners. Seidman does not argue at length for why practical reason is important, but I take it that his intuition is familiar. When I act, I am not simply moved by my desires; my actions reflect my

[63] Seidman actually argues that caring meets six desiderata for an account of self-governance. (Even more precisely, he calls these desiderata that an account of "valuing" must meet in order for valuing to serve the role of self-governance. Since Seidman ultimately argues that valuing just is caring, and since what he cares about is its ability to provide the authority of self-governance, I am interpreting these as desiderata for an account of self-governance.) I have only discussed the first two desiderata, since I take it these are the desiderata which provide the authority of self-governance. The other four desiderata are about ensuring that an account of self-governance has the proper scope, and so I have skipped discussing them for the purposes of brevity.

understanding of the reasons I take myself to have for acting. This inclusion of my understanding and rational capacities means the action comes from my distinctive agency. This is why my "standpoint" seems to be the ground of the authority for self-governance.

Despite this active element of Seidman's view – which once again is based on the centering of reason – Seidman's view is essentially one of governing from the self. I have particular, substantive cares, and when I act from reasons based on these cares, I act from myself. In other words, I self-govern when I determine my actions based on my substantive self. Like Bratman's view, this demonstrates that while acting on the basis of reasons is connected to human activity, what we need for governing the self is to have jurisdiction over what counts as a reason for us – in Seidman's terms, over our deliberative standpoint.

However, Seidman provides a perspicacious description of what is involved in this deliberative standpoint, and in human agency more generally: we are not just rationally deliberating *agents*, but emotionally laden *subjects*. Human agency involves not just decision making, reflection, and action, but the whole self – affective, emotive, embodied, etc. (This is something we see in many feminist critiques of autonomy.) As such, the proper domain for governing the self is this self, in all its complexity. Furthermore, Seidman's description of *seeing* as a reason connects these affective aspects of the self to our rational capacities. Recall how in section 1.2 above I described these rational capacities as grounding the richer world of meaning and significance which humans live in. "Seeing as a reason", and deliberative standpoints more generally, capture these aspects of reason and their corresponding forms of activity. (I take it this is what underlies Seidman's intuition that seeing as a reason can tell us when an action comes from the person himself.) We can

now re-phrase the above critique: governing the self will require us to actively shape this holistic "deliberative standpoint", and even more broadly, *how* we understand and imbue the world with meaning.

4.2: Shoemaker on Necessary Caring

In "Caring, Identification, and Agency" David Shoemaker characterizes cares quite similarly to Seidman. He understands caring to have emotive, desiderative, and often (but not always) evaluative elements. For Shoemaker, emotions are the primary element: to care about something just is to experience a range of emotions which track the object of one's care[64]. This does not seem to be substantively different from Jaworska's account of emotions. From this emotional susceptibility comes a set of desires to act in certain ways, and we may also be moved to judge that our object of care is valuable. These desires and judgments then shape the realm of our choices and reasons for action. Where Shoemaker and Seidman most diverge is in the role caring plays in self-governance[65].

Shoemaker argues that whenever we act *qua persons*, we are ultimately motivated to act on the basis of our cares (again, if we have taken the time to reflect). Of course, not *all* my actions are grounded in cares. But Shoemaker holds that in instances where my

[64] This is essential to his argument: he holds that we cannot help but care about what we do, and that this is precisely because we cannot change our emotions at will.

[65] Some brief clarifications about Shoemaker's terms and goals: he is explicitly concerned with autonomy ("Any robust theory of free agency must account for its two central features, namely, the availability of alternative possibilities and self-determination (i.e., autonomy). I here wish to focus on this latter feature"). Technically, he speaks most often of self-*determination* rather than self-governance. However, there seems to be no difference between the way he uses self-determination and the way I have been using the general term self-governance: self-determination involves being motivated by desires I endorse/identify with, and which therefore has authority for me.

action is not connected to any care, I act as a Frankfurtian wanton: I do not care what my will is, because the situation I am acting in and the choices I have before me do not involve anything I care about. The conclusion is that *all* "free agency [i.e., *characteristically human action*] is grounded ultimately in care."[66] Furthermore, Shoemaker intends this claim to have a surprisingly wide scope, since he holds that even weak-willed actions are based on cares[67]. It is clear that Shoemaker is using quite a weak sense of "self-governed" since it includes weak-willed action and is meant to include all characteristically human action. (One is left to wonder how useful such an enervated sense of self-governance can be.)

Shoemaker's justification of the authority of caring is a familiar one: who we are is constituted by our cares, and so they must have the authority to speak for us. In particular, he emphasizes the importance of emotions in our being coherent agents: without emotions, "one's decision-making landscape [would be] flat and without salience. With no emotional investment in what one might do, all options are on an equal footing—anything is possible . . . Losing one's capacity to care means losing one's identity as a coherent agent." Since who I am is "both *made possible by*, and *is a function of*, my emotional commitments [my

[66] This point is actually the beginning of a longer argument, whose main goal/conclusion is to prove that the authority at the heart of self-governance is primarily passively bestowed. Shoemaker goes on to argue that I cannot help but care the way I do, thus meaning that the authority of free agency is passive.

[67] It's simply that at the time of my action, I care about something more than what I judge I should care about. When my friend gives into smoking a cigarette even though he has been trying to quit, it's because in the moment he cares more about the relief smoking will offer him than his larger goal of quitting. Based on this, a final qualification is in order: what I am motivated to do is based on the strength of my cares at the time of action.

cares]" , these cares represent me and so provide the authority which grounds self-determination (emphasis added).

Shoemaker's view is clearly an account of governing from the self. Not only is self-governance based on the substantial self we currently have (more specifically, our set of cares), Shoemaker is insistent that we cannot change the cares we have. We may be able to take steps to change our cares – to dissolve our emotive, desiderative, and evaluative patterns which give weight to the object of care – but this is a slow process which we cannot effect in an instant.

There are two active elements in Shoemaker's account. First, he says that free agency "consists in both the necessitation stemming from care and one's *reflective awareness* of such necessitation" (emphasis added). However, the activity involved here is fairly minimal. It's essential to understand that Shoemaker envisions our "reflective awareness" of our volitional necessities to serve only an enabling role: that is, they *do not add to the authority* which these necessities bestow. What (typically) happens in critical reflection and decision-making is not nearly as active as we might think. I don't actually decide anything: I just "look inside" to discover what it is I already care about it. Once I have looked carefully at my self (or my situation) I simply know what it is I care about. My mind is "made up" for me. On Shoemaker's view, critical reflection does not actively do anything: it just clarifies my cares. The fact the reflective awareness is relegated to simply helping us discover what we care about indicates that there is very little room on his view for us to make changes to ourselves.

But – and this is the second aspect of activity – Shoemaker *does* allow that I may be able to change some of my cares. Significantly, this can only be done on the basis of some deeper care. Since my cares are my self on this view, this indicates that there is some room for me to decide what will be defining of me on Shoemaker's view – in other words, there *seems* to be some room for governing the self. But can Shoemaker's view actually *explain* governing the self in a satisfactory way? Unfortunately, it cannot. Since changes are still limited to my current substantive self, all possible changes I could make are based in the set of cares *already constitutive* of this current self. In fact, all I would be doing with such changes is getting clearer on who I really am. This is a move of discovering the substantial self I already am such that I can more effectively govern from it; it is not a move of deciding what this substantial self will be. Shoemaker is clear on this point: "Our freedom – our ability to do what we genuinely want to do . . . expands in proportion to the expansion of our self-knowledge." Who I am is already set; I just need to come to know it better.

This demonstrates that defining revision in terms of a "deeper" self will not work for philosophical autonomy. Put another way, if our current self restricts our capacity for revising the self *too* much, governing the self will never get fully off the ground. But this in turn means that we must be able to expand beyond the limits of our current self. We must have the ability to step partially outside our current substantial selves, and to develop cares and beliefs which are not already firmly rooted in our selves. How we could possibly do this is an essential question.

4.3: Frankfurt on Volitional Necessities

We have already looked at several of Frankfurt's views about self-governance. In his paper "Autonomy, Necessity, and Love" he turns explicitly to discussing autonomy. Frankfurt argues that a key form of autonomy is grounded in love (which I interpret as sub-species of caring). On his view, love is not essentially about feelings or beliefs, but about willing: the "heart of love" is "the more or less stable motivational structures that shape [an agent's] preferences and that guide and limit his conduct." This is similar to Seidman's view, since to see as a reason is to be motivationally drawn to certain courses of action. There are also similarities to Shoemaker, who emphasized that we most want to do what we most care about. Like these other caring accounts, Frankfurt believes I am defined by what I love. However, Frankfurt's view is unique in where it places the ultimate authority of self-governance.

Like other caring accounts, Frankfurt argues that because what I love is *defining of* and *essential to* me, it sets the standard of self-governance. Since love is volitional and defining of me, love determines what I must will, on pain of not being myself. As Frankfurt summarizes, "The necessities of love, and their relative order or intensity, define our volitional boundaries. They mark our volitional limits, and thus they delineate *our shapes as persons*" (emphasis added). The requirements which love places on us – the "laws of love" as it were – Frankfurt calls *volitional necessities*. When I act in alignment with my volitional necessities, i.e., when I act from love, I act from myself. In a word, I act autonomously.

So far, Frankfurt's view seems to be paradigmatically one of governing from the self: I have a substantive self, and when I act in accord with this substantive self I count as autonomous. But Frankfurt complicates things, for he claims that volitional necessities do not *themselves* ground the authority which makes my acting from them autonomous. Rather, this authority comes from *care for myself*. Love gives me a necessary law insofar as I can only act against what I love on pain of acting against myself. Since to love something means that I want to *act* in certain ways towards it, love will always be partially reflexive: it will always be partially *about myself* insofar as it is about how *I* need to act. And this means that loving something is tantamount to caring about myself. If I betray my beloved, I also betray myself. This brings us to the essential point. Frankfurt believes there is a "primitive human need to establish and to maintain volitional unity" – that is, to be a coherent agent. When I act against myself, this creates "a rupture in [my] inner cohesion or unity". It is, in other words, not simply acting against myself in that I go against what I care about; it disrupts my state of having a stable self at all. To betray myself is therefore to disrespect myself as a person. Since we want – indeed, need – to respect ourselves, we have a vested interest in being faithful to what we love. According to Frankfurt, *this* is the real ground for the authority of love/volitional necessities: "It is our basic need for *self-respect*, which is very closely related to our need for psychic unity, that grounds the authority for us of the commands of love".

This makes Frankfurt's account unique among structural views. Structural views typically argue that once we have found *X* psychological states that represent the self, these states automatically ground the authority of self-governance. But Frankfurt (in

Frankfurtian fashion) incorporates an essential reflexivity into this authorization. He could have stopped with the point that what I love is defining of me, and then had the authority of love fall out of its role in constituting me. But he goes a step further and argues that the real authority comes from a fundamental attitude I have towards myself. I care about having – about *being* – a self in the first place, and it is *this* care which makes what I love lawful for me.

Frankfurt's account could have been more flexible than the other accounts of self-governance we have looked at. The common problem with psychological accounts is that they claim the self just *is* a certain kind of state – intentions, values, cares, etc. These states necessarily bring certain contents along with them, and whatever particular content I have (e.g., whatever I happen to love) is defining of me. Frankfurt's pinpointing the self as caring *about the self* seems to insert a fundamental dynamicity into this picture, for this kind of abstract self-relation is not fundamentally connected to a particular content. To be clear, by "fundamentally connected to a particular content", I do not mean that other psychological accounts hold that only certain contents will do: these accounts are intended to be substance neutral, after all. Nevertheless, they require particular content to be "filled in" in order to function. To care is to care about something *in particular*; to treat something as reason-giving is to treat something *in particular* as reason giving. In contrast, to care about the self I have is *not* attached to a particular content in this way. It is open-ended, and seems to allow room for me to decide *what* will be defining of my self – room, that is, for governing the self.

But Frankfurt does not go in this direction. He argues that a person cannot change what he loves at will[68]. This gives us the final piece we need to fully understand volitional necessities. They are *necessities* not only because they make categorical demands on us, but because *we cannot help but have them*. We cannot help but love what we love. With this, Frankfurt makes his account a solid governing-from-the-self-one, and even seems to preclude the very possibility of governing the self. I cannot do anything to alter what I love; loves may come and go, but this is independent of me.

4.4: Reflections on Caring Accounts

While two out of the three accounts we looked at precluded the possibility of governing the self, caring accounts are not intrinsically incompatible with governing the self. In each case (Shoemaker and Frankfurt's) the incompatibility was based on the additional element that we cannot change what we care about at will, and on giving this point a primary place in the account. While it is undoubtably true that we cannot

[68] Frankfurt says: "Love may appear or disappear; one beloved object may be replaced or joined by another. Changes of these kinds alter the configuration of the will. But the fact that they are changes in the will does not mean that they are up to us. In fact, they are not under our deliberate volitional control. . . It is not up to [the person] whether he is intimately susceptible to the object that he loves." Frankfurt briefly suggests one possibility for change: "It may sometimes be possible for a person to manipulate conditions in his environment or in himself so as to bring it about that he begins or that he ceases to love a certain object; but this does not imply that for him love is a matter of free choice." In this sentence Frankfurt both recommits to his position, since the most a person could do is try to induce changes in himself (which may not be successful), and admits a weakness in his account: a person might not want to love what he loves, in which case it seems strange to say he is autonomous. The mere possibility of change does open a potential space for governing the self: but this is not what Frankfurt is interested in here. As such, his account in this paper is one of governing from the self.

immediately decide to care or not care about X, we are not completely helpless when it comes impacting our cares. Such changes are never instantaneous, but we can affect them.

Despite this compatibility, caring accounts cannot explain governing the self, for the reasons explored above. We much be able to impact our cares, and we must be able to do this in a way which does not essentially limit us to our current set of cares (as Shoemaker suggests). It makes sense to say that I must take some content (perhaps cares) in myself for granted in order to have a position from which to reflect on and shape other parts of myself (again, Neurath's boat-style); however, for us to have the possibility of governing the self, this means we must be able to move freely between the perspective of various cares. I cannot simply claim that some care is fundamental to me such that all authority unquestioningly comes from it, for this would fall right back into governing from the self. But this means that it cannot be the care itself which grounds authority. It is precisely this undetermined element which caring accounts leave unexplored.

Conclusion

All of the accounts discussed have been structural in the sense defined above: they require that the person's will have the proper form, or relation between relevant parts, in order for action to count as coming from the person – that is, for the person to count as self-governing. Furthermore, these accounts do not require the person to undergo any additional procedure to authorize these actions or the motivations behind them: having the correct form is sufficient. The array of structural accounts we have looked at demonstrate that such views can involve sophisticated forms of human activity and agency. Many give central roles to reason, understanding, and deliberation. Nevertheless, an essential passivity

lies at the heart of all these accounts. The person starts with a core set of substantial content which is defining of the particular self they are, and structural accounts take for granted that this content sets the standard for self-governance.

This passivity which remains despite complex forms of activity clarifies the heart of philosophical autonomy. What governing the self requires is that we are active with regards to the very forms of activity that are uniquely human: that we take up the ways we exercise our reason, the affective and deliberative patterns that our understanding of the world (i.e., the ways we imbue/extract meaning from it) charts for us, and this larger understanding itself. Governing the self, in short, requires not just activity but *meta-activity*. This term allows us to make sense of the fact that although many of the accounts just discussed had elements of activity, it was inadequate for the purposes of philosophical autonomy.

We began with the hypothesis that structural views would only be capable of accounting for governing-from-the-self. This hypothesis turned out to be correct. Structural accounts hold that the authority of self-governance is *passively* bestowed by a certain subset of elements in the person's psyche – be it cares, self-governing policies, or even the sub-set of heterogenous elements which cohere together. There is no need for the person to actively authorize this content. But this means that the particular content of the relevant sub-set is taken for granted. Structural accounts, therefore, are intrinsically ones of governing from the self. They share the same intuition: if we can find what is *substantively defining of* the person, the person must be self-governing when they act from this core substance. What we need for an account of governing the self goes beyond this: we need

to be able to impact what is defining of our substantive selves, not simply to act in accordance with the most essential substantive selves we already have. In chapter 3, we will consider accounts which have elements of such meta-activity.

Chapter 3: Surveying the Literature, Part Two

Procedural, Substantive, and Externalist Accounts

Introduction

In the previous chapter we looked at structural views of self-governance, and discovered that these can only be governing-*from*-the-self accounts. More importantly, we pinpointed the peculiar kind of activity needed for governing the self: *meta-activity*, or the ability to decide what form one's activity will take. When a person is meta-active, they do not simply act based on the substantive self they have; they choose the substance of this self which causes them to act in particular ways. It is this which differentiates a view of governing the self from one of governing from the self. In this chapter, we will turn to looking at views which have elements promising for such meta-activity. This will clarify what is necessary for this unique form of activity.

We will begin by looking at *procedural* views. Recall that procedural accounts, like structural accounts, hold that a person is autonomous if their will has the correct *form*; however, procedural accounts also require the person to have *undergone a procedure* to arrive at this form. Because of this, procedural accounts ostensibly involve a form of activity which purely structural views do not: they require active participation from the agent, instead of simply relying on the form the person's will happens to have.

Weaknesses in procedural accounts led many to develop what I will call *independent procedural* views. Such views require not just that a person undergo a procedure, but that this *procedure itself meet some external standard*. While (mere) procedural views include implicit elements of governing the self, independent procedural

views seem to be explicitly trying to provide an account of governing the self. By subjecting the procedure the agent undertakes to stricter standards, these accounts hope to impose limits which will guarantee that the decision the person makes in undergoing this procedure is genuinely their own; in short, that they were not illicitly influenced by outside factors. (Of course, much rides on what characterizes illicit versus acceptable influences.) This seems to be exactly what is needed for governing the self.

Both procedural and independent procedural views are, like structural accounts, substance neutral: they place no limits on the content that one can autonomously will. This is a *prima facie* strength of such accounts: whether or not one counts as *self*-governing seems to be independent of any external evaluations of the content one wills (e.g., whether it is morally acceptable or not). But weaknesses found in both procedural *and* independent procedural views lead some authors to conclude that merely formal accounts do not have enough teeth to guarantee autonomy. *Strong substantive* views include such requirements on the content one wills. We will consider each of these views in this order.

Section 1: Procedural Accounts

1.1: Active Hierarchical Accounts

Several of the views we looked at last chapter had hierarchical elements: they gave some element of the agent's psyche ascendancy over another element. (Bratman's self-governing policies, which determine which of the person's multiple motivations should be given status as a reason in deliberation, are a good example.) But hierarchical *views* hold that it is *the higher-order nature itself* of some specified mental state – be it a desire, attitude, or whatnot – which grounds authority. I call the following views *active*

hierarchical – and I categorize them as *procedural* accounts – because they require the person to *actively take up* this "higher-order" standpoint in order to endorse (i.e., decide to identify with) lower-order elements of their will.

Hierarchical views originated from Frankfurt's paper "Freedom of the Will and the Concept of a Person". In this paper (and the next we will briefly look at) Frankfurt was explicitly concerned with moral responsibility, not autonomy. When we assess his views for their capacity to ground philosophical autonomy we must remember that these criticisms do not serve as diagnoses against the account's intended purpose.

Frankfurt argued that a person actively participates in producing an action when the *first order motivation* which drives him to act is endorsed by a *second order volition*. A second order volition is more than a second order desire, which would simply be the wish to *have* a first order desire; a second order volition is the desire that the first order desire be what *actually moves* one to act. Frankfurt was clearly talking about a form of self-governance: he was trying to delineate the set of desires which genuinely belong to the person such that when he acts on them, it can be said that *he* acts.

Frankfurt's view leads us down the path towards active decision making and reflection. There is some ambiguity in this initial paper, as discussed in the Introduction (footnote 12); to summarize, my understanding of Frankfurt's position is that we consciously *endorse* second order volitions without necessarily arriving at them through conscious *deliberation*. But the very nature of *higher-order* attitudes indicates at least minimal *reflection on* lower order attitudes. Furthermore, Frankfurt himself speaks of "decisive commitment" to first order desires. Interpreted this way, Frankfurt's account is

one of at least partial self-constitution. The person consciously commits himself to one of his first order motivations, thus making this desire fully "his own". In short, he decides what his self will be. Hierarchical views are thus quite promising for governing the self.

However, once we take a closer look we find that Frankfurt's initial view was much more lax. He explains that "a person may be capricious and irresponsible in forming his second-order volitions and give no serious consideration to what is at stake. Second-order volitions express evaluations only in the sense that they are preferences. There is no essential restriction on the kind of basis, if any, upon which they are formed"[69]. The "reflection" implicit in higher-order volitions can be exceedingly minimal; they can be made essentially on a whim. Frankfurt acknowledges that many of his examples "may suggest that [higher order] volitions . . . must be formed deliberately", but he corrects this misapprehension: "the conformity of a person's will to his higher-order volitions may be far more thoughtless and spontaneous." This may seem to contradict his idea of a "decisive commitment", but in fact, Frankfurt characterizes this idea primarily in terms of lack. We commit to a lower-order desire when "there is no reservation or conflict" about wanting *this* desire to move us to action, and when "there is no room for [further] questions" about this. A "decisive commitment" is much less rigorous than it first sounds.

Frankfurt's view in this paper is therefore too weak to work for philosophical autonomy, although it is richly suggestive of ways it could be made more robust (for example, by adding stricter requirements for reflection and active decision-making). Such leniency makes sense given that Frankfurt's intention was to make sense of the grounds for

[69] Frankfurt makes this point in a footnote.

moral responsibility; its inadequacy for autonomy demonstrates the stricter requirements for the latter. The essential problem is that second order volitions are given such flimsy grounds. I can decide simply on the basis of a caprice of the moment. On most views, such a decision would not even be governing-*from*-the-self, since it could be based on a mere whim instead of on my deeper or more enduring self. For our purposes, the main worry is that such a decision is not truly within the person's control. What I happen to prefer or desire in any moment is most likely influenced by things beyond my control – not just what I was socialized to want or to value, but also what I was taught to see as inappropriate and therefore don't even consider as an option. Beyond this, what I happen to unthinkingly want is probably based on innate interests, which again were simply given to me. Furthermore, the "decision" involved can apparently be so unreflective that one wonders why we should call it a decision at all. In short, higher-order decisions are promising since they seem to allow me to act on and shape myself; but we need to set stricter parameters for what grounds such a decision if it is going enable me to act on myself in a genuinely meaningful way[70].

[70] This point is related to two criticisms of Gary Watson's. Watson pointed out that there seems to be an infinite regress of possible higher orders, which indicates that second (or even third, fourth, or nth) order volitions do not automatically have authority-conferring status. If we need second order volitions about first order desires to give authority to those desires, then don't we also need third order volitions to grant authority to second order volitions, and so on? Relatedly, Watson pointed out that a second order desire (and any higher order desire) is in itself simply another desire, just with a special kind of content. Both criticisms show that hierarchy itself cannot be what grounds the authority of self-governance; that is, it cannot explain how I am able to act on myself in meaningful ways. Frankfurt actually foresaw this regress, and he attempted to cut it off by emphasizing that our commitment to the second order volition must be "decisive". In short, he implicitly acknowledged that it was the decision doing the work. Watson presses Frankfurt on the weakness of his conception of decision, just as I did above. The

One obvious suggestion is to require the decision to be based on active reflection. Frankfurt attempts this path in "Identification and Wholeheartedness"[71], where he tried to redeem the notion of decisive commitment. Here, he claims that we are driven to reflect on our first order desires when we see some problem with them: when we are ambivalent about whether to endorse or reject them, or when they conflict with other desires we are committed to. This reflection ends when we have resolved the issue which drove us to reflect in the first place. If we are fairly confident with the answer we have lighted upon and expect that any further reflection would merely confirm this answer, ending the reflective sequence is not arbitrary[72].

Frankfurt's view here is more demanding since it adds the requirements of active, conscious reflection and decision. Furthermore, he explicitly claims that this decision is *self-constituting*: through the decision, we *make* certain desires and motivations a proper part of ourselves. Decision, Frankfurt argues, is essentially something we do to ourselves. (To be clear, by "self-constitution" Frankfurt does not mean that we create the desires we have. To paraphrase him: while we may not be responsible for having certain characteristics in the sense that we did not cause them, we are responsible for our character when we take responsibility for certain of our characteristics such that we incorporate them

---

conclusion is the same: we need a more robust account of what goes into this decision and commitment to make sense of its ability to ground authority.

[71] Although this paper has "wholeheartedness" in the title, the view here is not to be confused with Frankfurt's coherence view which also emphasizes wholeheartedness. See chapter 2, section 2.1 for more detailed discussion.

[72] Frankfurt claims that such a decision "resounds" through all possible higher orders because we are confident that further reflection would be pointless.

into ourselves. While many desires occur in us, it is only in deciding to endorse them that they *become* us.) Frankfurt's view in this paper, therefore, seems to make good on much of what was so promising in his initial hierarchical account. It centers acting on the self such that we can shape this self.

But even in the case of these firmer requirements, reflection simply by itself is not be enough for philosophical autonomy. Frankfurt does not look critically at the reasons why a person would decide the way they do. The process of reflection itself could be problematically shaped by socialized or indoctrinated values, beliefs, desires, and so forth. This criticism is not new. Irving Thalberg points out that our higher-order volitions might not be grounded in self-authority: they could be the result of societal conditioning, or "a cowardly second thought"[73]. Many feminists have lodged basically the same complaints: when we have been socialized into accepting certain beliefs, values, and so forth, then typically our higher-order, reflective stances reflect not what *I* want at some "authentic" level, but what society has *taught* me to want. In such cases, first order desires seem to have *more* authority than higher order volitions (at least, if it is an account of governing *from* the self we are after). In sum, we need to have stricter standards for not only the *amount* of reflection undertaken, but for the *quality* of this reflection. Several of the accounts discussed below attempt to provide such standards.

Not all of these criticisms may demolish Frankfurt's view, which was only meant to be one of moral responsibility. For instance, John Martin Fischer argues that Thalberg was unfairly critiquing Frankfurt's account against the higher standards that would be

[73] "Hierarchical Analyses of Unfree Action"

required for autonomy[74]. However, they are all useful criticisms for our current project, which *is* concerned with autonomy.

1.1.1: Reflections on Active Hierarchical Accounts

Although the hierarchical accounts we have looked at here will not suffice for governing the self, they still seem to involve elements which will be necessary for it. Governing the self, by definition, requires *acting on* the self. Furthermore, it will necessarily require *reflecting on* the self, as we struggle to come to terms with the ways the contents of our selves have been given, and to take an active role in shaping this self. Finally, this means that we must make an active, conscious *decision* about the selves we will have. Acting on the self is the only way to ensure that we are not simply "going along" with the selves we already have, but are actually taking steps to choose this self. Since governing the self requires acting on the self, it will necessarily involve hierarchy in at least this sense – although one can argue whether this will necessarily involve second order volitions. But reflection will need to meet some requirements in order to ensure that the decision it issues counts as genuinely coming from the person herself. Using Frankfurt's terms, what we need is for the process *behind* the identification to also be genuinely the person's own. What precisely these requirements are needs to be determined.

A criticism of Gary Watson's against Frankfurt's initial paper complicates several of these points[75]. According to Watson, people are not primarily concerned with their own desires, but with what concrete action to take in the world: that is, their practical focus is

[74] "Mission Creep"

[75] "Free Agency"

not primarily directed at the *self*, but at the *world*. On the one hand, this point is less relevant for considerations of philosophical autonomy, which is something we can assume comparatively few people have achieved, as opposed to moral responsibility which we are inclined to attribute to most adults. (Again, Frankfurt and Watson were concerned with the latter, not the former.) Nonetheless, Watson's point gestures towards a concern that philosophical autonomy will involve excessive self-absorption. As an initial response, I do not think this need be the case. Our self-understanding and awareness is typically tied up with how we understand the world. In reflecting on what we think we should do – a first order question – we are also implicitly deciding which of our motivations we think should guide us – that is, we are also thinking about ourselves, even if only indirectly. Nonetheless, it *does* seem that philosophical autonomy will require us to think explicitly about the self a lot. This is a concern which I believe my account of philosophical autonomy will be able to address, as discussed in Chapter 4.

In cases where our ability to reflect has been largely co-opted by problematic forms of socialization, many may be suspicious of the ability to reflect. This may lead some to think that we should emphasize getting in touch with authentic first order desires. But again, simply acting on first order desires is not enough for philosophical autonomy, since innate traits were also merely given to us. We need to reclaim our ability to reflect, not give up on this ideal as hopelessly contaminated.

One final idea worth remembering from Frankfurt is the thought that *caring about the kind of self I have* is essential – to my personhood in general, and to self-governance in particular. (It is interesting, given these high-stake concerns, that Frankfurt initially

claimed we could form higher order volitions so cavalierly.) Since developing one's philosophical autonomy is a difficult task which will almost surely need to be consciously undertaken, it seems clear that caring about one's self will be a key part of the *motivation* to take up this project. But Frankfurt's idea goes deeper than mere motivation; it embeds caring about the self at the very heart of personhood, and hence at the very heart of the self I hope to govern. I believe this captures much of the internal richness which we think make persons unique among living beings. This will likely play a role in governing the self.

Section 1.2: Evaluative Accounts

One of our major themes has been the potential importance of *reason* for self-governance. In the last chapter, I identified a few elements of human action which reason makes possible: it allows for *new degrees and forms of control*; it allows us to *justify* our actions to ourselves and others; it gives us a *richer relationship to ourselves* and a *richer understanding (or interpretation) of the world*.

The importance of reason is often cashed out in terms of *values*. Values seem to capture several of the elements reason makes available – in particular, the ideas of *justification* and *richer meaning*. Agnieszka Jaworska offers an explication of values which nicely illuminates these connections[76]. She claims there are three key marks of values. 1) We take our values to be *correct*, "or at least correct for us". From this follows several corollaries: we feel that if we no longer valued what we currently do, this would be "an impoverishment, loss, or mistake"; we can usually give an articulation of *why* we believe our values are correct; and thus our values are open to rational criticism and revision. 2)

[76] "Respecting the Margins of Agency: Alzheimer's Patients and the Capacity to Value"

Our self-worth is typically entwined with how well we live up to our values. 3) Our values are frequently independent of desires to have (or avoid) certain experiences, and hence express concern with something wholly outside myself[77]. These three characteristics embody the valences reason gives to distinctively human action. 1) has to do with justification; both 1) and 2) have to do with the ability to act on, or at least reflect on, ourselves; 2) demonstrates a richer self-relation; and all three embody the richer world of significance humans live and act in, based in our cognitive capacities for enlarged understanding. Values thus seem to capture much of what makes for characteristically *human* actions, and they capture this in a way that makes clear how such action is more robustly the *person's own*.

Jaworska herself thinks that although values can ground a minimal form of self-governance, they are insufficient for autonomy. Such "full-blown autonomy involves not only acting on one's own principles and convictions, but also the ability to scrutinize these principles and to revise them in light of critical evaluation, so that they are well-articulated and robust", and this means that "capacities [other than the ability to value] are necessary to further develop and perfect autonomy". Jaworska thus highlights precisely those concerns which, taken to their logical conclusion, necessitate an account of governing the self.

We will look briefly at two views which hold that values involve more robust and reflexive activity than Jaworska's explication, and which therefore seem promising for a

[77] Of these three key marks, Jaworska only argues the first is actually necessary for valuing; the second is sufficient, and the third is merely indicative.

more potent notion of self governance. Both the accounts we will look at are procedural in the above sense that they require the person to undergo some process to arrive at their values (this process being the way these accounts are "more robust" than Jaworska's).

1.2.1: Watson on Platonic Values

Evaluative views were first suggested by Gary Watson, who presented his view in contradistinction to Frankfurt's hierarchical one. His key move is the idea that there are two sources of motivations: desires and values[78]. These are differentiated not in terms of their content (e.g., lower-order vs. higher-order desires), but in terms of their source. Our values are what we, "in a cool and non-self-deceptive moment", *judge to be integral to a fulfilling and good life*. So understood, values are based in *reason*. (To be sure, Watson does not hold that they are based in reason *alone*, but rather that rational reflection is the necessary ingredient which *makes* something a value.) After such a value judgment, I then come to desire the thing I value. This means that values are a kind of desire but cannot be reduced to simply desiring something. In contrast, desires independent of such a judgement of goodness are *just* desires.

Watson argues that a person is best represented by their values, and so it is values that ground which of her actions count as self-governing. This means his view has aspects of governing from the self. However, Watson's view is best understood as an account of governing the self because of where it places the authority of self-governance. The authority of self-governance comes not from the values themselves, but from the fact that these values are based in my *judgment* that X is worthwhile or correct in some way (or at

[78] "Free Agency"

least, correct for me). In other words, what gives values their authority is that *I make a decision* to have these values[79]. It is because I make a judgment that the value becomes defining of me: I shape myself. In short, the standard of self-governance is determined not by the self I already have, but the self (or the life) *I decide* is worth having[80]. Furthermore, the fact that the relevant judgments are about what is worth having means that these decisions are robust – they are not just based on a whim. This form of governing-the-self is thus more rigorous than Frankfurt's initial view of second order volitions.

Watson's view is therefore rather promising for governing the self. However, it runs into a familiar problem: what if the decision process behind our making certain value judgments is itself tainted by illicit influences? Socialization instills certain values into us from a very young age; how can we ensure that these values do not perniciously influence our subsequent reflection? In short, how can we be sure that our judgments are freely made? Watson's account does not tackle these issues. As such, it alone cannot provide an account of philosophical autonomy.

It also seems that the category of "values" is too restrictive, and the proper domain for governing the self is more encompassing. Depending on how expansively one wants to interpret "valuing", actively shaping myself means shaping more than just my values. It

[79] To clarify, the decision need not *create* these values ex nihilo – the decision may rather be an endorsement of what I find I am inclined to value. But in either case, what makes something a value is this decision.

[80] We can compare this with Ekstrom, discussed in chapter 2, who similarly had aspects of both governing the self and governing from the self, but who defined authority in terms of governing from the self.

involves impacting my broader interpretations and understanding of the world, my deliberative processes more generally, and the affective aspects of myself.

1.2.2: Charles Taylor on Radical Re-Evaluation

In "Responsibility for Self", Charles Taylor does not espouse *values* specifically, but *evaluation* is crucial to his view[81]. Taylor argues that we are responsible for ourselves when we live in accordance with what we find to be most meaningful or valuable. His account is a procedural one, since it emphasizes the extensive process by which we come to discover and fully articulate what we find to be most meaningful. Nonetheless, Taylor's view is largely one of governing *from* the self, since we *discover* what is most important to us, rather than shaping our essential selves.

Taylor shares many of Frankfurt's key intuitions. He takes up Frankfurt's definition of a person as a creature who evaluates its own desires, but goes on to clarify which terms of evaluation are genuinely distinctive of such a creature. Taylor claims there are two ways we can evaluate a desire: we can *weakly* evaluate it, in which case we simply judge how much we want the desire's corresponding objects; or we can *strongly* evaluate it, in which case we judge whether the desire is worthy in some way (be it good, virtuous, profound, refined, or so on). While weak evaluations are merely quantitative, strong evaluations judge

[81] Although Taylor uses the term "responsibility", he does not seem to be using this term in the sense of moral responsibility. He is not concerned with interpersonal practices of praise, blame, and holding accountable; he is concerned with the relationship one has to oneself. Here's a characteristic quote: "The human subject is such that the question arises inescapably, which kind of being he is going to [be] … He is not just de facto a certain kind of being, with certain given desires, but it is somehow 'up to' him what kind of being he is going to be." I therefore take Taylor to be concerned with self-governance which is more closely aligned with autonomy.

the quality of the desire according to some richer, more meaningful standard. This means that we can *articulate*, or give a *rationale for*, strongly evaluated desires in a way we cannot for weakly evaluated desires. This is why strongly evaluated desires are connected to my being a certain kind of person: if a creature cares about the kind of self it is, it is strongly evaluated desires which are essential to defining this self.

If we are to be responsible for our selves, then, it seems we need to be responsible for our *strong* evaluations. Taylor argues that the key move here is articulation. An articulation is an "attempt[] to formulate what is initially inchoate, or confused, or badly formulated". As we struggle to explain in qualitative terms what is important to us, we try on different articulations. Articulations can be more or less faithful to what they are trying to clarify, and so we are compelled to refine and clarify them. This requires an ongoing process; indeed, the deeper and more important something is to a person, the more difficult it will be to articulate, and the more work he will need to do to refine this articulation.

This is what leads to Taylor's concept of *radical re-evaluation*: "The question can always be posed: ought I to re-evaluate my most basic evaluations? Have I really understood what is essential to my identity? Have I truly determined what I sense to be the highest mode of life?" This is *radical* re-evaluation because it means that nothing is beyond revision: nothing is safe from being asked "have I articulated this correctly?". This is particularly true of our deepest values. Since they are the deepest, they inform all the others - and because they are deepest, they will be the hardest to articulate. This means that on the one hand, I will be compelled to re-evaluate and refine my understanding of them; and on the other hand, this redefinition will bring the foundations of my selfhood into question.

This radical re-evaluation is what leads to a person who is *fully responsible* for himself – what we can call fully self-governing.

Taylor's account initially appears to gesture at governing the self. This is what the idea of being responsible for our strong evaluations pushes towards: by reflecting on the evaluations which are most defining of personhood, we can critique and revise them, thus shaping our own personhood. But Taylor ends up giving an account which seems to be a deep version of governing from the self. The criteria we have for questioning our evaluations is essentially "are these evaluations true to my deepest self"? This is perhaps best described as a "self-discovery view".

Taylor captures much of what is appealing about centering values and rational capacities. His idea of strong evaluations captures not just the ability to justify our actions, i.e., to act for reasons, but the rich world of significance humans live in and the complex ways of relating to ourselves entangled with this. He presents a sophisticated version of authenticity, and once again demonstrates that a view of governing from the self can involve complex forms of activity. To remind the reader, I do not want to deny the obvious value which authenticity (and governing from the self) has, nor deny that striving to be more authentic is a worthwhile project. But authenticity is different from governing the self. Governing the self is concerned with being fully responsible for the self: it wants to make sure that I am not simply outsourcing who I am to external forces and influences. Although we have mainly focused on the potential pernicious influence of socialization, innate traits are similarly inadequate for governing the self since they are also simply given to me.

Taylor's idea of responsibility for oneself is extremely relevant for governing the self. To be responsible for oneself is to not just let oneself be determined by external forces, which "work through" the self as their vehicle. This requires self-understanding, as Taylor makes clear with his emphasis on articulation; a self-understanding which allows us to meaningfully impact who we are and how we act. But deep authenticity cannot allow us to take full responsibility for the self in the way philosophical autonomy requires.

Section 2: Independent Procedure Accounts

Procedural views insist that to be self-governing, a person's will needs to not only have the right form but to have undergone a certain process to arrive at this form. This makes them promising for governing the self since they emphasize a process by which we impact the substance of our will. But as we have seen, procedural accounts are not sufficient for governing the self. The basic problem is that we need this process by which we shape our selves to *itself* be attributable to us. Without any additional constraints, there is no way to ensure that the process not been indelibly shaped by outside forces. *Independent* procedural accounts add such constraints in an attempt to ensure that the process is suitably independent.

2.1: Dworkin on Independent Authenticity

The grandfather of independent procedural accounts is Gerald Dworkin. He argues there are two elements to autonomous action: authenticity and procedural independence[82]. These correspond to the motivations I act on and the process by which I authorize these motivations. "Authenticity" means, as usual, that the motivations are "genuinely my own",

[82] "The Concept of Autonomy".

while procedural independence means the process by which I have come to make these motivations my own is itself genuinely my own. Dworkin adopts Frankfurt's idea of second order volitions as fulfilling the authenticity requirement, and for procedural independence he emphasizes the reasoning behind these second-order volitions[83]. The emphasis on process is what make Dworkin's view, like all procedural views, promising for governing the self; the emphasis on an independent process is the new element he introduces. On Dworkin's view, it is the agent's *active and independent reasoning process* which forms the ultimate ground of authority.

But what are the requirements which will ensure that our reasoning is suitably independent? The main point Dworkin focuses on is that it must not be manipulated: "our reflective and critical capabilities need to maintain their integrity*"*. This is a good start, but we need more details. What are the criteria for the integrity of our reflective and critical capacities? Dworkin rejects the idea that integrity requires that such faculties be unshaped by socialization. Just because our ability to reason and think critically has *shaped* by others does not automatically mean that these faculties have been "*manipulated"*. This seems correct, since we only learn to reason at all under the guidance of others. Dworkin wants to allow for positive forms of socialization, and he is right to do so. But how can we tell the difference between social influences which promoted our reasoning faculties, and those

[83] It's important to note that Dworkin's addition of this requirement is not necessarily an implicit criticism of Frankfurt's (identification-only) view. Frankfurt was talking about a form of agency central to personhood and moral responsibility, and Dworkin is talking about autonomy. Dworkin may have agreed with Frankfurt's account as sufficient for its original purpose, and simply proposed the addition of procedural independence as necessary for autonomy specifically.

influences which distort them? These are precisely the details which Dworkin leaves to be filled in. The next few accounts we look at propose different ways of filling in these details.

2.2: Christman on Personal History

John Christman approaches the question of how we can ensure the independence of the procedure in a unique way: he suggests we need to focus on the extended history by which we came to acquire certain motivations[84]. This contrasts with Dworkin (and several others) who held that although we cannot control the motivations we are initially given, we can subsequently endorse them and make them our own. John Christman argues that by focusing on this history, we can do away with the requirement of endorsement altogether: so long as the *way* we acquired a desire meets certain requirements, we do not need to actively identify with it for it to count "as our own". Christman proposes several requirements meant to ensure that we had sufficient control over the process of acquisition. Interestingly, he holds we can count as having had sufficient control even if we did not *actively direct* this process.

Christman suggests three conditions to this end. (1) The person (a) *did not resist* the formation of the desire because she was "attending to" this process, OR (b) she *would not* have resisted if she *had* been attending to it. By "attending to", Christman simply means that the person was aware of and monitoring the process. (2) An agent's lack of resistance must have taken place under *conditions that allowed for self-reflection*. The person *could have* become aware of "the beliefs and desires that move her to act" and developed a sufficient understanding of their larger implications and meaning – that is, of *the larger*

[84] "Autonomy and Personal History".

significance of what happened (what she *let* happen) to her. This condition is meant to address the possibility that a person's lack of resistance could itself have been manipulated if her ability to fully understand what was happening to her was curtailed. (3) The person must have been capable of (i) *minimally rational* and (ii) *non-deceptive self-reflection*. Condition 3 is partly meant to further specify what is required for condition 2 – that is, what is required for the agent's lack of resistance to have been *free of manipulation*. It is also meant to place requirements on more internal obstacles to self-reflection, such as neuroses and pathologies.

Christman's view is essentially based on "normal cognitive functioning" as applied to processes of desire formation. If I have (minimal) normal cognitive functioning – if I am self-aware, minimally rational, and non-delusional – I will be aware (or *could have become* aware) enough of the process by which I acquire certain desires and preferences to be able to influence this process if I so desire. I will be in a position to go along with the change (even actively cultivate it) or to resist it. It is significant that on this picture I can count as having controlled this process even if I did not actively direct it. All that matters is that I *could* have impacted it if I wanted to. Christman's picture is largely counterfactual.

It is also surprisingly passive. It emphasizes my ability to monitor and supervise what is happening to me, without requiring that I actively direct or even be fully aware of what happens to me – I only need to be able to *become* aware. I short, my consent (my endorsement) is constituted by my minimally aware complacency. The things which undermine autonomy are those influences or circumstances which vitiate this ability to attentively monitor changes in myself, and which thus cut off my ability to respond. These

circumstances are harmful because they hurt my reflective abilities; they make me "less able to evaluate [] *from my own point of view*" the things which happen in me (emphasis added). But this "reflection" and "evaluation" is minimal: it does not require me to *actively reflect*, or even to be reflectively aware of what is happening to me, but simply the ability to *become* reflectively aware of what is happening to me. This passivity is rather surprising in an account meant to ensure that the person had proper control over his desire formation.

Christman was explicitly motivated by the need to have an account of autonomy that addresses oppressive socialization; that is, by precisely the kinds of concerns we have been raising. His hope was that by turning to the history behind desires, we could address these harmful influences. However, Paul Benson compellingly argues that Christman's view fails to address this central concern. The key problem is that oppressive socialization can impact the self which is minimally rational and self-aware. To use one of Benson's key examples, a young woman who has already been socialized may be aware of all the implications involved in acquiring a desire to be beautiful – for instance, that it causes her to objectify herself, and to value her appearance over other aspects of her personhood – but precisely *because* of her socialization, she embraces these aspects. She *wants* to be objectified, and she genuinely believes that appearance is one of the most important values a woman can embody. In short, Christman's account falls into the same old problem: socialization can perniciously influence the very requirements he poses to guarantee autonomy. The reasons why we might accept certain changes in ourselves may themselves be illicit; "my own point of view" may have already been shaped in problematic ways.

Alfred Mele and James Stacy Taylor have both offered criticisms which, while less directly relevant to our project, highlight some important points. Mele shows that even when an agent approved of the historical process by which she came to have a desire, this does not guarantee that all *subsequent instances* where this desire moves her to act are themselves autonomous[85]. Although the agent might approve of the desire in general, she might not approve of acting on it in each particular instance. This means Christman is wrong to completely discount the relevance of contemporaneous endorsement. Relatedly, a person may have approved of the process by which she acquired a desire, and yet no longer want this desire to continue to motivate her. Endorsement of the process is not sufficient for endorsement of the motivations forever after. Based on these criticisms, Christman added a fourth criterion: that we need to be "minimally rational" regarding the desire in question, meaning that the desire does not cause us to act in ways that conflict with other motivations we want to be operative in deciding any particular action[86]. Christman thus ends up acknowledging that a historical account cannot be sufficient to ensure the agent's autonomy regarding the desire in question; we need a synchronic account of autonomy as well.

Finally, James Stacy Taylor has pointed out that Christman's account does not even provide a *necessary* condition for autonomy[87]. Someone may disagree with the process by

---

[85] "History and Personal Autonomy"

[86] "Defending Historical Autonomy: A Reply to Professor Mele"

[87] "Introduction", *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy.*

which he came to have a desire or interest, but then subsequently decide, despite its problematic origins to make this desire a part of his authentic self. While we would want to be sure that this new process of endorsement is sufficiently independent, it seems plausible to me that this is compatible with governing the self. Indeed, I think that this is the typical case which governing the self will operate on. We will always be given an initial set of desires, values, and interpretations. To be philosophically autonomous we need to be able to address these pre-given contents, possibly incorporating them subsequently. But this will require an explicit act on my part: first of reflection, and then of endorsement or rejection. We need, in short, the very act of endorsement which Christman was hoping to avoid.

Christman's account qualifies as an independent procedural account because he tries to specify the conditions under which we can say a person has *independently* endorsed a desire. However, both the *procedure* and the *standards of independence* he proposes are too minimal to capture the agent's active participation in the formation of the desire. Minimal rationality and minimal awareness are not enough, since such rationality and awareness (and the interpretations and values bound up with these) can themselves be negatively impacted. We are still lacking an answer for when rationality counts as "distorted" and when it counts as "healthy".

2.3: Meyers on Autonomy Competency

In her book *Self, Society, and Personal Choice*, Diana T Meyers offers a novel approach to independent procedural accounts. Instead of emphasizing one key process which the agent must undergo in a certain way, she claims that autonomy comes about

through the exercise of a host of different skills. The ability to use this wide range of skills she *calls autonomy competency*. The authentic/autonomous self *comes into being through the agent's exercise* of this competency. (53). The self is therefore dynamic and evolving. Meyers' account appears excitingly propitious for governing the self. However, the picture of the self she presents is still rather self-contained, and therefore remains primarily limited to governing from the self.

Meyers defines "competency" as "a repertory of coordinated skills that enables a person to engage in a complex activity" (56). Her articulation of it as a set of skills highlights that a competency is not an overly intellectualized, reflective, or self-centered process – it can be outwardly focused. There are three core skills necessary for autonomy: self-discovery, self-direction, and self-definition. Self-discovery is the ability to *know* who you are. Meyers is a bit ambiguous about this, but I believe that self-discovery involves two distinct components: understanding the deeper motives which drive you so that you can change them if need be; and uncovering who you want to be, what you most deeply value and desire. Self-direction is the ability to *act on* your authentic values, desires, principles, and so on. Finally, self-definition is the ability to *change and create* what your authentic self is. Meyers seems to argue that there are three main components of self-definition: openness to possibilities, sensitivity to one's own responses, and a coalescing of identity. So understood, self-definition is an extended process, which requires genuine engagement not just with ourselves, but with the world.

Of these core skills, Meyer's idea of self-definition is the unique addition. It is this skill which opens up a space for governing the self. However, if we look more closely at

how Meyers describes the process of self-definition, we discover that it does not bear the fruit of full-blown philosophical autonomy. When it comes down to it, self-definition seems to just be self-discovery. To be sure, Meyers clearly *wants* self-definition to be more than just self-discovery. In particular, she indicates that it requires some sort of *decision* on our part. As we have emphasized, governing the self would require such an active decision. But what would we base such a decision on?

Meyers provides two options. 1) We can decide based on outside standards and values. But which external standards are relevant here? Social and cultural standards can be useful, but of course they can also be autonomy undermining. We need more of an explanation as to where such standards come from, and why they provide a justifiable basis for decision. Meyer's second criterion is less promising: 2) we can self-define based on our *autonomously endorsed* desires, values, and such. But in this case, self-definition just reduces to self-discovery: it is based on what we *want* to value, desire, etc. Saying that we base self-definition on self-discovery is to leave mysterious the primary act of self-definition which was supposed to make it a unique form of reflexive activity. The "openness" Meyers appears to value therefore seems to reduce to openness to new desires I find in myself, or wants and desires that my current, conscious self-conception does not account for. The changes she allows for are ones that I simply *find* in myself; they are not changes I *enact* in myself. To see this, we need only to observe that in each case that Meyers discusses of self-definition, we find the justifying basis for the change is that the person is satisfied with the change. By emphasizing how well an action fits with our selves – and in particular, emphasizing our feelings of satisfaction with this fit – Meyers makes her

account rather self-contained. Everything becomes indexed to what I find appealing, to what fits me. But if all that differentiates the traits we endorse from those we don't is that we *like* the former because they feel natural, comfortable, or intuitive, there seems to be no real role left for self-definition. We are no longer creating or deciding what we endorse, as self-definition suggests – we are simply discovering it. There is no substantial contribution for self-definition to make. We are no longer governing the self; we are governing from the self.

Although Meyer's account holds promising seeds for philosophical autonomy, the picture of the self she presents remains self-contained and remains limited to governing from the self. This may be a surprising conclusion, since Meyers ostensibly wants to avoid assuming innate traits have the necessary authority for autonomy. She argues that merely "clearing away" the effects of socialization to uncover innate traits is not enough for autonomy. In particular, she emphasizes that I may not like some of my innate traits. Nonetheless, on her view the authority of autonomy comes down to what *satisfies* me, what I feel is right or natural for me. So long as I can recognize and respond to this using my autonomy skills, I count as autonomous. While these might not be innate or immutable traits, they are things I simply find in myself. Once again, this kind of authenticity-based autonomy has value – but it is not philosophical autonomy.

A brief suggestion may help to illustrate how philosophical autonomy goes beyond this authenticity-based autonomy[88]. Sometimes it is precisely those experiences which do

[88] John Fischer asks if we could respond to Meyers by saying that authenticity and autonomy, although closely related, are not the same concept. I certainly agree that philosophical autonomy and authenticity are different concepts (although I will explain in

not feel comfortable or natural which open up the greatest possibility for *radical* change – change not based simply on self-discovery, but on substantial growth. Whenever we come into contact with something that is not intuitive to us, that does not immediately feel right or align with what we have previously discovered about ourselves, this very novelty presents us with something which we could not provide to ourselves – a chance to radically change who we are. When we are exposed to new paradigms, new ways of understanding and living, then suddenly we are no longer limited to just one way of being. This exposure is what opens a space for even greater autonomy – assuming, of course, that we are not closed off to it. And it is those paradigms which are the least intuitive, the least fitting to us as we are, which open up the biggest possibility for change and growth: that is, the biggest possibility for governing-the-self. But precisely because these possibilities are not ones we would naturally think of, or choose, for ourselves, we rely on outside forces, experiences, and persons to present them to us. In this way, governing the self may require us to be open and responsive to not just ourselves, but the world. This is only a suggestion at this point; we will return to it in a later chapter.

2.4: Reflections on Independent Procedural Accounts

The main problem of independent procedural accounts is how to specify the standards which will ensure independence. Of the views we have looked at, Meyers' fares the best at addressing the impacts of socialization because of her emphasis on an open-

chapter 5 how they can be related). The question is if Meyers is wrong to assert that the kind of autonomy *she* is interested in is essentially equivalent to authenticity. I'm inclined to think that some kinds of autonomy – especially some forms of governing from the self – are equivalent to authenticity.

ended and dynamic process. Even if we take X for granted in one instance, since this is placed in context of the ongoing exercise and cultivation of autonomy skills, we always have the possibility to overcome X. Unfortunately, even if an independent procedural account did successfully pinpoint the standards which would guarantee the independence of one's reasoning, without any additional elements it seems such an account would still only be one of governing from the self. This is because independent procedural accounts are usually silent about the role of internally given traits. Again, Meyers' account fairs best regarding this aspect since she wants to explain how we can reject even innate traits. Unfortunately, her view still relies on what I simply happen to be satisfied with, and remains a view of governing from the self. The question of how we can go beyond innate traits or internal feelings of satisfaction is particularly intractable. I characterized Meyers' account as being too "self-contained". This suggests what we need is to get *outside* of the self somehow. The next category of accounts can be understood as allowing us to do this, for they introduce the idea of an external standard.

Section 3: Substantive Accounts

So far, we have only looked at *formal, substance neutral* accounts. These views hold that autonomy is found in the organization of the psyche alone. They do not place any kind of restrictions on the content a person can autonomously will. We discussed in chapter 2 that there are compelling reasons to adhere to a such a substance neutral account. To put restrictions on the kinds of content which one can will in order to count as self-governing would seem to be arbitrary. After all, it's *self*-governance: so long as you are relating to

yourself in the right way such that you are providing authority to your choices, shouldn't you count as autonomous?

But there appear to be some good reasons to impose restrictions on what counts as self-governance. The main worry is that purely formal accounts might simply be unable to account for all the ways that autonomy can be illicitly undermined. In particular, oppressive socialization seems to necessitate posing standards which may be independent of the person. Because the effects of socialization can be so deeply ingrained, it might seem that no purely structural requirements could guarantee that the agent has sufficiently "escaped" their influence. Reflecting, critiquing, basing one's endorsements on reasons or on some previously endorsed standard, values, cares . . . no purely internal requirement seems immune to being shaped, potentially perniciously, by socialization.

Substantive accounts hold that in order to be autonomous, one's will must meet external standards of some kind. These stricter requirements are meant to safeguard autonomy in ways that more formal accounts struggle with. I take it that all substantive accounts of autonomy will agree that it is necessary for the will to have the proper form, and so they are *also* structural or procedural. In other words, substantive accounts agree that it won't be enough to have the "right" content; you must will this content in the right way.

Following Catriona Mackenzie and Natalie Stoljar[89], we may distinguish between weak and strong substantial views. A *strong* substantial view places direct limits on the

[89] Mackenzie, C. and N. Stoljar "Introduction: Refiguring Autonomy," from *Relational Autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self*.

kinds of content one can autonomously will. For example, you might be restricted to content which is compatible with moral requirements, or which is well-based in reality. In contrast, a *weak* substantial view provides criteria for autonomy which go beyond entirely formal views, but which do not directly limit the kinds of content one can autonomously will. For example, you might be required to have a minimal level of self-esteem or self-respect. Weakly substantial accounts go beyond the neutrality of purely formal accounts because they hold that there are certain standards for autonomy which one could fail to meet *even if* one's will had the right form. We will briefly consider weakly substantive accounts before turning to strong substantive accounts.

3.1: Weak Substantive Accounts

Weakly substantive views have all taken the form of emphasizing the importance of certain *self-regarding attitudes.* Such views have their roots in feminist theory, and are (once again) proposed in an effort to combat difficult cases coming out of situations of oppression. The basic intuition is this: unless an agent has a sufficient level of regard for herself, she will be greatly hampered in her ability to be self-governing. If someone believes that she is incapable or unworthy of pursuing her aspirations, or that her goals should always take second place to supporting others, her ability to genuinely self-direct her life and to shape the kind of person she is will be greatly limited. These self-regarding attitudes cannot be accounted for on a view which only emphasizes abilities to self-reflect, reason, or live by one's "most central" self: a person could meet any of the standards proposed by formal accounts while still failing to view herself with the proper regard. In

other words, she might have the right relation to *specific mental states*, but fail to have the right attitude to her agency *considered as a whole*.

Varieties of self-regarding attitudes have been proposed. Paul Benson talks of *self-worth*, and argues that an agent has lost this sense of worth when she no longer "trust[s] herself to govern her conduct competently"[90]. In these cases, she no longer trusts her very agency. Benson notes, that this need not preclude her recognizing that her will and her actions are her own, that is, that they do *come from her*: it's simply that she does not trust her ability to properly govern her will and her actions[91]. Similarly, Trudy Govier emphasizes *self-trust*, describing this as believing that you can rely on yourself[92]. This involves trusting your ability to make judgments; to plan courses of action and execute them; to interpret situations correctly; to abide by your basic convictions and values; to know what your basic motivations and beliefs are; and in general to be able to cope with both external and internal challenges. If one lacks self-trust, one will not be able to govern one's life or one's self effectively[93]. Robin Dillon argues for the importance of *self-respect*, and indicates that this means seeing as *significant* my own particular "needs, desires,

---

[90] "Free Agency and Self-Worth."

[91] Benson eventually parses self-worth in terms of believing that one can properly live up to normative standards which grant that they are worthy agents: for (a simply) example, that one is sane or respectable enough to be an agent.

[92] "Self-Trust, Autonomy, and Self-Esteem."

[93] As Govier puts it, "If we are insecure in our sense of our own values, motives, and capacities, we cannot think and act effectively. We will bend too easily to the suggestions and criticisms of other people, cave in in the face of social convention, lack initiative in overcoming obstacles."

projects, and so on" such that I give weight to them when considering what the best thing for me to do is[94]. She also argues that this involves seeking to *understand* myself, since this involves seeing my own desires, wants, and perspectives as worth investigating, and potentially worth acting on. Since oftentimes oppressive cultures enforce false pictures of reality (including categories of persons) on us, this quest for understanding oneself can be liberatory. Perhaps the most exhaustive self-regarding attitudes account has been proposed by Joel Anderson and Axel Honneth[95]. They argue that the three key attitudes are self-trust, or the willingness to explore our feelings, perspectives, and desires; self-respect, the view that my needs and concerns provide legitimate reasons for acting; and self-esteem, the belief that what I am doing, and who I am, is worthwhile. All of these are essential to believing that I have the requisite authority and right to govern and direct my own life.

These theorists make good points. These self-regarding attitudes are perhaps best described as a background sense of my own worth and abilities which shapes the range of possibilities I take to be open to me. The lower my sense of worth and competency, the more the range of possibilities I will consider for myself shrinks. However, it's fairly clear that self-regarding attitudes can at most form the necessary conditions under which philosophical autonomy can take root and develop. Simply having a sense that I am capable and worthwhile does not yet mean that I am currently governing my life, much less governing myself. Nor do they give insight as to what such self-governance would involve.

---

[94] "Toward a Feminist Conception of Self-Respect."

[95] "Autonomy, Vulnerability, Recognition, and Justice"

Self-regarding attitude views do not work as an account of philosophical autonomy. We will need to go beyond them.

3.2: Strong Substantive Accounts: Reality Tracking

We will look at two strong substantive views which are remarkably similar. Both Susan Wolf and Paul Benson emphasize the ability of the agent to differentiate true from false and right from wrong. I will therefore call these *reality-tracking* views. Wolf is concerned with self-governance involved in moral responsibility, while Benson is focused explicitly on autonomy. However, their views essentially dovetail. Wolf focuses on the contrast between sane persons who can *recognize* reality (even if they have not yet) and insane persons who lack this ability altogether. Benson focuses on people who cannot currently recognize real reasons, but presumably still have the underlying ability to recognize them, and those who *can* currently recognize real reasons. In other words, Benson seems to be concerned with a subset of Wolf's "sane persons". This tracks nicely onto the fact that Wolf was concerned with moral responsibility, whereas Benson was concerned with autonomy. (We could potentially view Benson's account as an extension of Wolf's: someone who merely has the *capacity* to track reality is "only" morally responsible, but someone who is *actually* tracking reality is autonomous). These views are *strongly* substantive because they require that one's will only have certain content: namely, *true* content. (Technically, Wolf is a bit more lenient– she requires only that we are not intrinsically incapable of having the correct content).

3.2.1: Wolf on the Sane Deep Self

In her paper "Sanity and the Metaphysics of Responsibility", Susan Wolf argues for what she calls the sane deep self view. The "deep self" is Susan Wolf's term for a perspective common to three philosophers: Frankfurt, Watson, and Taylor. Wolf explains that these theorists all agree that "the key to responsibility lies in the fact that [for] responsible agents . . . it is not just the case their actions are within the control of their wills, but also the case that their *wills* are within the control of their selves in some deeper sense." This is the "deep self", since it is the self which controls not just actions, but the *will*. In other words, this deep self seems to be the realm of philosophical autonomy. Indeed, we have discussed all three of these philosophers' views as having at least inchoate elements of governing the self. (Wolf's articulation also demonstrates why it has been worth our while to look at views of moral responsibility in addition to views of autonomy.)

However, as Wolf points out, we face an infinite regress. If the "deep self" which governs the will is *itself* given, the core appeal of the deep self – that we are in control of our wills – will be vitiated. Do we then need a *deeper* self? As Wolf succinctly puts it: "Even if my actions are governed by my desires and my desires are governed by my own deeper self . . . Who, or what, is responsible for this deeper self?" Wolf's description makes clear the worry that no matter how "deep" my reflection gets, it seems I will *always be taking something for granted*, something which was simply given to me and which therefore did not come from me. This is, of course, the problem of governing the self – if it didn't come from me, how can it have authority?

Wolf points out that while this infinite regress is a theoretical possibility, the fact of the matter is that it is a psychological impossibility. After a certain point, the reflection will become so abstract that we finite beings cannot carry it out. For Wolf, this is not a problem, since she claims that there must ultimately be some deepest self which grounds all the rest, and logically this deepest self must be simply given. But philosophical autonomy cannot rely on a given "deepest self", and so rules out this answer from the start.

Wolf likes the deep self-view, so long as we understand it to bottom out in a deepest self. But she thinks it leaves out an essential element. This further element is *sanity*: the ability to have one's beliefs and value guided by reality (indeed, controlled by reality, as Wolf strongly puts it). This includes ethical reality – the ability to be able to tell right from wrong. All of us have deep selves which are unavoidable; but if we are insane, then we have deep selves which are unavoidably mistaken.

Why should the fact that an evil despot's deepest self is unavoidably mistaken, whereas we were lucky enough to have deep selves which are unavoidably *not* mistaken, make a difference to moral responsibility? While sanity does not give us additional control over our deepest selves, it gives us a species of control which the evil despot does not have. Since we can tell right from wrong, we have an essential resource in our deepest selves which allows us to critique and revise our selves. The despot is thus unavoidably evil in a way a sane bad person is not. Because he cannot do anything to change his evil self, he is stuck with it. A sane bad person, on the other hand, *does* have the ability to change himself – and because his badness is *not* unavoidable, he is *responsible* for this badness in a way the despot is not.

Wolf concludes that this means the actual requirement for free agency is *self-correction*. We have seen that self-creation is impossible, and therefore too strong. But we have also seen that mere self-revision is too weak, since "the selves who are doing the revising might themselves be . . . products of external forces". Self-correction provides us with a middle ground. It is the ability to revise ourselves on the basis of a solid standard: the standard of reality, and of right and wrong. As Wolf puts it, "only someone with a sane deep self – a deep self that can see and appreciate the world for what it is – can self-evaluate sensibly and accurately". Wolf's account thus captures the essential motivations for adopting a strong substantive account, and is paradigmatic of this type of view.

There is a seeming tension in Wolf's view: she claims that we cannot change our deepest selves, but she also allows for revision. How does this work? I do not think this is an irresolvable tension, but more needs to be said. But the aspect of Wolf's view which is most likely to draw disagreement is that she assumes there is a *moral reality*. This leads to a series of familiar problems: how do we determine what this moral reality is? How can we know what the "true" standards of good and right are, or which thick[96] ethical terms we should be using? But without an agreed-upon standard, it seems we will never be able to tell if someone is genuinely autonomous or not. More importantly, from a first-person standpoint I have no way of knowing if the standards I am using are correct or not. How can I know *I* am sane? And how can I know that my revision, on what is to me an obvious ethical reality, is not itself a kind of insanity? Given that, once again, socialization colors

[96] I'm referring to Bernard Williams' distinction between "thick" and "thin" ethical terms.

our ethical understanding, what is obviously "right" in one society may be obviously "wrong" in another. There is no clear way to track ethical reality, and thus no clear standard for "sanity" here.

3.2.2: Benson on True Reasons

As the title of his paper "Autonomy and Oppressive Socialization" suggests, Paul Benson is concerned with one of our biggest themes: the impact of socialization, specifically on our reflective capabilities. It has been a recurring hope that our reflective powers can allow us to step back from socially inculcated motives, and independently decide whether to accept them or not. But Benson points out that this is not how socialization normally works. It does not just insert motivations into us; it makes us "*internalize its standards*". It "insinuates its lessons into [a person's] most stable views of what they are and ought to be as persons". Indeed, this is the very mark of successful socialization: that the socialized person *accepts* these standards *as correct*, as natural. This means that reflection, hoped to address the influence of socialization, is not up to the task.

Benson wants to allow that not all forms of socialization are oppressive. Indeed, many seem to be autonomy promoting. What makes oppressive socialization *oppressive* is that it impedes a person's ability to properly grasp, appreciate, and respond to reasons. Harmful socialization "limit[s] in well-organized ways what sorts of reason to act persons are able to *recognize*", and can "render[] them unable to *take seriously* reasons" for doing things differently (emphasis added). But in order to be autonomous, according to Benson, we must have at least the ability to *recognize and appreciate* the range of good reasons. It is precisely this ability which harmful socialization systemically impairs.

This is a strong substantive view because it presupposes that there is an objective standard for the reasons we need to be able to recognize. On Benson's view, there are certain considerations which just *are* reasons. This especially clear when he speaks in terms of falsity: "[W]hat feminine socialization aims to instruct women about the value of their appearance is untrue. It is not true that women who deny feminine looks preeminent importance are lazy or selfish or unhealthy or neurotic or not real women". Oppressive socialization limits the reasons we can see by giving us a false view of reality. This means that the reasons we need to be seeing are those which correspond to some ethical or evaluative reality. In short, the autonomous will must be responsive to reality in this way.

Benson's view is quite similar to Wolf's, and so all my above comments apply to him[97]. I think there is something intuitively correct about the idea that autonomy requires us to track and base our actions on "right" or "real" reasons – that is, reasons which accurately represent the world and moral reality. The problem is how we can confidently set the standard for what is *ethically and morally* "right".

Benson is explicitly concerned with oppressive forms of socialization, and argues for his account in terms of how it can address this issue. Given this, presumably Benson hopes that if someone *can't currently* recognize the real reasons she has, she can *learn* to

[97] It's worth noting how their two views do come apart. Wolf focuses on the contrast between sane persons who can recognize reality (even if they have not yet) and insane persons who lack this ability altogether. Benson focuses on people who cannot currently recognize real reasons, but presumably have the underlying ability to recognize them, and those who can currently recognize real reasons. In other words, Benson seems to be concerned with a subset of Wolf's "sane persons". This tracks nicely onto the fact that Wolf was concerned with moral responsibility, whereas Benson was concerned with autonomy. Combining their views, a woman who criticizes other women for dressing "too slutty" is morally responsible for this belief, but she is not autonomous.

recognize these reasons: she can learn to break with her socialization and to become autonomous. This is an advantage Benson has over Wolf: he does not claim that someone either has the ability to recognize reasons or not (he just *is* "sane or insane"), so presumably we can improve our ability to recognize true reasons. The question is how this is possible. If the person has no internal resources because she has so deeply internalized the norms and values of her society, then how can she learn to critique them? One suggestion is that she may be able to learn from others, and thus overcome her lack of internal resources. But what if she has internalized these false values such that they form a key part of her self-worth and identity? In this case, she will likely be internally resistant to such attempts to get her to "see things differently".

3.2.3: Reflections on Strong Substantive Accounts

Reality-tracking views have aspects which seem quite promising for an account of philosophical autonomy. Because they propose an external standard which is independent of socialization, they seem to provide a promising ground from which we might be able to escape socialization to independently govern our selves. By being in touch with "reality", with what is "true", I can be the primary agent in control of who I am and what I do, without outsourcing this to other people's opinions and perspectives. In this way, strong substantive accounts provide a clear pathway to answering the question "What is the standard for when our reasoning capacities are "supported" vs. "twisted" by socialization?" The answer is our ability to perceive truth.

Unfortunately, the advantages of reality-tracking accounts are tied up with their issues. They provide a solid standard from which we may be able to genuinely govern the

self by bringing in objective truth, specifically moral truth. By relying on objective truth, we are given a place outside societal ideology from which to govern ourselves. But the problem is precisely this: how could a person come to know these objective truths if they did not already? The solidity of this standard – truth – was meant to provide us a sufficiently independent space to govern the self, but it is undermined by our uncertain ability to make contact with this standard.

A further concern is that even if we agree that there are some moral "facts" which are universally accepted – e.g., murdering someone for fun – this realm of uncontroversial facts seems rather small and would not offer much guidance for governing the self. Strong Substantive accounts are also silent on the issue of innate traits. Are we to assume that these accounts take innate traits to have unproblematic authority for the individual, so long as they are compatible with moral reality? This seems to be Wolf's view, since she holds that we cannot change our deepest selves. But as we have seen, innate traits cannot provide the authority governing the self requires.

Section 4: Externalist Accounts

All of the accounts we have discussed thus far have been *internalist*: the standards for self-governance they proposed only place requirements on the individual's will or psychology. But there are also *externalist* accounts, which involve standards for the external circumstances a person is in. On these views, autonomy is not simply a matter of what or how the person thinks, relates, or acts on herself: she also needs to be in the right sort of circumstances.

Notice that this is different from claiming that a person needs to have the right kind of upbringing in order to become autonomous – to have been raised in a family or community that supported the development of capacity to become autonomous. In one sense, these are external circumstances which were necessary for the person's autonomy. But such external factors only give the person the capacity to *become* autonomous; whether or not he actually *is* autonomous may be entirely up to him. The current exercise or manifestation of the capacity is different from how we came to have this capacity. In contrast, Marina Oshana argues that autonomy requires having *actual, de facto* control over your own life[98]. Whether a person has such control is not simply a fact about her psychology, but about the situation she is in[99]. Parsed in terms of relational autonomy, an externalist account is a *constitutively* relational account, as opposed to a merely *causally* relational account; it holds that right relations with others are essential not just for the development of autonomy, but for the very state of being autonomous. (Oshana also refers to her account as socio-relational because she largely emphasizes how our relationships with other people impact this control.)

Oshana's account of autonomy is quite an ambitious one – I am not sure many people would count as autonomous on her view. But this is likely a benefit of her account, at least taken in a context of social justice. If autonomy requires these external factors,

[98] "Personal Autonomy and Society"

[99] To be more precise, Oshana distinguishes between autonomy of preferences (that is, whether your desires, values, and so forth are autonomously chosen) and autonomy of persons, and argues that in order to be an autonomous person you must be in de facto control of your life.

many of which are structural and institutional, and we genuinely care about autonomy, then we will be motivated to make changes so that more people can enjoy the conditions necessary for autonomy. This seems like an essential goal for modern societies, and Oshana's account can add theoretical backing to this view.

Externalist accounts pose a question for philosophical autonomy. Governing the self requires being able to choose the values, beliefs, and so on which will be defining of this self. We have been mainly concerned with how the external influences of socialization impact a person's ability to exercise control over their self at this deep level. But how essential is it for philosophical autonomy that we are actually able to translate our authentic motivations into action? Perhaps surprisingly, I think it is inessential. To see this, we can look at one of Oshana's own examples called "the conscientious objector".

The conscientious objector is a pacifist who goes to prison for refusing to fight in what he believes is an unjust war. As such, he loses many concrete opportunities to live as he chooses. He is stuck doing what the prison schedules dictates for him. According to Oshana's view, he has lost his autonomy because he has lost *de facto* control over his life. While the objector has lost *de facto* control over his life, this does not seem to touch the heart of what is important for philosophical autonomy. Assuming his values are his genuinely his own (they have been issued from his governance of himself), this means that going to jail and losing *de facto* control is being philosophically autonomous for him. In

other words, the primary issue for governing the self is *not* being free to act in as many concrete ways as possible.[100]

To be sure, *self-direction* is an important part of autonomy more generally: I do not seem to have successfully *governed my self* unless the principles and values I have set down for myself actually lead me to action (and structure my personhood more holistically, i.e., my internality). Nevertheless, successful self-direction may lead me to take particular actions which then put me in circumstances where I have less de facto control[101]. (It is also important to remember that governing the self is not in the first instance concerned with self-direction).

---

[100] Oshana might reply that this response indicates that one of her other examples - the subservient housewife - is also autonomous, since although she is in a position where she gives up de facto control, she does so in accordance with her authentic values and motivations. But I do not think the case of the objector and the case of the housewife need to be collapsed. There are multiple routes we could go here. One potential route is this: the housewife values subservience itself – this is the content of her value and her will. In other words, she values precisely the giving up of de facto control. This is not the case with the objector. What he values is refusing to participate in something he sees as evil. The consequence of this is that he loses de facto power, but this is not the content of his values. This would be a substantive view, since it holds that the relevant difference between the housewife and the objector is the content of their values, and so it would face its own problems. I am not claiming this is the right answer, and I think explaining how precisely the conscientious objector and the subservient housewife diverge will be an important point for philosophical autonomy to address. Nonetheless, I think there will be a way for philosophical autonomy to deal with this issue.

[101] As this example shows, there is an important difference between someone who is imprisoned as a result of their successful, *self-governed* self-direction and then has their de facto control curtailed, and someone who is randomly imprisoned and thus has their de facto control curtailed. In the first case the person's past autonomy (the choices they made faithful to their values) impacts the meaning of the loss of de facto control, such that it is not *just* a curtailing of their freedom; it is a curtailing which *signifies* their ability to self-govern, to choose their own values and remain faithful to them.

Externalist understandings of autonomy have their value. In particular, they are useful insofar as we want to promote conditions under which everyone has *de facto control* over their own lives – conditions which we should promote. However, externalist accounts are concerned with an essentially different question than philosophical autonomy. They want to know when a person is in fact free to exercise control over their own lives, and so focus on external circumstances. But this means they have little to say about where authority lies in the individual, which is essential for adjudicating internal conflicts of the sort fuller kinds of autonomy are concerned with. As such, we can put these accounts aside.

Conclusion

I started out with a simple enough suggestion in chapter one: we govern the self when we do not simply take the self we already have for granted, but actively shape this self. In the course of the past two chapters, it has become clear just how difficult this is to achieve. I need to not just be active in some characteristically human way, as the views in the last chapter emphasized; I need to be meta-active in the sense that I am deciding on the underlying structure which determines the characteristically human actions I take. This means that at the very least, I must *consciously reflect* on who I am and who I might be, and then *actively decide* on what content and form to give this self. This is what active hierarchical and evaluative views try to capture. But such reflection and decision is not enough to govern the self, for the very standards which we use to reflect may themselves be given by socialization, which impacts on some of the deepest levels. We need a criterion which tell us when reflection is suitably independent from detrimental outside influences. This is what independent procedure and strongly substantive accounts attempt to provide.

Finally, even more "innate" traits, such as predispositions and interests I simply find myself with, are merely given such that they don't come from me. These more "natural" aspects of the self have a role to play in authenticity, but cannot in themselves ground philosophical autonomy. The only theorist so far to consider this concern is Meyers, whose view nonetheless remains one of governing from the self. What all accounts are missing is a story of how I be genuinely meta-active: how I can determine the shape my activity (my agency and subjectivity) will take without outsourcing this very decision to something which will decide for me, thus undercutting the very point of being meta-active. There is still no story for how I can be fully responsible for my self.

These considerations are complicated by a further one: it may be that such "given" aspects of the self are essential for my being an agent at all. Both Buss and Jaworska point out the fact that I *just am* predisposed to pursue, care about, or be interested in certain things is necessary for action to get off the ground. Without these initial inclinations, I would be at a loss for what to do. Furthermore, both Frankfurt and Berofsky have argued that sometimes the very things I don't choose are the most important and enriching aspects of the self: love being a primary example of this.

Governing the self, at least if we interpret the concept in all its intuitive strictness, seems almost impossible to achieve. Does this mean we should weaken the concept, or even abandon it altogether? I believe there is a way to preserve the key motivation behind governing the self while allowing it to be compatible with all the above. This is the project of the final two chapters.

Chapter 4: A New View of Governing the Self

Section 1: The Dilemma

In Chapter 1 I argued that we should distinguish between governing *from* the self, which presupposes a more or less stable substantial identity, and governing the self, which involves deciding what will be definitive of this substantial self. It is this latter concept of governing the self which is central to the uniquely robust kind of autonomy I have been calling philosophical autonomy. In what follows, I use these two terms – governing the self and philosophical autonomy – interchangeably.

In Chapters 2 and 3 we looked at the current state of the literature on self-governance. Some of these accounts were primarily ones of governing from the self; some had implicit elements of governing the self; and some explicitly tried to allow for the possibility of governing the self. In chapter 2 we looked at view of governing *from* the self, and determined that the core concept needed for an account of governing *the* self is *meta-activity*: activity which determines how I will exercise my agency. In chapter three we looked at accounts that had elements of meta-activity, but ultimately concluded that none of the existing views provides a satisfactory account of governing the self.

I believe this shortcoming is not due to any obvious failure of the accounts discussed; rather, it demonstrates just how difficult governing the self is. We started out with a simple idea: we want to be able not only to act from the substantial self we already have, but to shape and decide what this self is in the first place. The core problem is that the shape of any currently existing substantial self has been largely given to us, and thus cannot provide legitimating grounds for the more thoroughgoing kind of self-governance

involved in philosophical autonomy. There are two main issues driving this concern: socialization and (more or less) innate traits.

Socialization is the more obvious culprit of the two – most obviously oppressive socialization. As noted in chapter one, socialization does not simply insert beliefs and values into us: rather, it teaches us that certain ways of using our agency are correct, natural, or should simply be taken for granted. It "co-opts" our agency such that we only direct it in certain ways. This means that our selves are largely decided for us. To govern the self requires that one decides for oneself how one's agency will be directed: the beliefs, interpretations, and worldviews which will shape one's understanding and the basic horizon of possibilities for action; the values, ethical code, and priorities which will guide one's deliberation about which actions to take.

The problem arises when we try to determine what precisely is required to self-determine the contours of one's agency. Put simply: how do we get past the influence of socialization? Socialization does not just give us "basic inputs" (values, desires, and so on); it also determines the way we evaluate these basic inputs. This means that simple reflection on which desires, motivations, values, beliefs, and so on we want to be definitive of our agency will not do, because this reflection itself will likely be informed by the values and beliefs we were socialized into. What is needed is for this reflection itself to be independent, and not illicitly influenced by outside pressures.

But the problem then becomes what the standards for "independence" are. What is required for reflection to be sufficiently "free" of the impacts of socialization? We could say that reflection is sufficiently independent once we have gotten in touch with our deeper

authentic desires or interests. (Recall that innate or intrinsic traits is the second issue of concern; I will say more about this in a moment). But as Paul Benson rightly points out, even our most stable and deeply rooted desires are themselves often shaped by socialization. Alternatively, "independence" could be defined by those desires or interests which are based in reason; that is, values which are based on a rational judgment of what is good or worthwhile. "Reason" and judgments of "the good" may appear to set a standard independent of the norms of the culture one was raised in. But this falls into the same predicament– part of socialization is precisely shaping what we take to be reasons and what we see as worthwhile or good in the first place. We could perhaps go up a level and reflect not just on our current selves and the beliefs, values, and so forth which are integral to them, but on the standards of reflection themselves. But once again, this higher-up reflection will be influenced by socialized values and beliefs. This is the phantom we cannot escape from – no matter how abstract or "higher-order" our reflection goes, it seems it will always presuppose some standard from which to judge, a standard which we simply take for granted and therefore did not choose.

Even *if* we were able to escape the influence of our past socialization, on what grounds could we authoritatively choose a new perspective? How could we be sure we were not simply being influenced in a new way? We might be tempted to assert that there is some independent, objective standard of truth and value, and that what is needed to govern the self is to be able access this objective truth. There are two problems with this

approach. The most obvious one is the proposal of an objective *moral* truth[102]. This is a notoriously controversial claim. The second problem is how meeting this external standard would contribute to self-governance. Given that moral truth is not something I choose, what does being able to follow the moral law have to do with self-governance? There *do* seem to be non-moral objective truths of practical rationality: viz, prudential principles such as always taking the means to a desired end. However, such standards of instrumental reason cannot help us with governing the self. They tell us what we must do *given* that we have certain ends; but their ability to help us *critique* the ends we have is quite limited. Because practical rationality insists our set of ends must be consistent, it enables us to see that if ends A and B are inconsistent, we must choose to give up one. In this way, it can force us to decide which ends are most important to us; in other words, it can help us with the project of authenticity. But as I have emphasized, authenticity is not the same as philosophical autonomy. The latter cannot rely on innate desires. Because practical rationality requires some presupposed content for its principles to operate on, it offers no richer resources for critiquing our desires[103].

Perhaps we have taken a wrong turn. The question "On what grounds can we legitimately choose a new perspective?" has an intuitive answer: I base it on what I deeply

---

[102] Since this standard is meant to help us exercise our agency fully and is therefore a *practical* standard, the idea that there is an objective *moral* standard is central to these views. In other words, we cannot do only with the more palatable idea of an objective theoretical truth.

[103] Another potential kind of objective standard are *aesthetic* standards. However, these are just as controversial as objective moral standards, if not more so, and just as subject to social and cultural influences.

and authentically desire. People typically have at least some innate characteristics – psychological pre-dispositions, natural interests and proclivities, and the like. Of course, these things can *also* be shaped by socialization – for instance, certain interests may be discouraged, or the cultural environment may not provide opportunities for certain talents to be developed and expressed. As remarked previously, even our most seemingly authentic desires can be shaped by socialization. While these points are well made, let us assume that we do have *some* innate desires, interests, and predispositions (a not wholly unreasonable assumption). These more innate factors may not seem to be a problem for governing the self precisely because they are *not* the product of socialization. If we are naturally inclined to them, they must come *from* us, and so they must automatically have the requisite authority for self-governance.

However, the mere fact of their innateness is inadequate to establish the authority of these traits and predispositions. To govern the self means that I decide which parts will be essential to and defining of myself. Simply having certain tendencies and interests is not enough to establish this. One way to explain this is that we typically have a mess of disjointed and conflicting predispositions. To be a reasonably effective agent, we have to create some sort of order out of them. We need a plan for dealing with them – which ones to prioritize, and in which circumstances. (This is a Bratmanian point.) This order must be imposed on them, and so they cannot be the simple grounds of their own authority. Diana T. Meyers provides a parallel argument: what if we do not like certain innate aspects of ourselves? Surely tendencies we wish to reject do not have authority to speak for us. While this point is about traits we do not like, its relevance extends to all innate traits: any

authority they have is not provided by their innateness. Innate tendencies cannot be the ground for governing the self.

One obvious response to this is to adopt a "governing the self from the self" view: we decide which interests and dispositions have authority for us, and how to organize them, on the basis of our *deepest* interests. After all, as we are deciding which traits to endorse and incorporate into ourselves, this decision itself must be based on something. Isn't the most appropriate basis for this decision our own deepest interests or cares? The problem is that if we are committed to the idea of thoroughgoing control over the substantial self, we cannot take *any* innate interests for granted. To do so is to outsource the problem of self-definition to something which was simply given to us. A view of "governing the self from the self" is still primarily a view of governing *from* the self.

This demonstrates just how abstract governing the self really is, and clarifies what precisely is involved, the central insight it is getting at. At the heart of the idea is control over my own agency. I do not let its shape be determined by socialization, and I do not "outsource" it to the things I am naturally inclined to prioritize. To govern myself requires that I take nothing for granted – not even the things that I am innately inclined towards, since these were also just given to me. This is the central point of governing the self, and so it's worth repeating: the goal is to have thoroughgoing control over my own agency, and how I exercise it. This is what is meant by the term *meta-activity*.

But this hope of thoroughgoing control is itself a problem. If I do not rely on anything to guide my decisions about what substance to give myself, how could I even

begin to make such decisions?[104] To get action going at all, it seems that I must *not* be neutral. I must simply *find* myself interested in certain things, or else I will not have any idea what to do with my agency (and no motivation to do it)[105].

Every route we take seems to lead to a dead end. Philosophical autonomy places great demands on us, and these demands appear impossible to meet. Governing the self requires me to be fully in charge of my activity, how I exercise and express it. But it seems I can never be fully in charge of my agency in this way, because at any kind or level of reflection or decision, I must take something for granted. Governing the self appears to require an impossible sort of agency: that I occupy a completely neutral "view from nowhere", or that I am a pure, self-constituting spontaneity that can act on myself without relying on anything else[106].

In summation, governing the self as we have conceptualized it so far seems both *theoretically impossible* and *practically undesirable.* It is empirically unlikely that I could

---

[104] This is a point made by Sarah Buss in "Autonomy Reconceived".

[105] I will also need to have some abilities to pursue some of these interests; and *which* abilities I happen to have will also be merely given to me. Furthermore, the capacity to exercise these abilities often depends on social/cultural structures which are beyond my control. In short, the ability to act at all requires multiple contingent factors which I did not choose. In what follows I will mainly be focusing on the contingencies internal to the will – the things I simply happen to be interested in or find compelling.

[106] I am certainly not the first to make the point that self-governance seems to require that I be a pure, self-acting spontaneity. For example, Galen Strawson in "The Impossibility of Ultimate Moral Responsibility" argues that this requirement of pure spontaneity means we cannot be ultimately morally responsible. Robert Kane's idea of "ultimate responsibility" (from *The Significance of Free Will*) also relies on the idea of a person as "the ultimate creator" of his own ends – and in contrast to Strawson, strives to show how this is possible.

ever transcend my socialization or the historical situation I am born into. Even if I could, innate traits do not themselves give the required sort of authority. And if I *could* theoretically occupy a neutral reflective space, this would make practical reflection and action impossible. The contingent aspects of substantial selfhood which are worrying for philosophical autonomy are the very aspects which seem to be necessary to get agency going in the first place.

One final concern which I have yet to mention, and which I can only adequately address next chapter, is that it would be *politically problematic* to insist that we could change any contingent aspect of the self, since there are oppressed groups which are trying to be recognized and respected precisely for their contingent qualities (e.g., gay and trans persons). To say that they could change these aspects of their selves would be to make it possible to use this argument for oppressive purposes, and may also seem to disrespect these identities as being "merely" contingent and therefore unimportant.

This chapter will be primarily focused on explaining how governing the self is theoretically possible. The following chapter (Chapter 5) will address details of the view, including how it makes room for the practical necessity of contingent aspects of substantial selfhood and the political desirability of actively embracing certain contingencies. In what follows, I will argue that we need adopt a different understanding of the sort of control required for governing the self. There will be three key elements of my view, which will turn out to be inseparable elements of the same core attitude.

I will proceed as follows. First, I will diagnose why governing the self seems impossible: I will argue that the core problem is the underlying notion of control we have

been implicitly using. In order to salvage the notion of governing the self, we need to recharacterize the control necessary for governing the self in terms of an attitude of perpetual openness (the first aspect). However, this attitude of openness requires as its condition the attitude of taking one's self seriously (the second aspect). This second aspect enriches our understanding of what precisely is involved with the first. Finally, taking one's self seriously requires a further attitude as its condition: holding oneself responsible to the world (the third aspect). This third aspect, once again, enriches our understanding of the first aspect. This chapter therefore unfolds as both an argument (first for why perpetual openness is the answer to our dilemma and then of why openness requires taking one's self seriously and holding oneself responsible) and an extended description of what perpetual openness is. Finally, I will end by explaining how this core attitude of perpetual openness counts as governing the self. This will involve us refining our notion of activity – something we will be in a much better position to do once we have all three aspects of perpetual openness on the table.

Section 2: Perpetual Openness

If governing the self seems impossible, this indicates we must revise how we understand this concept and the notion of control at its heart. I initially defined governing the self as *deciding what will be defining of my substantial self*. This suggests that the only way to be in control of our substantial selves is by standing on untainted ground from which we can administer our decisions: untainted in that it is free from influence and therefore does not illicitly outsource our authority to anything else. This definition therefore assumes a kind of control which can only be exercised from an Archimedean point. The conclusion

from our above investigations is that there *is* no Archimedean point. There is no self that exists beyond or uninfluenced by socialization, the norms and concepts and reasons of culture, or innate interests, abilities, or dispositions.

Governing the self must be impossible – *if* it does in fact require such an Archimedean point. Is there a way to understand governing the self which does not require this? What is needed is to revise our notion of control, which will in turn revise our understanding of what it means to govern the self. The alternative I will sketch in this chapter will not itself provide a full account of governing the self – Chapter 5 will include more details – but it will tell us the necessary standpoint we must inhabit.

2.1: Avoiding the Archimedean Point: An Initial Proposal

It is helpful to remember what we are trying to *avoid* in governing from the self. In ordinary cases, we are never *simply* controlled by external influences, since successful socialization effectively teaches us how to exercise our own agency. The issue is with the deeper causes behind why we choose to exercise our agency in the ways we do. In this respect, it seems clear that oftentimes people *are* largely controlled by socialization. Philosophical autonomy wants to avoid our agency being co-opted and outsourced in this way. It asks us to be more actively involved in the exercise of our own agency, as opposed to passively exercising it along the guidelines that we have been given: it requires us to be meta-active. What we are thus trying to avoid is *meta-passivity*. When I say we "*passively* exercise our agency along the guidelines we have been given", I do not mean that we are *entirely* passive in such instances. We may, in fact, appear to be quite active. For example: if a man brought up in a society with certain ideas of masculinity such as that "real" men

do not show any emotion (except anger), denigrates other men who do show emotion (perhaps by telling them "Don't be a pussy"), he is clearly acting. Nevertheless, there is a deep sense in which he *is* being passive: the *ways* in which he decides to act are determined for him. He only exercises his will along these pre-given values and guidelines. In this way, the shape and structure of his will is decided for him. Although though he is active with regards to his particular actions, he is meta-passive with regards to this deep structure. In what follows, I will use the term *meta-passive* to refer to this kind of higher-order passivity.

My suggestion is this: while we can never completely "escape" external influences to a place of pure, self-constituting activity, we can still take up an *active stance* towards these things. This requires perpetual openness. Such openness means that we are always a) aware of the likelihood that we are being meta-passive (in the deep sense explained above), b) open to learning about and reflecting on the ways we are meta-passive, and c) committed to sincerely engaging with what we have thus learned. Such openness is essentially *a meta-attitude towards the whole self*. This does *not* mean that you adopt a "view from nowhere"; rather it means that *you view every part of yourself "in quotation marks"*, seeing each as only provisional. By remaining *genuinely open to changing any part of ourselves* (provided a compelling enough reason – we will come back to this essential point later), we ensure that even if we are currently in the grips of an external influence, its grip is *incomplete*. Openness allows us to avoid being completely beholden to these external factors[107]. (This

[107] The idea of perpetual openness may sound similar to Otto Neurath's analogy of the ship. Originally taking aim at foundational theories of knowledge, Neurath argued that like sailors at sea who are unable to take their ship out of the water to comprehensively evaluate and fix it, humans are immersed in systems of knowledge which we cannot step outside of, and which we therefore cannot critique all at once. All we can do is piecemeal

need not involve being completely indecisive or never having any genuine commitments – in fact, as we will see in section 3, it actually requires us to make certain commitments and decisions.)

Our understanding of the constitutive aim[108] of governing the self therefore shifts. It is no longer focused on *defining* my substantial self – a project which ostensibly has a concrete end goal. Instead, it is focused on *taking up and being more actively involved in my own agency* – an open-ended and continuous project. This is actually a natural shift; we were pursuing the possibility of defining one's substantial self because we were concerned

---

revision, questioning certain parts of our knowledge while taking other parts for granted. We can then revise the parts we took for granted, but only so long as we take some *other* part for granted. Neurath's point can easily be applied to self-constitution. We are always immersed in our current substantial self, and we cannot get completely outside ourselves to decide which parts to keep and to modify. We must always hold a certain part of ourselves fixed from which we can evaluate the other parts of ourselves.

Neurath's metaphor of the ship is hopeful: it allows us to simultaneously acknowledge the difficulties of governing the self while also leaving room for us to work on ourselves. It may look almost exactly like the account I am offering. However, Neurath's metaphor by itself does not tell us what governing the self consists in – it simply indicates that there is space for it. It tells us that revision is possible. But the mere *ability* to revise is not what I am arguing is the seed of autonomy; it is the *willingness and the commitment* to revise. This is an attitude towards the whole self which we must actively take up and inhabit at all times; it is not simply making particular revisions when it becomes clear that they are necessary. Perhaps most importantly, the driving motive of openness is *not* primarily to make your existing substantial self more well-integrated and sound on its current substantive terms, but to become more actively involved in your own agency. This means that revision does not end when we are "satisfied"; our commitment to openness must be ongoing. We must be continually on the lookout for relevant feedback, including when this feedback indicates we may need to make a major overhaul of our selves *even when* this self in internally consistent. The fact that the "standards for success" are *not* internal coherence or satisfaction, that being open requires you to be open to wholly new perspectives that illuminate new forms of passivity within the self, is a large reason why this avoids turning into an account of governing from the self.

[108] By "aim" I mean the defining goal of governing the self: to talk about governing the self is to talk about pursing this goal.

about passivity within agency. We have simply clarified that the project of being more actively involved in my own agency is in fact the primary concern. However, this is an important shift. I cannot "win" the game of autonomy by achieving a final, wholly autonomous substantial self: I can only strive to become *more* autonomous[109]. For this reason, it is most accurate to speak in terms *practicing* autonomy instead of *being* autonomous.

This might seem like a poor consolation prize. Are we merely giving up on an unvitiated notion of governing the self? Once again, we must revise our understanding of activity. It's true that activity in the sense of pure, self-constituting and self-contained spontaneity does not exist for humans; for us, activity is always engagement with an outside context and outside forces. But this is not necessarily an inferior form of activity. A being that was entirely self-contained would be immune to anything but its own forms of activity, but its forms of activity would therefore be essentially static. What it did (including the basis for what it did) would be formulaic; almost, dare we say, passive. A quasi-example of this are animals which do act mainly on instinct (i.e., "lower" animals which do not need

[109] We can explain this idea by contrasting it with the more familiar concept of identification. Identification is the idea that by endorsing a desire (interest/value/etc.), we make it a part of ourselves. If we stick with the initial definition of governing the self as "deciding what will be defining of my substantial self", then identification seems to be precisely what is needed – provided we can find a solid ground from which one can say with genuine (i.e., uninfluenced) authority "This is who I am." Of course, we have seen in extensive detail that such a ground is impossible. This means we must give up on the hope of being able to definitively decide once and for all what will be defining of myself. But this does not mean we need to give up on the goal of being more actively involved in our own agency. My proposal of perpetual openness is that we never identify with something so completely such that we would refuse to consider it in a new light. Identification, and the action on the self it embodies, no longer holds the key to autonomy.

to learn certain skills from parents.) In some sense, these animals are more "self-contained" than we are: they act according to internal "principles" which were not given to them by other creatures. Of course, as I have argued, they are nevertheless largely meta-passive precisely because their instincts are pre-determined for/given to them. A creature which gave itself its own principles would have to occupy the Archimedean point unavailable to humans. But the pure self-contained activity such a being would have is not necessarily a "higher" form of activity; it would be unchanging, and this static-ness is an aspect of passivity[110]. Thus, while the human and imperfect form of activity I am suggesting autonomy consists in necessarily involves elements of passivity[111], it also involves a form of activity which we would not have if we *were* able to be perfectly self-constituting: the ability to dramatically change our internal principles. Perpetual openness is essentially an open-ended engagement with oneself – engagement which recognizes and embraces the

[110] Two ideas posed by rather different philosophers seem relevant here – both, interestingly, about death. Derrida claimed (in paradigmatically esoteric style) that to know oneself is death. There are multiple possible interpretations, but the one I favor is that he meant that to know oneself completely would constitute the end of genuine engagement with the world. You would know exactly what you would do in each circumstance, and simply do it; life would become mostly automatic. This would put us in a state of mere existence as opposed to active living. Bernard Williams seems to be circling around a similar idea with his argument in "The Makropoulos Case". Here he argues that after a certain amount of time, a person would have had all the meaningful experiences possible for her given the substantial self she has, after which point life would become repetitious and unredeemably boring. This always seemed to me to less an argument against immortality and more an argument against having a static identity. Both Derrida and Williams seemed to be in tune with the dangers of a self which was too self-contained.

[111] in that [1] we start from a certain point, [2] we rely on outside feedback to help us become autonomous (more on this in a moment), and [3] our responsive revisions based on this feedback are thus necessarily partial.

possible need for radical re-definition. This open-ended form of activity should not be thought of as a consolation prize; it is in some ways a more genuine form of activity than pure spontaneity would be.

Of course, when we first take up the attitude of openness, we will still be largely meta-passive. Simply being open, while essential, is merely the first step. Autonomy will be a long and ongoing process of learning about (and re-shaping) the substantial self. Nevertheless, the mere adoption of the attitude of perpetual openness *itself* already constitutes *more active engagement/involvement* with your own agency. By accepting that you are almost certainly in the grips of certain influences and therefore meta-passive, you loosen the power that these things have over you. You are no longer completely beholden to them precisely *because* you stand at the ready to genuinely listen to disconcerting or confusing (perhaps even embarrassing) feedback about your current substantial self. This is what I mean when I say that by viewing your substantial self "provisionally", you take up a *meta-attitude* towards your whole (substantial) self. You no longer completely stand "on the side" of your current substantial self; instead, you are "on the side" of your potentially more active self. Although the attitude of perpetual openness will not by itself allow us to become significantly more autonomous, it is at the heart of philosophical autonomy, and it is intrinsically significant.[112]

[112] My proposal may seem vacuous. Am I suggesting that the way to govern the self is simply that we continuously try to govern the self? In a sense, this is indeed what I am suggesting. This might seem unhelpful. The very question we have been trying to answer is how we can govern the self, what the "standards for success" are. Simply saying that the standard is that we keep trying to meet the standard appears to sidestep the central question. But this appearance is misleading. As I have tried to explain, openness is a meta-attitude towards the whole self which is a form of activity *in itself*.

2.2: Details about Openness

I have argued that perpetual openness can replace the impossible Archimedean point. We can avoid being ineluctably controlled by external influences *not* by having access to a pure and neutral space untainted by them, but by being continually open to different viewpoints which allow us to learn about ourselves and the ways our agency is being determined for us. This requires that we alter both our notion of activity and our conception of governing the self. The goal becomes not defining my substantial self once and for all, but continually engaging with the ways I am exercising my agency. We will now look more closely at what this idea of perpetual openness involves.

There are two aspects to openness: what is *being* opened, and what is being opened up *to*. These aspects correspond to two points I made above: the attitude of openness (the first aspect) *in itself* constitutes a new kind of activity, and yet this attitude by itself is not enough – it needs feedback provided to it (the second aspect) from outside the self. I provide the first aspect, and depend on the world (including and especially other people) for the second. While openness itself already constitutes an important "levelling up" of activity insofar as it involves a genuine willingness to revise myself, I need substantive feedback about my self which suggests ways I can revise: feedback about the particular ways I have been meta-passive in my own agency and/or feedback which gives me greater self-understanding more generally. This is what I am being open *to*. The actual instigators of this learning experience can only be *perspectives outside of our current substantial selves*.

Let's look at this process more closely. The goal is dealing with the specific, substantive ways we are meta-passive. The first step towards this goal is simply to become aware of the specific ways we have been meta-passive. Even this initial step presents substantial progress, since part of our passivity is precisely that we are unaware of their influence: their co-opting of our agency is invisible. We make them visible when we interrupt their smooth functioning through us. This can happen in any number of ways: when we are presented with a perspective that our current ways of understanding cannot make sense of; when we are shown that our way of doing things is not universal or simply natural; or when we are presented with a situation where our values seem inadequate to guide us. Anything we are confronted with which we cannot make sense of given our current paradigms of beliefs, values, etc., throws a wrench in our current forms of agency and subjectivity, "breaks" it in a way that opens up space in which to question it.

For example, if one believes that there are only two genders, each one with different natural tendencies and abilities, and one lives in a society where everyone else believes this and behaves accordingly, then one will never question this and so this outsourcing of one's agency and subjectivity will never be brought to light and addressed. But if one is exposed to different cultures with different expectations about gender roles and identities – or if someone in one's own culture claims that this paradigm does not work for them – then all of a sudden one can no longer take this for granted. When we encounter something which causes a glitch in our system, a break is caused which makes visible the ways we have been exercising our agency (and more generally, shaping our subjectivity). They no longer

operate smoothly through us, their influence unseen and unchecked. This visibility gives us space from these things, a space in which to question them.

As the above example demonstrates, often times this break will come from outside of us. But it can also come from within us. If there are parts of our own experiences that do not make sense within what we have been taught, this also provides a break. In other words, a fragmented self can be a boon to governing the self. If the man raised in a culture of toxic masculinity described above experiences psychic pain because he cannot properly appreciate and express his more vulnerable emotions, he may come to see that this indicates something is wrong with this masculine ideal. If he is a "real man", he would not be experiencing such vulnerable emotions; the fact that he does experience them, if he accepts it, causes a break in his ability to see this ideal as unproblematically "natural" or "true". (Cases like this are where authenticity and philosophical autonomy coincide.)

To sum up: what we must be open to is any perspective which does not fit with our current way of being an agent and a subject. Often, if not usually, it is the perspectives of other people which open up a "gap" which allows us to question things we previously took for granted. (Even in cases where one has an "outside perspective" within oneself, one often needs the support of others to validate this perspective enough to take it seriously.) Philosophical autonomy on this view is interactive. This open-ended engagement with oneself (which is our revised notion of activity) essentially requires engagement with the

world: with other people and the various *social* worlds these people belong to, and potentially with the new ways of relating to the natural world that these others open up[113].

But this "breaking of one's perspective" is not enough to solve the problem of governing the self. Simply showing a person that their perspective has been given to them by the contingencies of their upbringing (for example) does automatically break the influence of this external factor. Often, the person will simply double-down on their perspective, claiming that others with different perspectives are wrong. This demonstrates once again the importance of the first aspect of openness: the *attitude* of being open. Things can only impact me in a way which causes me to *genuinely question* my self if I am first *sincerely open* to them. This is what perpetual openness is: it means that I am *committed to genuinely engaging* with perspectives different from my own. I must not simply tolerate or learn about different perspectives, but take them seriously such that I allow them to potentially inform and impact my own. Anything less than this commitment of sincere engagement means I am not actually open. This essential attitude of perpetual openness something only I can provide. But only the combination of the two aspects – my being open, and what I am open to – can allow me to become more philosophically autonomous. I need other people and their perspectives to present me with concrete possibilities of moving beyond the ways I have been meta-passive.

There has been a major question shadowing this discussion which I have yet to addressed. Being exposed to a different perspective simply introduces a *different* external

[113] Ways of relating to the natural world are socially learned and reinforced, so we can learn to see different kinds of value in the natural world by learning about other cultural perspectives.

influence. On what grounds would I choose to endorse one or the other in a way that would count as governing the self? The solution to this problem is found as we delve deeper into what is required for adopting this attitude of openness. Perpetual openness is the standpoint we must take up in order to (begin) governing the self. But why would we adopt such an attitude in the first place? The answer to this is the second core element of philosophical autonomy: to adopt the attitude of openness requires that one *take one's self seriously*.

Section 3: Taking One's Self Seriously

Before I can argue for why perpetual openness requires taking one's self seriously, I must first explain what I mean by "taking one's self seriously". I will start by articulating what the "*self"* is which we are taking seriously, and then turn to what it means to take this self *seriously*. By learning more about the self, we will also be able to draw a much more detailed picture of what precisely openness involves.

3.1: The Self

Let us start with an initial, relatively uncontroversial definition of selfhood. To have a self is to have a *rich kind of interiority* which we can assume only persons have (leaving open the possibility that persons need not be human[114]). (In what follows I will use the terms "person" and "self" interchangeably and use the terms "organism" or "creature" to refer to non-persons.) This interiority could be explained in terms of some combination of *agency* – the uniquely sophisticated ways humans act – and *subjectivity* – the uniquely sophisticated ways people are affectively enmeshed in the world. These concepts are

[114] The relationship between the concept "person" and the concept "self" is a complex one, which I will not be able to address here.

helpful and I will make use of them. However, I propose that we conceive of this interiority in terms of involvement in two relationships: a relationship to *oneself* and a relationship to the *world*. In many ways, these are two aspects of the same relationship since they mutually inform and depend on one another.

There's an apparent equivocation of terms I must first clarify. I am claiming that part of selfhood is being in a relationship to oneself. But if the self *is* this relation, what is the self I am relating *to*? The initial answer is the same sort of interior instincts and inclinations which all creatures of a certain sophistication have: desires, drives, pro-attitudes, a general orientation to the world, and so forth. It's not quite correct to call this a "self" in my sense of the term; we could perhaps call it a "proto-self". To have a relationship to myself is, in a primary instance, taking up a certain relationship to this proto-self: viewing my desires and drives in a certain way, prioritizing some and suppressing others[115]. A large part of socialization is teaching us how to relate to our proto-selves in this way[116]. However, after a certain point in normal development (at least in modern societies), it often *does* make sense to talk of my having a relationship to my *self* in the full sense of this term. In this case I am taking up a relationship to the way I am relating to myself: I am questioning or investigating which desires or values I am prioritizing, for instance. (This kind of multi-layered relationship is central to issues of self-government.)

[115] This is a Frankfurtian point. Frankfurt was right to emphasize how I relate to my own desires, but to make this relationship the whole of selfhood is too narrow.

[116] Of course, socialization is also about teaching us to managing our proto-selves in relation to the social world – how we present and conduct ourselves in front of others. In other words, managing the proto-self involves managing it in context of a social world.

I will use the phrase "relationship to oneself" in an open-ended way which can refer to either of the above senses.

One's relationship to oneself involves how we regulate and produce *actions* and how we interpret and regulate our *emotions and cares*: 1) a unique form of agency and 2) a rich form of subjectivity, respectively. Regarding 1: all organisms act on their instincts, but persons have the ability to distance themselves from such things and deliberate about what to do. This seems to be not just a different kind of activity, but activity in a greater form. To decide which desires we will act on means that we more actively participate in producing our actions than creatures do. Importantly, this means the agency distinctive of personhood both requires and reflects a depth which mere organisms do not have: it *requires* the depth of being able to hold yourself apart from immediate impulses and sense data, and it *reflects* the substantial beliefs and values behind the choice you ultimately make. (Persons need not always act in a way that demonstrates this distinctive agency *par excellence*. I can decide to eat ice cream simply because I want to, and I can decide to eat ice cream even if I value not consuming too much sugar [a weak-willed action].) The reasons, considerations, and motivations persons base their actions on involve beliefs, values, principles, and so forth. As we have discussed, these are first given to us by society, and they are typically bound up with how we interpret the world. (This is one way one's relationship to one's self draws on one's relationship to the world.)

My relationship to myself also includes how I *interpret and make sense of my internal world*, the *meanings and emotional import* these interpretations have, and how I *regulate my* internal life based on these. Again, the guidelines for how I do this are first

given to me by society. What resources for understanding my internal world – my thoughts, feelings, and cares – does society make available to me? What kinds of internal experience and attitudes does it values, and what does it disvalue? How does this structure my internal experience and attitudes (or cause me to try to change myself)? What are the standards I measure my worth against?

One's relationship to the world is, most broadly, the way one *interprets* it. This includes what we might call "scientific" or "theoretical" understanding aimed at truth, such as developing complex knowledge of causes and effects; conceiving of minds other than our own; developing vivid pictures of what these minds are thinking and feeling; and general conceptions of how these other minds and persons function. We therefore live in a much more expansive world than creatures do[117]. Aside from this theoretical kind of understanding, things also *mean* more to us. The world affects us more because of our emotional depth – an emotional depth which has a mirror in the world. We conceive of values and significances beyond the obvious ones of pain and pleasure; we laden objects, situations, and other creatures with elaborate connotations. We can summarize that one's relationship to the world encompasses what is taken to be *true*; the *meanings* one sees; what is taken to be *significant* (and thus what has *emotional import*); what a *just or right*

[117] Even people and societies which provide largely incorrect answers to these questions have still been driven to answer them. The incorrectness of any theoretical understanding of the world, or of other people, does not speak against the expansiveness of the world lived in.

ordering of the world looks like; and the *possibilities and reasons for action* one sees in a situation.[118]

These two relationships are not truly separable. Corresponding to our complex interpretations of the world around us, we have more complex desires, interests, and abilities. Our beliefs about the "correct" ordering of the world inform how we interpret and regulate ourselves. The complex meanings we see in the world structure our deliberations. Our actions reflect our understanding of the complex meanings of the world we live in; in turn, the meanings we see in the world are dependent on the projects we have set up for ourselves. We typically only get a solid sense of our individual identities insofar as those identities are supported and confirmed by concrete things in the world – status, material possessions, specific roles and relationships. In short, our relationship to ourselves and our relationship to the world mutually inform one another[119].

---

[118] It's worth noting that creatures also have a relationship to the world. This is in fact one way to understand what makes them creatures (as opposed to static objects), as well as what makes a creature the particular kind of creature it is. An elk has a very different way of relating to the world than a wolf. Different features of their environment stand out to them: they have different drives and desires; and to the extent that they have similar desires, different ways of fulfilling these desires tend to immediately present themselves. Like persons, the relationship creatures have to the world is bound up with the (proto)selves they have.

[119] It's worth noting that this understanding of selfhood makes room for a point many feminists have made: namely, that socialization is not an entirely negative thing. Although we have been suspicious of all the ways socialization can hinder autonomy, socialization is actually necessary for us to become selves at all. We must be guided in the development of our ability to relate to ourselves and the world in all the rich ways distinctive of persons. The depth of selfhood can only be grown in a human society. This means that we always and necessarily learn to be a self by first learning how to be a particular self.

Selfhood is the result of sophisticated cognitive capacities – and the cultivation of these capacities by being raised in human society[120] – which allow us to take up a rich relationship to the world (the ways we interpret and imbue it with meaning) and a rich relationship to ourselves (the ways we interpret our inner experience, and the ways regulate and exercise our agency and subjectivity). To be a self is to be *essentially involved* in these two relationships. Put differently: a self is a being which is enmeshed in a complex of meanings and significances which has the twofold direction of my self and the world. What I have referred to throughout this dissertation as one's "substantial self" can therefore be defined as the *particular ways* one relates to oneself and to the world – the particular meanings and interpretations one gives, lives and acts in. We learn to interpret the world and ourselves a certain way based on what we have been taught and based on the desires and cares we have been encouraged (or discouraged) to cultivate[121].

Thinking of the self in terms of a *relationship* is important for several reasons. Firstly, it highlights the ways in which we do *not* have control over these relationships. like any relationship, we have to "work with" the other party. There is a hard limit to the world *and* to the selves we have been given which we do not control and must simply accept. I

[120] Although we have been suspicious of all the ways socialization can hinder autonomy, socialization is actually necessary for us to become selves at all. We must be guided in the development of our ability to relate to ourselves and the world in all the rich ways distinctive of persons. The depth of selfhood can only be grown in a human society. This means that we always and necessarily learn to be a self by first learning how to be a *particular* self.

[121] It may be worth noting that our relationship to the world seems to come first. As an infant, one's initial desires and interests dispose one to relate to the world in a certain way, but as we are raised we learn to relate to these internal factors in certain ways.

cannot "interpret" my way out of gravity or of the particular body I have been given.(Once again, these limits mutually inform one another: if I do not have the use of my legs, this will impact the ways I live in the world.) But talking of a relationship also highlights the ways we *do* have control. Within the hard limits of our given selves and situations, there is much room for diverse interpretations and alternate meanings. To a large extent, we can choose what the "terms" of these relationships are. If I am born without the use of my legs, this could be interpreted as an evil deformity, a pitiable deformity, or as just one more variation of human genotype. Of course, one interpretation will likely be more supported by the society (i.e., the *world*) one lives in; once again we see how the two are intertwined. Thirdly, to speak of the self in terms of a relationship highlights that the ways one gives meaning to the world and exercises one agency is an *ongoing activity*. We *relate* to ourselves and to the world; it is something we *do,* and therefore something we can change.

Most importantly, speaking in terms of a relationship emphasizes two essential and reciprocal aspects of selfhood: *distance* and *involvement*. There can only be a *relationship* if there are two individual things involved; they must be differentiatable. The rich interiority of selfhood is predicated on this distance. I am distant from my basic instincts and desires (and later, can become distant from parts of my substantial self) such that I am no longer beholden to them. I can give them different meanings and values, and thus craft (and re-craft) a substantial self out of them. I am also distant from the natural world such that I am not enmeshed in immediate sense data. I can conceive of experiences and truths much broader than my immediate ones, and can imbue this world with meanings which go beyond those connected to pleasure and pain. These distances are what allows for agency

and subjectivity, interpretation and meaning. In this sense, to be a self is *already* to be involved in a significant kind of openness.

But although this distance is essential to selfhood, it does not imply true separation. So long as I have been raised in human society and have therefore learned to be a self, I *must* relate to the world and myself in some way. And since I must relate to the world and myself in some way, this means I must have *some "*substantial" self. This does not mean my substantial self is unchanging across time (this would be diametrical to my whole project); rather, in each moment of action or commitment I must take up *some* way of relating. The point of making these relationships the very definition of selfhood is to emphasize both the distance, and its attendant open-endedness, *and* the inseparability, and the attendant necessity to make some sort of stand.

Of course, the motivation which drove us to philosophical autonomy was precisely that this particular self is largely *given* to us. Now that we have a general understanding of the self, we can return to our main argument about perpetual openness and taking one's self seriously.

3.2: Taking the Self Seriously

I have argued that perpetual openness is the attitude we must take up in order to govern the self, and indicated that taking the self seriously is necessarily a part of this attitude. Now that we have a picture of selfhood, we can turn to articulating what it means to take this self *seriously*.

The phrase "taking one's self seriously" could be understood to mean "taking the project of living up to the substantial self I currently have" seriously – that is, taking the

particular interests, desires, values, and perspectives I have on the world and myself seriously. Insofar as I would be taking my substantial self *seriously*, this would be a project of building greater integrity: making sure my internal and external life is aligned with my values and interests. There is some room for how to interpret "substantial self" here: it could mean the self which was largely given to me by socialization, or it could mean my deeper, "authentic self" (whatever that might mean). In the case of the latter, it would be an instance of governing from the self. But in either case, it will not work for governing the self.

We must return to the definition of selfhood I just proposed. To have a self is to be necessarily involved in complex relationships to the world and to ourselves. The essential trait of selfhood is this twofold fact: I need not have any *particular* relation, but I must have *some*. To take my self seriously requires focusing primarily on this twofold fact. I will therefore use the phrases "taking oneself seriously" and "taking one's self seriously" to denote separate attitudes. To take *oneself* seriously means that you are identifying with your current substantial self, and you are taking this substantial self seriously. In contrast, to take *one's self* seriously is to stand somewhat apart from your current substantial self, and to take your selfhood as such seriously. The trick to getting this right is balancing the two aspects of the definition.

Let's start with the first aspect: to take my self seriously requires that I *not* take the *particular* relations which are defining of my current substantial self so seriously that I neglect the fact that selfhood is not to be defined by these particular relations. (To give an example, a conventional suburban mom might believe that all people need to have children

in order to be genuinely fulfilled. This assumes that this particular kind of meaningful project is essential to personhood as such.) It thus requires recognizing that the substantial self I currently have – the specific ways I am taking up these relations – are not necessary aspects of selfhood. There are other ways of being a self. The main thing this entails is recognizing that I have *been* actively relating to the world – in other words, that *I* have been setting the terms of these relationships. The world I've been living in and the self I have are not pre-given essences which I have been neutrally observing and responding to – they are given their particular character because I have *interpreted* them a certain way[122]. Recognizing that my current substantial self is not essential to selfhood does not automatically mean that I have "overcome" this self. However, just by recognizing that I have been actively relating to the world and myself in a certain way, I have already broken a large part of the external influence's control: the *invisibility* which allowed it to exert its influence so seemlessly. Precisely because I will no longer see my original views on the world and on persons as "natural", or "just the way things are", I am no longer simply beholden to these ways of seeing. I have broken my total immersion in them. This realization will not immediately allow me to see *all* the particular ways the world and myself has been colored by; nevertheless, by recognizing the general fact, I will be more

[122] I could introduce here a concept which will become very important later – taking responsibility for one's self. To recognize that one has been living in a certain way and being a certain kind of person because one has been relating to one's self and the world in particular ways is to see oneself as responsible for these relations – not in the sense that one originally chose them, but in the sense that one has been continuing to give them power. This of course means that one can chose to alter these relations.

open to recognizing these particularities when they are brought to my attention. I will be in a better position to give due weight to the first part of the definition of selfhood.

Grasping that the form my current self takes is not essential to being a self as such (the first aspect) is the first step towards taking my self seriously. It is also the first step towards perpetual openness. Indeed, most of what I have said so far is a way of redescribing the preliminary attitude of openness. But it has not yet gotten us to the *robust* attitude of openness – the kind of openness which compels us to take other viewpoints seriously.

This is where the second aspect of selfhood becomes important. While no particular relation is essential to selfhood, I nevertheless must have *some*. Since we need to relate to the world/ourselves in *some* way, this means that we must have some *particular* relationship to them. In other words, the attitude of openness cannot simply be refusing to take on a particular form. If I do not relate to the world and myself in some way, I will not be able to act. If I do not relate to the world and myself in some way, I will dissolve the rich interiority which makes me a self to begin with. In short, I cannot refuse all particular content. But if I must have some content, why can't I just accept the content I already have, and be happy with the substantial self I have been given? It seems possible for me to recognize that I needn't have had the self I do, that there are other legitimate ways of being a self, and yet double down on this substantial self. (After all, I will already be biased in favor of it.) Put most precisely: given that I must have some content (the second aspect), what relevance does recognizing that no content is essential to selfhood as such (the first aspect) actually make to how I take up my selfhood?

The problem is making sure that I give due weight to both aspects of selfhood simultaneously. If I focus too much on the non-essentialness any particular content, I might shrink away from committing to any content. If I focus too much on the necessity of some content, there seems to be no reason to change my current substantial self. But if I focus on both of these aspects *equally*, then I will be forced to confront the question of what content I will give myself: I must decide who I will be. Insofar as I simply accept who I am, I will be outsourcing terms of my relationships (to the world and to myself). To take my self seriously is to take ownership and responsibility of this dual relationship, to be committed to being more actively engaged in them. This means I actively take up the question "who will I be?" and commit myself to finding a "good" answer[123]. *To take my self seriously* is to take *this question* – "Who shall I be?" – seriously, such that I am committed to trying to find the best answer. It is *not* taking a particular answer – that is, a particular, substantial self – seriously, *except* insofar as it is the best answer I have yet found[124].

---

[123] Here's a different, perhaps more intuitive way to describe this: To take my self seriously means that I think it matters how I fill myself in. And if I think it really matters, I will believe that simply giving myself content on a whim, or based on the contingencies of how I happen to have been raised, or on the predispositions I happen to have been born with – none of these things can serve as a sufficient ground for my choice.

[124] With this definition in hand, we can understand the previous argument in a new light. To focus on the fact that no content is essential to you and to thereby hold off from claiming any content as part of you is to emphasize the first aspect of selfhood – that it is not tied to any particular content – at the expense of forgetting the first. This is not yet to take the question seriously, because you are not yet seeing it as a question you must answer. To see that you must have content, and so the content you already have might as well do, is to emphasize the second aspect at the expense of the first. This is also not yet to take the question seriously, because you think that it does not matter what answer you give. I must give both aspects of my selfhood equal weight. Only once I have done this

One might wonder if there is another way to give proper weight to the non-essentialness of any particular content: couldn't I instead adopt the perspective that it doesn't really matter what concrete form I give myself? In this case, I would be taking the question "Who shall I be?" less seriously in that I'd give less weight to any answer I gave. There seem to be two main ways this could manifest: one could live in the whims of the moment, *qua* Frankfurtian wanton; or one could adopt a set of longer-term projects/a longer-term identity while nonetheless seeing these projects as ultimately arbitrary. Both options would preclude any kind of genuine commitment. While the second involves longer-term projects, insofar as one was taking the non-essentialness of any particular content seriously, any commitment involved would be shadowed by a sense of irony ("I do not have to do this, but I guess I will"). But this would be at odds with taking my self seriously. Indeed, it would be akin on giving up on the possibility of taking myself seriously. In fact, this actually goes against taking the second aspect (I must have some content) seriously, since it would entail that I would have to hold myself at a distance from any content I gave myself. (There is a seeming tension here with my previous emphasis that we must see any answer we give as "provisional", a tension which allows me to clarify what I meant by this statement. Seeing my current answer as "provisional" does not mean that I give it lightly or without consideration. It does not mean I am not committed to what I believe is right in my answer. It simply means that I am humble enough to see that my current answer is mistaken or incomplete.) Holding oneself at a distance from the content

---

will I have seen the question of my self as worth answering, as needing an answer which is, in some yet-to-be-determined sense, a good answer.

one adopts, as one does when one thinks it doesn't really matter what content one has, isn't compatible with taking oneself seriously. It is a way to avoid the task of taking oneself seriously.

But perhaps, one *could* see the adoption of one's projects as totally arbitrary, and yet still be seriously committed to them. A nihilist or Nietzschean might find meaning in the very strength of spirit required to affirm something even while believing that it did not matter whether they affirmed it or not. I have two responses. 1) We must remember that we are concerned with finding out what is involved in *governing the self* – and this attitude of finding meaning in self-assertion is a dead end in this respect. At most, self-assertion would be a project of authenticity: discovering what I find most meaningful and then devoting myself entirely to this. The ground for what I did would be based on what I happened to find meaningful, the proclivities and psychological dispositions I happen to have. I may transform these by giving them an extra layer of meaning by seeing them as uniquely mine; but nonetheless who I am is largely given to me. I simply double down on it. 2) This Nietzschean response doesn't take seriously enough the first aspect of personhood: that no content is essential to it. It takes my selfhood as defined by certain content, and takes pride in doubling down on this content.

To take myself seriously, therefore, is to think that it matters what content I give myself, such that I can do a better or worse job at this task. The obvious question now becomes, what decides the standard of a "good" or the "best" answer? We will see a more complete answer in a moment, but for now we can say at least this: since selfhood is essentially open-ended (the first aspect), this means *we cannot give any answer which*

contradicts this aspect of self-hood. Answers which claim that you simply are a certain kind of person and that's all there is to it would be ruled out. This includes answers like "I am X (ethnicity, religion, nationality etc.), and therefore I *just am* (or I *just have*) Y (ability, values, interest, etc.)". (The "just" is doing most of the work here: as we will discuss in Chapter 5, there are ways to emphasize the importance of the contingent identities one has; the issue is whether these properties are essentialized such that they definitively close off possibilities.) Put another way: it's essential to remember that the primary form of autonomy is taking the question of my self seriously, and so any answer that would be inconsistent with taking the question seriously is precluded.

One surprising corollary of this is that answers which imply a denial of this aspect of selfhood (that is, their ability to reflect on and remake themselves) to others are *also* off-limits. This comes simply from the fact that if I properly understand what is essential to my own selfhood (the two aspects), I will understand that it doesn't attach to any of my contingent qualities. If I claim that someone of a certain race, gender, or so on simply *is* a certain way, I am making of them a static creature; I am denying them the second aspect of personhood (the fact that they are *not* essentially bound to any particular content). But by doing so, I am holding my *own* selfhood hostage to *my* particular race or gender (and so on). I am saying that the fact that I have the ability to complexly relate to the world and myself, and can reflect on and remake myself, is because what is *most* essential about me is a brute fact I cannot change. The purported foundation of my sense of self and dignity

contradicts this very selfhood[125]. Therefore, a proper understanding of my selfhood entails a proper understanding of selfhood in others.

If I take seriously the question of who I am such that I want to give a good answer to this question, and I am aware that how I currently understand myself and the world are just one way of taking up these relationships, then this means I will be aware that I almost certainly have a limited view of things. In short, I likely have some content which is "incorrect". This means that I must be open not only to changing my initial answer (the answer which was provided for me via socialization), but open to other perspectives which show me how I might fruitfully change my current perspective. In short, taking myself seriously necessarily leads me to the attitude of *perpetual openness*[126].

Interestingly, taking one's self seriously effectively means taking up the project of philosophical autonomy – that is, seeing the problem which the concept of philosophical autonomy captures as a genuine problem which I *personally* need to address. One need not conceive of this problem in the precise terms I have used. One simply needs to a) feel the tension of the dual fact that my selfhood (my existence) has been shaped and limited by things which are purely contingent to me and which therefore are not essential to my

[125] A parallel point: when one believes that one's ability to be autonomous and self-directing is guaranteed by the kind of creature one is, one will be less likely to take seriously the ways in which one is not already autonomous and self-determining. In short, one will not have the requisite humility for perpetual openness.

[126] In technical terms, taking my self seriously is both *necessary and sufficient* for perpetual openness. This is because taking my self seriously is constitutive of perpetual openness. In turn, perpetual openness properly and fully understood means taking my self seriously. As I will discuss in the section 4, this need not mean that I primarily or even overtly think of this project in terms of my self.

selfhood, and yet I cannot escape having some particular form; and to b) feel the urgency of addressing this tension (answering "the question of my self") in some adequate way.

It is this standpoint of taking one's self seriously that gets us to perpetual openness. If we did not see that the contingencies of how we have been taught to have a self are not actually essential to our selfhood; that we nonetheless must have some concrete form; and that it is therefore significant *what* concrete form we give ourselves, we would not be perpetually open to new perspectives and useful feedback in the ways I described above. We might tolerate different perspectives, but we would not be motivated to take them up and allow them to impact us. The prerequisite of perpetual openness is believing that it matters how I answer the question of my self: that it matters who I am.

But this attitude requires something further: something which also fills a significant remaining lacuna in my view. I have argued that one must be motivated to fill out the details of one's self in way that is somehow "good" or "correct". But what sets the standard of "good"/"correct"? These concerns lead me to the third and final core element of my view.

Section 4: Taking the World Seriously

To believe that who I am matters such that I want to provide a good answer to the question "who am I?", I must believe in a standard outside of me. More specifically, it requires believing in an *objective* value or standard of correctness. By "objective", I simply mean that you cannot be a relativist through and through: you cannot think "anything goes". To think that *anything* goes would mean that it would not matter what content I gave to myself, which would preclude taking the my self seriously. But we must be careful. When I claim that taking oneself seriously requires believing in some sort of objective value or

standard of correctness, I do *not* mean that one needs to already have a particular standard in mind – quite the opposite, in fact. It requires believing that there is *some* standard of correctness which we must continually strive to understand and live by. It requires accepting the fact that we are finite and largely meta-passive, and so we can never take for granted that we simply *know* this truth and have nothing more to learn.

But this means that I must *also* believe *other* things matter. Believing in an objective standard just means believing that certain things matter independently of my perspective on them. But this requires believing that things outside of me matter[127] *in themselves* such that I can relate to them in better and worse ways. This is essential for the possibility of recognizing that I could get things wrong. I am beholden to things outside of me such that I do not get to arbitrarily decide on my own the best way to relate to them.

Believing that things matter independently of me is therefore necessary for the possibility of governing the self. Without this belief, there would be no ground for an objective standard I should respond to, and so I would have no basis on which to decide what to make of myself – no basis other than the contingencies of innate traits and/or the society and culture I happen to have been born and socialized into. There would be no way and no reason to reach beyond myself in the way governing the self requires. There would only be room for governing *from* the self.

[127] In fact, this dependency was already implicated in the definition of the self we are working with. This definition holds that the self is essentially a relation to the world and to oneself. Therefore, to believe I matter just means that I believe how I relate to the world matters; and this requires believing that the world *itself* matters such that there are better and worse ways to relate to it. This includes believing that other people matter.

I have emphasized that we must believe there is an objective standard of correctness. But we must always remember the corollary point that we cannot assume we already know the complete objective standard to live by. Rather, we must believe that one exists and that we are compelled to discover it. This is, of course, simply what it means to look for truth – to refuse dogmatism such that you are always open to learning more, and to adjusting your perspectives. To look for truth is to accept that there is a standard which exists outside of you (though not necessarily independent of people/this world) which you might not have fully grasped yet, and which yet sets the standards for your beliefs. Similarly, the kind of belief which is necessary for taking oneself seriously is the conviction that there is standard of correctness which we must strive for, even if we do not know it yet. Indeed, as I have been emphasizing, this belief that we are almost certainly missing parts of it is *central* to this quest. I am therefore committed to uncovering this standard. It may be useful to draw an analogy here – there is clearly a difference between people who are interested in being right, and people who are interested in the truth. Those who are simply interested in being right are often closed minded, willing to stubbornly defend their positions, because they cannot stand to be wrong. But those who are interested in the truth will be enthusiastically open to new data and perspectives, because their goal is not essentially to be right, but to know the truth. This combination of intellectual humility with belief in the worth of knowing the truth is precisely what is needed for taking the self seriously. We are not completely lost, however: our guide is the simple axiom that people (and most likely other creatures and even some inanimate objects) matter.

Here's a parallel argument. If I did not believe the world and other people mattered, I would not believe how I *impacted* them mattered, and so what I *did* would not matter; and if what I did didn't matter, I would not believe *who I am* matters, and so I would not be inclined to question who I am beyond whether or not who I am makes me happy/satisfied. I would not be inclined to take up the radical position of openness which is required for philosophical autonomy. Again, philosophical autonomy is predicated on my believing that things matter independently of me, that I do not simply get to decide what their value is. (Of course, I must also *care* about these things that matter, or else I will not let them impact what I make of my self.) Therefore, to take myself seriously is equivalent to taking the world and other people seriously such that I am motivated to "do right" by them – and to discovering what precisely "doing right by them" means and involves.

What do I mean by "doing right by" the world and other people? What kind of "objective standard of correctness" does this involve? Because selves are essentially active beings, and because what we are concerned about here is governing how we exercise this activity, we need a specifically *practical* standard. The standard we are looking for must therefore be, at least in part, an ethical standard: a standard that tells us how to treat (and not treat) other people *and* other things in the world more generally. Since (some) things in the world, and people specifically, matter in themselves, this means that there is a standard for how I should treat them. But since selves are also essentially meaning-embedded beings, this standard applies not only to how I treat people and things in the world, but also how I understand them and the significance I give them.

One might ask, "Why must these standards of correctness be ethical ones? Why must the "better" or "correct" way to relate to other people have anything to do with ethics?" The answer is that the question at the heart of philosophical autonomy – "Who should I be?" – requires this kind of standard. If the standard was not ethical, it could not provide us with any guidance for how to act, how to manage ourselves, what ends/values to see as worth pursuing, and so forth. This is a classic philosophical point: if there is only theoretical/scientific truth, then it seems the ends we chose are up to us, and so the "truth" can only give us instrumental guidance. If this were the case, I would be right back at governing *from* the self: I would have to simply accept whatever ends I happen to have as ultimately setting my practical principles (probably with some modification to smooth over internal inconsistencies). To govern the self, to be perpetually open to what I do not already see, requires that I believe not just in a standard for theoretical truth, but a standard for ethical truth.

Similarly, one might object that being concerned about how I relate to others need not commit me to being concerned about how I impact them; i.e., that this need not involve any *ethical* concerns about others. But this misunderstands that selfhood is not just about how I interpret, understand, or assign meaning to myself and the world; it's also about how I act based on this "knowledge". Thus, how I relate to others, as part of my selfhood, is also about how I treat them. More to the point, this objection misunderstands that the kind of "interpretation" I am talking about here isn't just theoretical, but involves my whole subjectivity: what I care about, what my emotional self looks like, my values, and so on. Being a self is being in relation to the world and other people in this deep way. To care

about my *self* therefore means that I care about how I am *emotionally and motivationally bound up with others*.

The forgoing discussion allows me to clarify a bit more what I mean when I say things *matter* (specifically, that they matter *independently* of my perspective on them). Things carry "in themselves" certain requirements for how I must treat them if I properly understand their significance and worth. For example: if I come to see that a mountain top has value (aesthetic, ecosystemic, perhaps cultural), then I do not get to say that since in *my* perspective it has none I can blast it apart to access coal underneath. If I believe the mountain top has value independently of my perspective, then I am compelled to treat it in certain ways. (At the least, I must give it a certain weight when deliberating about potential actions.) A simple way to express this idea: if things matter, I must adjust my interpretations and actions in response, as opposed to adjusting things so they suit my interpretations and actions.

Again, I cannot assume that the ethical standards I already believe in are the correct ones – I must be open to revising these standards. The point is that I must believe there *is* an objective ethical standard, and I must be committed to uncovering and living by it.

A clarification: I mean "ethical" in the sense that is broader than "morality": that is, in the sense that is not solely concerned with a set of duties which dictate actions we must or must not take. This is *not* to say that there will not be moral standards we must strive to understand and meet. Since there are better and worse ways of relating to people, including how I treat them, this means that standards we might more properly call "moral" will also appear. (Many of these will be obvious: e.g., if people matter, I probably shouldn't

murder them just for fun.) Nor am I saying that these moral standards are secondary to the ethical standards we uncover. But since we are concerned with governing the self, where this is concerned with the content of the self, this goes beyond the moral in the narrower sense of duties. It has to do with who I am in all the deep senses of this – the meanings I give the world and myself, what I value (or at least appreciate as valuable), what I pursue, etc.

So far I have been speaking mainly in terms of how I must relate to other people. But I have also said that my relationship to *the world in general* will have standards of "correctness". It involves an attitude I will call (vaguely but hopefully evocatively) "honoring the world". Honoring the world means acknowledging, and then acting in a way that does justice to, all the nuances and complex meanings of a situation. Again, the idea here is that in order to believe there are better and worse ways for me to relate to the world, I must believe I am beholden to it in a certain way. Once again, this "relation" involves not just how I treat the world, but how I might understand, appreciate and value it. (This is because my selfhood involves not just agency but subjectivity.) This is what I hope to get at with the term "honoring".

Other people are essential to learning about how I can honor the world. People – especially those with very different perspectives from me – can provide valuable insights into aspects of the world which I have previously dismissed or devalued. Often (perhaps usually) honoring the world involves engaging with the meanings that other people have either *seen* or *put into* it. To explain why this is, we should remember that we do not "access" the world from a neutral standpoint, but from a particular standpoint. In other

words, the world we relate to is already interpreted in a certain way. Now, I do not think we need to go so far as to say that people *constitute* the world such that there is no external reality independent of us (as indicated above, there seem to be *some* "hard limits"). Nevertheless, there are many ways of interpreting the world; ways of emphasizing different aspects of the world and imbuing these with meaning and significance. Many of these meaning we will want to reject on ethical grounds, but many of them are simply different ways to relate to the world. These different ways of relating will often -though not necessarily always! – reveal new forms of value and significance which we did not see before. For example, many indigenous cultures emphasize the importance of deep involvement with a specific place and environment which is often missing in modern western societies. Engaging with these cultures may thus teach us the importance of paying close attention to the specifics of our physical surroundings.

Perpetual openness means that we are open to these different ways of relating to the world and living a meaningful life. Since we are raised in particular societies, we have likely been taught to only see one set of meanings. Being open to different forms of meaning therefore allows us to live in a world of open-ended richness, to add layers of meanings to the ones we are already familiar with. There are multiple ways in which we can be open to new meanings: based on different cultures (or subcultures); based on kinds of value I might not be predisposed to engage with (e.g., if I am predisposed to like art, I might see sports as a waste of time); based on individual experiences which differ from mine (e.g., family upbringing); different sociological or historical interpretations of a situation (e.g., systemic/structural racism is real).

To bring this back to "honoring the world": the idea is that since I have only a partial view of the world and all the values/meanings it has to offer, there are many forms of these I am missing. I will not therefore act in a way that respects all these forms of meaning. The resources I currently have within myself (in terms of dispositions, values, desires, etc.) will not allow me to fully honor the world, and so I must grow my perspective and understanding to accommodate it[128]. Once I do, I will simultaneously open up unforeseen possibilities for myself, unforeseen forms of human good. Failing to honor the world means refusing to see meaning beyond where I originally see it[129].

My point that we should always be open and responsive to different perspectives gives rise to an obvious worry: what about perspectives which are obviously wrong, morally or otherwise? Do we need to be receptive to the perspective of slaveholders, for instance? This brings us back to the belief that other people matter, which I have argued is necessary for the project of philosophical autonomy. Perspectives which contradict this basic belief upon which the whole enterprise of philosophical autonomy rests are automatically disqualified. To briefly reiterate: the core attitude of autonomy is taking the question of myself seriously, and this precludes any answers which impede taking this question seriously. Positions which deny selfhood (and its attendant concepts such as

[128] Will I be able to accommodate *all* worthwhile meanings? Certainly, I will not be able to enact/directly engage with *all* possible meanings in my own life. These are questions I will address in Chapter 5.

[129] And since the world is connected to people, failing to honor the world also involves failing to honor other people. This once again points us to the fact that respecting people in the way necessary for philosophical autonomy requires more than just tolerating differences; it involves genuinely engaging with these differences such that I allow them to impact me and potentially change my perspective.

dignity and autonomy) to another person require me to misunderstand my own selfhood in a way that hinders taking the question of my self seriously. I must accord to other people the same kind of selfhood that I have: they actively relate to the world and themselves in certain ways, but need not relate to it in the way they do – they have the ability to revise themselves. We can therefore reject many obviously morally wrong perspectives, and do not need to be open to them.

With this said, there *is* still some value to engaging with these unacceptable perspectives. The goal of honoring reality is to build up a better and better understanding of the world, people, and possible meaning. In other words, we don't just want to be ethical, we want our ethicalness to be grounded in genuine understanding. To engage with how these perspectives see the world and other people can be instructive to developing this ethical understanding. More specifically, *what* they get wrong and *how* they get it wrong can be instructive. Our rejection should be based on this understanding. In fact, grounded rejection is a *form* of responsiveness. In this way, our very rejection is a way of honoring the world.

The fact that taking myself seriously requires taking the world seriously leads to what I take to be an advantage of my account: it does not require a primordial preoccupation with myself. This might seem contradictory; let me explain. Taking myself seriously does not require that I am so concerned with myself such that I am only incidentally concerned with the world. I must be concerned with and attentive to the world for its own sake. This is part and parcel with believing that things matter independently of me, that there is an objective ethical standard I must strive for. This means that someone can count as "taking

one's self seriously" without being mainly focused on the kind of self one has. One could be primarily concerned with the world, with making sure that one is doing all one can to live ethically and honor reality, taking new perspectives and feedback seriously and adjusting one's perspectives and actions accordingly, and by doing so one would perfectly embody the attitude of taking one's self seriously. By engaging with the world and monitoring and modifying oneself, one is thereby also taking the question "who shall I be?" seriously, even if this is not one's primary concern. In other words, my view is decidedly *not* a navel-gazing one.

To be clear, none of this is a direct argument for why we should have an ethical consciousness. My point is the much humbler one that *if* one is concerned with being philosophically autonomous, then one must *also* be committed to having (and cultivating) an ethical consciousness. While I do believe that all of us *should* be concerned with being philosophically autonomous, arguments to this end are beyond the scope of this dissertation. But insofar as one does care about governing one self, one must also to take seriously one's involvement with the world.

Conclusion: The Dynamic Self

Let us retrace our steps. I argued that the concept of governing the self cannot be saved if it requires an Archimedean point. Since we cannot rely on anything solid, then the only way to escape the dilemma is to actively take up the attitude that nothing is solid. We then turned to the idea of perpetual openness. By being perpetually open and responsive to new feedback, we avoid being simply stuck in the ways of exercising our agency which we have been given. But openness requires taking one's self seriously: we will only actively

take up the perspective of being genuinely open to new perspectives if we are able to achieve some distance from the particular self we currently have *and* at the same time take responsibility for having *some* particular self. If we did not believe that who we are matters, we would not believe that it was significant that we get it "right" somehow; and if we did not take our selves seriously in this way, we will not be eagerly receptive to new perspectives in the way perpetual openness requires. But this is turn requires believing that the larger world, including other people, matters. Otherwise, there would be no need to take the question "who will I be?" seriously. One answer would be just as good as another, and I would not need to be perpetually open. Perpetual openness is therefore predicated on taking myself *and* the world seriously such that I hold myself responsible to objective standards. In turn, taking the world seriously naturally leads me to taking my self seriously and to being perpetually open. These three attitudes are mutually implicated and bound up in one another. As I claimed at the beginning of this chapter, they are really three different faces of the same core attitude.

We can now address a question raised at the end of section 2: how does being open to different perspectives lead to governing the self, since this simply seems to add one more option (the new perspective and the values/desires/etc it makes available) to choose from? How could we decide between this new perspective and the old one? The answer is that we choose the perspective (which will be a combination of insights from numerous perspectives) which is most aligned with honoring the world and other people. We do the best we can given our current understanding, and are committed to improving this understanding as necessary.

It is essential to remember that the standard of "correctness", which will be in large part an ethical standard, is not brought in arbitrarily from outside the concerns of perpetual openness. Rather, it is *constitutive* of openness itself. Governing my self *just means* that I actively *take up* my relationship to the world and *respond* to it in the ways required to honor the values and meanings it contains. (Presumably, it will be impossible for me to personally engage with all the meanings the world has; this is the main topic of chapter 5.) Since there is no Archimedean point, being brought outside my current substantial self is the only way to move beyond my natural and socialized limitations – i.e., the only way to move beyond passivity within the self. My suggestion of openness is essentially that what we need is not to go *deeper* into the self, but to expand *beyond* the self. And the only way to move beyond myself is to be *first and foremost* on "the side of" the world instead of on the side of my current substantial self. This is what the attitude of perpetual openness comes down to. Thus, taking the world seriously and being responsive to it is built in from the very beginning.

It might seem like we have strayed from the key concept of self-governance – how can this count as "giving myself the law" if I am basing my choice on the world[130]? As discussed in section 2 above, this does indeed involve an adjustment of how we understand "governing the self" and of its corollaries, "self" and "activity". To remind the reader, I originally defined governing the self (i.e., being fully active with regards to myself) as

[130] Explained another way: I originally introduced the notion of openness as a solution to a negative problem, viz., how can we escape being completely beholden to external influences? But an answer to this question does not yet seem to be a full, positive account of self-governance.

deciding what my substantial self will be in the first place. We could perhaps envision this as a chain of command – I decide not just the particular actions I take, but the reasons, values, etc. behind these actions, such that I am in charge at every step or level of the genesis of the action. The kind of "activity" here is thus one that commands from some centralized source and controls all radiating layers of the self; the corresponding notion of "self" is thus a self-contained consciousness at some internal seat of power. The notion of self-governance involved is that I set the terms on which I will interact with the world: by deciding what desires, interests, perspectives will be essential to me, I decide how I will interpret, respond to, and act in the world. By making sure that I and I alone set the terms of my relationship with the world, I thus maintain my self-containment.

The new notion of governing the self I am proposing is different. I want to say it is not only *necessary* given the impossibility of self-containment but is in fact a *better* option: that is, it captures forms of activity and value which the original picture leaves out. By definition, the more different a new perspective is from my current one, the less I will have a precedent for making sense of it or being able to respond adequately to it. It thus forces me to grow, to stretch my understanding (and often my compassion) in unprecedented ways. The only way I can do this is by being *sensitive and responsive* to what is outside my current understanding. This is thus a moment of peculiar vulnerability. But in being thus attentive to the particulars of the situation and the new demands it places on my ability to understand and respond to it adequately – by being focused on the world and *not* myself – it leads me to create new resources in myself and to use them in ways not pre-given to me. This kind of activity is decidedly not self-contained in the way of the above picture.

But it is a form of governing the self, precisely because it involves me grappling with my self in a way that does not simply reduce down to governing from the self I currently have.

There might still be some doubt as to whether this counts as self-governance. The answer is that we must properly understand the weight of taking one's self seriously, i.e., of taking up the attitude of perpetual openness. Perpetual openness is a meta-attitude, as is taking one's self seriously: they force me to bring under consideration the particular ways I have been relating to the world and to myself. To take on this meta-attitude is in an obvious sense to step outside myself – that is, outside my substantial self. Of course, I can never completely step outside myself. This means that I am "identified" (if I may be allowed to appropriate this word) with this meta-attitude such that this is the standard I am holding myself to, the standard which tells me the defining project of my identity. Therefore, to be perpetually open means that I am indeed governing according to a standard which I am giving myself; I am self-governing. Although I am "taking my cues" from the world – and indeed the commitment is precisely to thus take my cues from the world – the standard is the standard of my self.

Philosophical autonomy is perpetual openness – and this means an *ongoing commitment* to be receptive to feedback. Governing the self, therefore, becomes primarily about grappling with the self by engaging with the world. It means being willing to learn about one's current substantial self from multiple sources; being sensitive to the nuances of new perspectives and what they have to teach you about how you have been understanding the world and yourself; being willing to ask yourself "what forms of meaning, value, ethical truth have I been missing?"; and *responding* to this feedback.

Governing the self is unavoidably bound up with sensitivity, engagement, and responsiveness. The point is not to decide on a substantial self once and for all, but to become more actively involved in your own selfhood, which requires you to be continually engaged in the process of learning from and about other people, the world, and yourself. The world and other people are my partners in governing my self.

Despite the revision, this account of governing the self still retains an element of the original definition: by continually grappling with my self, I am continually redefining myself. Indeed, the re*definition* is essential. As discussed in sections 3 and 4, I must take *some* stand (relate to/understand/act in the world in *some* way). This is part of what it means to take the world and myself seriously. In other words, I must take what I learn in moments of growth and incorporate them into my larger, longer-term perspectives: I must incorporate them into a (semi)stable self. Where my account differs is that it insists that this is an open-ended, dialogical process. This means that "perpetual openness" is perhaps not the most accurate way to describe the attitude as a whole: it leaves out the longer term concrete changes. What is involved is continuously *monitoring* and *managing* your self – that is, the ways you are relating to the word and to yourself. By continuously inhabiting this position of overseeing and adjusting your self, you are governing your self in the best way that an unself-contained creature can.

I will call this the *dynamic self* view. It is dynamic for multiple reasons: 1) my autonomy requires me to engage with other people/the world, i.e., it is not something I can achieve on my own; 2) it requires me to always be willing to revise myself and thus 3) it means my substantial self is a continuous work in progress; and finally (and at the deepest

ontological level) it holds that the "Deep Self" which philosophical autonomy wants us to secure is fundamentally defined by this very sort of activity: self-re-definition that happens in conversation with the larger world and with others. This activity is the kind which allows us to do the most justice to ourselves, in the sense that we are committed to not outsourcing our selves to anything or anyone else. (To be committed to not outsourcing one's self is another way to describe taking one's self seriously.) The dynamic self can never "be" autonomous, or have autonomy as a property; it can only practice autonomy.

I have argued that these three attitudes, and the dynamic self they are centered around, are key to an account of governing the self. But this is not yet to describe how these attitudes actually work in practice. I have also not addressed how my view is compatible with the other core critique of governing the self. Recall that there were two critiques: one theoretical, and one practical. I have effectively tried to address the theoretical one, but the practical one remains about how contingent aspects of the self actually seem necessary to agency remains. In the next chapter I will take up these questions.

Chapter 5: Taking Responsibility for One's Self

Introduction

In Chapter 1, I first introduced the core concern which motivates the need for an account of governing the self: I want to have control over not just what I *do*, but who I *am*. If I truly desire to govern my self, I will not be satisfied with simply accepting that I was simply born or socialized a certain way. I will not want to outsource the deciding factors behind how I exercise my agency (my values, principles for actions, worldviews) to things beyond my control; I will want to avoid being *meta-passive* in this way. In Chapters 2 and 3, we discovered just how difficult coming by the necessary control over our selves is. In Chapter 4, I proposed an account meant to address these difficulties. By being perpetually open to learning about and from other perspectives, I can continuously learn about and strive to overcome the ways in which I have been meta-passive. Being perpetually open requires taking my self seriously and taking the world seriously. Considered together, these three attitudes – perpetual openness, taking my self seriously and taking the world seriously – are the attitudes at the heart of governing the self. To take on this attitude is to be a dynamic self, ever-growing as one sensitively responds to new feedback.

However, as it currently stands this account is inadequate. The problem is that as finite, mortal, embodied creatures, there seem to be certain limits on our substantial selves which we cannot simply overcome. While there is much leeway for how we can relate to the self we have been given and the world we have been placed in, each of us faces certain hard limits. A deaf person cannot simply decide to "relate to themselves and the world" as a hearing person would; a 5' 2" women cannot simply decide to try to become a linebacker

for the NFL. There are also fuzzy limits: for example, while I can learn to appreciate things that I initially had no interest in, typically I still find myself being interested in some things more than others. And, of course, there are societal limits: restrictions placed on us by institutions, infrastructures, and the ways other people treat us based on the social categories we fall into. Such limits seem to place constraints on who I can possibly become. As such, they seem to place constraints on my ability to govern myself.

This chapter is about the complexities of dealing with these limits in a way that still counts as governing the self. The overarching theme is that limits are not in themselves incompatible with philosophical autonomy; the relevant question is how we *take up, engage with,* and *inhabit* these limits.

Section 1: The Core Problems

Let us begin by laying out the core complications which the various kinds of limits present for my account. Some of these I will be able to address here, but most of them will require me to add on to my theory of governing the self.

Problem #1: As both Jaworska and Buss have noted, the fact that I am limited actually seems to be essential for getting action off the ground[131]. For many (if not most) actions I take, if I had no initial interests or concerns to guide me I would remain paralyzed

[131] Buss makes this point very explicitly in "Autonomy Reconsidered": "every deliberation is necessarily responsive to some nonrational influences: practical reasoning would be impossible if we did not simply find ourselves taking an interest in certain things, and if we did not simply find ourselves attributing greater significance to some of these things than we attribute to others". In "Caring and Internality", Jaworska argues that secondary emotions, and especially caring, seem necessary to get agency started: without such things to direct me interest towards certain possibilities for action, I would be at a loss for what to do with myself.

with indecision. In other words, the fact that I simply find myself, as a contingent fact, being *partial* to certain things is an important enabling condition for action and for life – and therefore for being a *self* at all. This point might be taken to indicate not only that governing the self is impossible, since I must rely on pre-given content in my substantial self, but that governing the self is contradictory to action and to selfhood since it indicates that this pre-given partiality should not be relied upon.

To a certain extent, this first problem turns out to not be a problem at all. Jaworska and Buss are completely right that I must have some particular and partial standpoint to start out with, and this is compatible with my remarks last chapter that I learn to become a self by first learning how to be a particular self, since I learn to be a self by being raised in a certain culture. The relevant point for governing the self is that we need to have the ability to reflect on and *change* the ways we are partial. (I am reminded of Descartes in the woods: since I need to start by going in some direction, it's good, indeed *essential*, to start by going in any direction. The key is to be able to change this direction.) In this sense, the fact that we are given initial, partial selves is quite fortuitous; the key is to view these selves as a starting point. Furthermore, the goal of my account is *not* to become *im*partial. To be completely neutral would be at odds with taking the world and myself seriously. To take the world seriously is to believe that there are better and worse ways to understand, relate to, and live in it (which, as discussed in chapter 4, requires believing an objective standard exists which I must strive to know); to take my self seriously is to believe that I need to give myself some particular content, i.e., relate to the world and myself in *some* particular

way. As such, some kind of partiality is in fact essential to governing the self. The key is that this partiality cannot be based solely on contingent grounds.

However, this does not fully address the core concern. The fact is that I am a finite, mortal person, and so I can only do so much. If I genuinely expand my perspective to learn to appreciate all the various forms of meaning that other people have discovered or placed into the world, won't I be overwhelmed with choices? In other words, being restricted in the sense of contingent partiality still seems to be an important and positive condition for human action (and therefore selfhood); it still seems necessary to simply find myself being interested in or caring about certain things. How can this possibly be compatible with governing the self? To properly answer this problem, we need to reformulate it in more precise terms. This leads to the next three problems.

Problem #2: How can I take up my *hard individual* limits in a way that counts as philosophically autonomous? To clarify, a "hard limit" is something non-negotiable about how I am constituted; an "individual" limit is something which is based in my particular individual constitution. The most obvious hard individual limits are physical ones: a deaf person, a paraplegic, a 5'2" woman, and a 6'2" man will all have limits given their physical bodies (as well as the social structures they live in which enable or disallow them to use their bodies in certain ways – but this is a separate, if not truly separable, question. See problem #4). Hard limits also include neurological facts about me which I cannot change at will – for example, whether I am neurotypical or neurodivergent. An autistic person cannot simply decide to be allistic – and there are important moral questions about whether it would be right to ask them to try (again, see problem #4 as well as problem #5). Hard

individual limits seem to set boundaries on the kind of self I can possibly be. They therefore seem to set limits on my ability to govern myself.

Problem #3: How can I take up my *soft* individual limits in a way that counts as philosophically autonomous? In contrast with a hard limit, a "soft limit" is one which I can change, but only to a certain extent. A good example is introversion vs. extroversion. An introvert can learn to engage in and enjoy certain kinds of extroverted interaction, but typically they will not be able to change that they are at bottom an introvert. Mental illness, at least some of the time, seems to be similarly "soft". An anxious person can gradually learn coping mechanisms to be more comfortable, but it's unlikely they will ever be as relaxed as their non- anxious peers. Natural proclivities and interests are another soft limit: I can learn to be good at many things with enough practice, but proclivity can give me a boost and interests can give me passion and drive. How likely I am to succeed at a particular endeavor is therefore influenced by these limits. Once again, these kinds of limits seem to present boundaries (albeit slightly more permeable) on how I can govern my self.

It can be tricky to determine whether a particular limit is hard or soft. For example, formative life experiences can have deep lasting impacts. While the brain remains remarkably neuroplastic throughout one's life, it can be extremely difficult to overcome, for example, an abusive childhood. Similarly, but more subtly, the kind of experiences we've had given the circumstances we grew up in and have lived in (e.g., the social groups and economic class we fall into) have a cumulative and lasting effect. It is not intrinsically important to know whether a limit is hard or soft: the categorization is only relevant because hard and soft limits will likely require different kinds of approaches to be properly

incorporated into philosophical autonomy. For now, we can just say that the two are most likely a continuum, and that depending on how hard or soft any particular limit is we will need to apply different strategies to make it compatible with philosophical autonomy.

Since we are finite creatures with only a limited amount of time, it is likely that once we have broadened our perspective enough through perpetual openness, even our hard and soft limits will not narrow down the range of options enough for us to know what to do. Given this, it makes sense to fall back on the things that we have been given – what I have called our contingent partialities. How this be compatible with philosophical autonomy? We have arrived back at problem #1.

Problem #4: How can I interact with *societal* limits in a way that counts as philosophically autonomous? This is the broadest and most complicated kind of limit so far. We should take care to differentiate between unjust social limits and neutral social limits. I will focus mainly on the former, before making some brief comments about the latter. Unjust societal limits are based on socially defined groups which significantly determine the possible options available to me. Race and gender are prime examples. Such limits are external in that they limit us "from the outside", both via how other people treat us and the larger societal structures and institutions. But they are also internal in the sense that they impact how I see, relate to, and value myself. Furthermore, socially defined groups, with their attendant expected social roles and statuses, will impact how I see the world around me, including how I view *others* in terms of *their* social groups. These are precisely the impacts of socialization we have been so keen to avoid. In other words, societal limits in their internal manifestations are prime examples of meta-passivities. In

their internal aspect, societal limits are therefore hopefully things we can overcome (although to be fair, in some cases this might be difficult enough that we may want to categorize them as *soft* internal limits).

But when I speak of "societal limits", I mainly intend to be referring to them in their *external* aspect: the fact that such meta-passivities are *externally enforced*. This includes a wide range of phenomena. Interpersonally, people are treated a certain way based on the social categories they fall into. Furthermore, society is structured in ways that re-enforce these social categories: ways that affirm the value of certain groups over others and the correlation of certain traits with certain groups. This includes economic structures and infrastructures which either enable or close off certain opportunities. Physical disabilities are a key example which demonstrates the importance of infrastructure: how much of the world a modern-day paraplegic can participate in largely depends on whether society has built the necessary infrastructure which enables them to participate. Such things are *limits* in the sense I am concerned with in that I cannot simply change them at will and they seem to significantly impact my options for the kind of self I can possibly have. If the society I live in is racist, I cannot simply change this by my will alone; and if societal racism is *systemic* and *structural*, this limits my concrete options for who I can become.

To be clear, my separation of limits into "individual" limits which are part of the individual constitution of the person and "societal" limits which are part of the structure of society is tenuous and provisional. The categories that we might say are "individual" to the person are often provided by the kinds of knowledge our society makes available: for example, someone being autistic or allistic is a distinction we can make because the society

we live in makes it. (To be sure, if we assume autism has a neurological basis then it will exist no matter the social categories currently available. But the *social significance* of being autistic changes based on whether a society recognizes autism as a category and what significance the society gives to this category.[132]) On the other hand, societal categories are typically based on characteristics that could be considered "individual" in my sense of individual constitution: biological sex (and the gender this is typically assumed to entail) and skin color (and the race it is assigned to) are good examples. While racism and sexism are societal and cultural problems, they rely on signifiers which attach to the individual to function. (The arbitrariness of these signifiers and the confusion caused by trying to consistently apply these signifiers are two key points of criticism. However, these criticisms are entirely consistent with my distinction, since they point us towards the fact that these categories have no substantial basis beyond social categories.)

Nevertheless, I believe that this distinction between individual and societal limits is a useful one to make. The two categories point us towards different problems, and hence will require different solutions. Perhaps a better way to parse the distinction is in terms of a person's self-understanding. Individual limits can be useful for a person as they try to make sense of their own experience in the world, while societal limits are injurious to the person. For example, a person who learns at a late age that they are autistic may find this category illuminating and freeing. It can allow them to make sense of various puzzles and

[132] A further point: material conditions can also make a difference to the significance we give to such things. For example, in a rural society with almost no technology and small family farms, an autistic person whom our society might consider to be especially sensitive to over-stimulation might have significantly less occasion to be overstimulated.

frustrations that they have encountered in their dealings with allistic people, and it can give them a sense that they are not fundamentally "wrong" but simply different. In contrast, a person who has been told that because of their sex they simply *are* a certain way will find themselves frustrated and limited as their internal experience does not fit this mold.

It may be a bit simplistic to say the people will always want to change societal limits and embrace individual ones. Nonetheless, we can tentatively say that internal limits are genuinely part of the person's unique constitution, and as such 1) learning about them gives the person empowering knowledge about themselves and 2) they present a genuine hard or soft limit on the person's possibilities for the kind of self they can become. In contrast, societal limits are imposed on the person, and as such they typically do not match up with the person's internal experience or the person's genuine limits. (Indeed, part of the problem of such societal limits is that they take seemingly "internal"/intrinsic characteristics like race and sex and claim these are deep characteristics which shape and determine the person at the deepest level of character traits and capabilities.) In short, they artificially truncate the person's ability for self-knowledge, self-respect, and self-realization. This is a problem for justice and ethics, but it is also a problem for governing the self. In part, addressing this problem has been a large motivation behind my account of governing the self, since I have been keen to avoid the meta-passivities caused by socialization. But again, we are here talking of societal limits in their external aspect. Since societal limits take the form of institutions and systems which cut off some groups from certain opportunities and possibilities, we are posed with the question of how an account of governing the self can work within these limits.

So far we have been discussing *unjust* societal limits. But there are also *neutral* societal limits: limits which cut off someone's options, but are not unfairly based on the devaluation of certain social groups. Depending on existing social structures, institutions, and technologies, some abilities will be supported and others won't. Among neutral societal limits we can distinguish between society-wide limitations and differentiated limitations. Examples of *society-wide* limitations include a 6'2", 245 pound Roman man who cannot become a linebacker because Rome has never heard of football, a person living in the neolithic era who cannot enjoy the enrichment of books, and a teenage living in the 1950s who cannot become a YouTube star. Examples of *differentiated* limits include the modern-day, 5'2" American woman who cannot become an NFL linebacker – not because football doesn't exist, but because someone of this size simply couldn't play. This is neutral because the petite woman is not being excluded because her abilities are being wrongly judged. She may in fact be an excellent, extremely strong athlete. The exclusion has to do with the fact that if she were included, the nature of the game would be radically changed. Simply put, petite women and large men cannot compete against each other in football without the whole enterprise collapsing. This is therefore a socially imposed limit, since football is a socially-defined enterprise, but it is not unjust[133]. Based on the nature of the activity itself, only certain kinds of people can compete against each other.

These limits demonstrate the extreme edges of our ability to govern the self: there are certain options for the substance of selfhood which I simply will not have access to.

[133] For this case to work as an example of a *neutral* social limit, I am also assuming that the woman would not be prevented from starting a women-specific football league if she desired and that women's sports are not devalued compared to men's' sports.

Both the Roman and the petite woman simply do not have the option of being a linebacker. What does this mean for philosophical autonomy, which compels us to be as much in control of our substantial selves as possible? Because these are non-negotiable limits (and because, being neutral, they avoid complicating questions of justice), they are parallel to the hard limits discussed in problem #2. Their solution will likely be similar.

Problem #5 is closely connected to unjust social limits, but is unique among the issues listed so far in that it presents us not with an *obstacle* we cannot overcome by our will alone, but with an *ethical* limit we should not cross. The problem is this: while it sounds nice to say that we should try to appreciate all perspectives and what they have to teach us, the fact is that we live in a world where certain perspectives are already given pre-eminence while others are actively and systemically devalued. Given this context of social groups ordered in an unjust hierarchy, it seems like there will be cases when one should not simply adopt the values and meanings of other groups, lest one reenforces the hierarchy. This problem is clearest when we take the perspective of those who are members of an oppressed group. We do not want to say those on the bottom should be perpetually open such that they simply adopt the perspectives of those in the dominant group. A clear example of this is the forced assimilation of Native Americans. In such cases where one's entire culture is targeted for eradication, it seems one would be justified in doubling down on one's original, given perspective.

In some ways, this is the most concerning problem my account of philosophical autonomy has faced so far, since it questions not just the possibility of governing the self, but the morality of it. Given that I have argued that governing the self, and in particular

perpetual openness, is only made possible by centering ethical consciousness, it would be particularly problematic if ethical considerations spoke against perpetual openness. To clarify, it is certainly correct for ethical considerations to speak against openness towards clearly unethical perspectives; this is a point we discussed last chapter. The worry is that in cases of systemic oppression, there seems to be reason to prioritize one's original perspective in a way that precludes openness. How can philosophical autonomy make sense of this?

While a full response to this problem cannot be made at this point, for now it gives me the opportunity to clear up a misconception. It might seem like the way I have characterized "meta-passivities" indicates that we need to completely abandon our old perspectives. In a real sense, this would result in a self just as arbitrary as the self I was initially given, since it would be equally determined by the self I just *happened* to be given: the only difference would be that the relationship is a negative one. But this is not what I am claiming. We can assume that however one was first taught to be a self, this initial perspective will have *some* redeeming aspects to it – it will allow us to see and participate in certain kinds of meaning and value. Being perpetually open does not require me to completely abandon all my initial perspectives. Instead, it requires me to think critically about them: to supplement them where they are lacking, to abandon them where they are ethically untenable, *and* to appreciate them more fully where I may have been taking them for granted. My account does not require someone to abandon their original perspective in favor of the perspectives of others. In this way, at least it does not automatically feed into unjust hierarchies.

There is a complementary worry that if those in privileged groups adopt the values of oppressed groups, they will also be contributing to the hierarchy. The main concern here is with cultural appropriation, where a person from a privileged group uses cultural artifacts from an oppressed group for personal gain while society as a whole continues to devalue the oppressed group – in effect extracting value from this group and re-enforcing the hierarchy. But one of the key characteristics of cultural appropriation is that it only takes the surface appearance of cultural artifacts without any deeper appreciation of the meanings that these embody. This is clearly *not* a case of openness in my sense of the term. To be open means striving to genuinely understand the deeper perspectives of a person from a different background and culture.

Addressing this concern also allows me to emphasize an important point we have not touched on yet: to genuinely understand the perspective of a different culture is to see the importance of *context*. Often cultural appropriation involves removing artifacts from the contexts which give them meaning; this is part of what debases them. To truly understand a different culture is to see that one cannot simply take cultural artifacts and use them as one personally pleases. Nevertheless, we are left with two big questions. First, when and how it is appropriate for privileged groups to adopt the perspectives of oppressed? Second, given that many values and meanings are tied up with contexts, how can we incorporate insights from them in the way that philosophical autonomy requires?

Of the five problems, 2, 3, and 4 and the different categories of limits they point to are the key ones we will need to deal with. As indicated above, a particular limit might fall into more than one category. For example, while physically I will simply have or not have

certain abilities – I might be born or become blind or deaf – what my options are in terms of what I can concretely *do* and the kinds of meaning and value I can engage with given these limits depends on societal structures and institutions. Similarly, while my abilities and interests considered in themselves only present me with soft limits, they may present a hard(er) limit if they are not supported by societal structures which enable their expression. In these cases, the problem will likely require more than one strategy: one to deal with the individual aspect of the problem, and one to deal with the societal aspect. How, then, can we make sense of these limits within an account of governing the self?

Section 2: Meta-Passivities vs. Limitations

The crucial move is to make a distinction between two concepts: meta-passivity and limitations. Since meta-passivity and limitations are strikingly similar in certain ways, I will use the term "limit" for the general concept which contains both of them and reserve the term "limitation" for the concept which is distinct from meta-passivity. To remind the reader, I have claimed we are meta-passive when we uncritically exercise our agency along guidelines that we have simply been given. This includes what we take to be reasons; the values we use to guide our decision making process; the larger worldview and belief system which forms the horizon against which we make decisions (*and* against which we act *without* needing to make a decision); and so on. All these deep factors structure the ways in which we exercise our agency (as well as how I live, experience, and express my *subjectivity*; for purposes of conciseness and clarity, in what follows I will speak primarily of agency). Although as a full human agent I am undoubtably active, when these factors

have been pre-given to me, I am passive on this deeper level. This is why I call it *meta*-passivity.

In an obvious sense, all the aspects of my deep agential structure which are meta-passive are *limits*. Based on how I understand myself and the world I live in, I am limited in what I consider to be worth doing; what I conceive of as *possible* to do; the ideals I strive for; and so forth. When such considerations are simply given to me, and especially when they are so deeply embedded in my world view that they are rendered practically invisible, they curtail the possible ways I can relate to myself and to the world. I am confined to what I have been taught. The point of perpetual openness is to allow other people to help me; someone different from me can make me aware of previously unseen possibilities. Understood this way, one could describe my account of philosophical autonomy as one of *overcoming limits*. There is some truth to this description. My account is indeed *expansionist*, since it emphasizes continually learning about the ways I have been blind or dismissive to other forms of value and meaning. The goal is to gradually synthesize a richer, more nuanced way of relating to the world and to myself that encompasses these various forms of value/meaning and the multiple aspects of reality they make available.

This description of my account as "expansionist" is worth dwelling on for a moment. Let me explain another way: the worry with the "limits" of meta-passivities is *not* simply that they were given to me. If *this* was the worry, then the solution would simply be for me to become aware of the meta-passive aspect and then endorse it or reject it. If I endorsed it, the limit would thereby be made acceptable. This would be totally satisfactory for an account of governing from the self. But from the perspective of a governing-the-self

view, we could always ask "*Why* did I endorse this limit? What are the *deeper* structures which lead to this decision, and am I meta-passive with respect to *those*?" In other words, we still would not have solved the problem at the heart of governing the self: the problem that any decision to "endorse" something (or otherwise make it defining of me) requires that I rely on something which I must simply take accept unquestioningly. The fact that these limits were *given* to me *is* problematic, but this cannot be solved by endorsement, because endorsement always relies on another similarly given limit. The only way out of this trap is to expand these limits: in other words, to make them more holistic and more complete, so that they encompass more of the possible meanings and values the world has to offer. In short, it turns out that the solution is to transform the problem. Since we cannot solve it by finding a non-given limit, we must re-articulate the problem so that it becomes about the *limits themselves*, and *not* about their origins. This is why expansionism is essential to my account.

This is also why my account of governing the self – which started as an ostensibly navel-gazing project which seemed to require me to retreat further and further into myself – turns out to require me to look outward in an attempt to encompass more and more of the meanings and values the world has to offer. I overcome my meta-passivities by being committed to engaging with the world as best I can, which requires me to expand my perspectives to see what I have been missing. (This is simply another way of stating that the way to solve the problem of governing the self is to see the problem as one of the limits themselves.) To expand my understanding of the world (which again, includes other people) is to expand my capacity to engage with the world, and therefore to enrich the

interiority defining of selfhood. As emphasized in chapter 3, this involves a change in how we conceptualize the deep control at the heart of governing the self. Instead of being primarily about defining myself, it becomes about fitting myself to the world and other people. Far from being a restriction, this "fitting" requires me to expand myself. In this way, developing a rich and expansive interiority – one which is responsive and sensitive to all the complications and nuances of the world – becomes paramount for governing the self.

To return to the main point: since meta-passivity imposes limits on us, it seems we could describe my account as one of "overcoming limits". But what then are we to do with the fact that there are certain limits we *can't* "overcome"? A deaf person has a certain set of options available to them: the limits of this range of options guides their choices and structures their agency. Conversely, a hearing person has a different set of options available to them. While the hearing person can participate in the world of sound and all the riches it has to offer, a deaf person can participate in deaf culture and all the riches *it* has to offer in a way a hearing person simply can't. Since these are *physical* contingencies (embedded in a social system, of course), we cannot escape the limits each one sets for us. Each one gives us access to only a concrete range of options. If governing the self is simply conceived of as "overcoming limits", it therefore seems that I can never truly govern myself.

The solution is to this problem is found in the realization that *not all limits are meta-passivities*: some are what I will *limitations*. To demonstrate the distinction, it will be helpful to draw an analogy. Consider the process of making a simple decision (setting aside

the complicated issue of whether this decision is "autonomous" in any sense of the word). When someone makes a decision, we can make a distinction between the *content* with which the decision is concerned and the *deliberative guidelines* one uses to make this decision. Take the choice between eating an apple or eating a piece of cake. The *content* of the decision is the apple and the cake. The *deliberative guidelines* one uses are any considerations or general principles which one uses to both frame the stakes of this decision and to reason one's way to a decision: for example, being on a diet to lose weight, wanting to enjoy oneself, or being concerned with getting proper nutrition. Being *meta-passive* is like being limited with regards to our deliberative guidelines. How I exercise my agency is, in a meaningful way, decided for me. But when I face *limitations*, it is like the content of the decision is decided for me, and I am limited in this sense. While each may be said to "structure my agency", they do so in different ways. Limitations set the boundaries of what is a concrete option for me. Meta-passitivies set the (current) boundaries of what I will consider doing – including any available options I cannot even see – and why.

Notice my use of the term "concrete". I use this word to indicate that even with only a few concrete options available to me, there will still be various meanings that could be assigned to the option, and thus a multiplicity of actions available to me. Choosing the apple could be any number of distinct actions: it could be one for health, or for meeting beauty standards, or for pleasure. This is an essential point. As agents and subjects, it is not just what we do, but *why* we do it and the meaning contained in this which matters. In short, it is not just actions, but the *meaning* of those actions which is the domain of meta-passivity.

But how are limitations and meta-passivities meaningfully different in a way that is relevant for governing the self? The answer is that meta-passivities are connected to our agency in a way that limitations are not. We do not control our limitations – they limit our choices, actions, and agency in a way that is not up to us. (This does not mean we are completely at their mercy – we will return to this important point later.) But meta-passivities are connected to things which *can* be up to us – they shape our decision-making process *and* the deeper *interpretations* which form the horizon of this decision making process. (The first could be said to be connected to our agency, the second to our subjectivity.) This second point is essential. Meta-passivities have to do with the sort of things which make us selves in the first place: the meanings, interpretations, categories that shape how we relate to the world and to our selves, and subsequently the actions we take in the world and on ourselves. Meta-passivities have to do with exactly the sorts of things that define and reveal the kind of substantial self I am. In short, meta-passivities operate on the level of the rich interiority which *makes* me a self in the first place. As such, when I talk of governing the self, I mean to talk of having control over the realm of things which are subject to being meta-passivities or not. In contrast, limitations set up the "concrete" options which are actually available to us. We can take up and engage with these options in any number of ways, but we must work within the range of what we are given by these limitations.

But, of course, things are a bit more complicated than this. Although the particular self I am is determined by the *meanings I* give to myself and the world[134], the range of

[134] more precisely, the meanings which shape how I relate to myself and the world.

possible meanings I can have is importantly constrained by these limitations. (For instance, although I can decide to find various meaning in atheletics, I will never be able to become an Olympic figure skater.) In other words, limitations *do* play a substantial role in the kinds of selves I have available to me. More precisely: even if we can give a range of possible meanings to ourselves and to our actions given our limitations, 1) these limitations will set the concrete form these meanings must take and this particular form will itself color these meanings; and 2) these limitations will likely cut me off from engaging with certain forms of meaning altogether. How can I claim to have deep control over my substantial self if my options for what this substantial self can be are pre-determined for me? If such limitations are to be compatible with philosophical autonomy, I will need to engage with them in a way that is somehow compatible with the attitude of governing the self.

I believe this can be done. The key trick to philosophical autonomy is in understanding the difference between meta-passivities and limitations, for each one requires a different kind of response. I will argue that while I must try to *overcome* the ways I have been meta-passive, I must try to *incorporate* my limitations. Each of these might take several different forms. We have already gestured to the various ways I may need to overcome my meta-passivities: some of my meta-passive original perspectives I will need to completely abandon (for example, prejudiced attitudes against others which make me blind to seeing them as sources and guides to new potential values); others I will need to synthesize with new perspectives (that is, I will need to recognize that both the new perspective and my original perspective "get something right", and I must to combine insights from both). But how can I incorporate limitations?

Section 3: Incorporating Limitations

I will suggest that there are two main ways we can incorporate limitations into the attitude of governing the self. The heart of both strategies will be *understanding how our limitations fit within the larger whole of meaning and value*. It is this broader understanding which allows us to engage with our limitations in a philosophically autonomous way. Developing this broader, encompassing perspective is precisely the point of perpetual openness, and is the heart of philosophical autonomy. Therefore, engaging with our limitations such that we fit them *within* this larger perspective counts as engaging with them in a way that fits with philosophical autonomy. This section will first elaborate what this broader understanding which underlies both strategies for incorporating limitations involves, and then move on to explaining each strategy in turn. For the purposes of this discussion, I will simplify things by only discussing hard limitations (that is, non-negotiable limitations in the individual's constitution).

A brief note: in what follows, I will be using the single term "value(s)" in place of "values and meanings". I do not mean to imply any specific theory of value by using this term. I only mean to indicate the various enriching ways we can live in the world and engage with other people and our selves.

3.1: The Broader Understanding

What does it mean to understand how our limitations fit within the larger whole of human meaning and value? What does adopting this attitude involve? Most essentially, it means I must always take the broader perspective I have gained/ am committing to gaining

through the attitude of perpetual openness[135] to be the heart of my identity; although I will have limitations particular to me and which will inform my substantial self-conception, nevertheless this broader perspective will still be the foundation of this substantial self-conception. Explained another way: last chapter, I emphasized I must foremost "identify" with the question "Who will I be?"; I can only "identify" with an answer insofar as it is the best answer I have articulated so far in the process of perpetual openness. I am now claiming that this best answer, formed from all the valuable insights and possibilities I learned from others, must form the core of my *substantial* self-conception – even as my substantial conception is filled out with projects more particular to me. Put simply, I must conceive of my substantial self firstly in terms of my ability to understand and engage with the broader realm of human values and meanings.

In practice, this means that although my identity will always involve particular and partial elements, this broader perspective will frame how I conceptualize and go about pursuing my more particular projects (and in the most extreme cases, cut off certain projects from consideration altogether). Because I will be first and foremost "on the side" of the "bigger picture" of meanings the world and other people have to offer, my concern with doing right by them will shape the way I go about my more particular projects. Being "on the side" of the encompassing whole of values means that I understand my particular

[135] By "perspective" I mean the framework which I use to understand/interpret the world and to deliberate about myself and my actions. By "perspective gained through the attitude of perpetual openness", I mean that this framework for understanding and deliberation is fundamentally shaped by what I have learned from others about what has/is a source of value (and what is *not* valuable) and the multitude of possibilities open for humans.

projects in this larger context. This might seem to require a troubling detachment from my personal projects, but this need not and should not be the case. The goal is to view these projects in terms of *how they fit into the larger whole of value.* This requires seeing these projects as *themselves* contributing to this whole, which in turn requires seeing these projects as being valuable themselves. What follows will make this clearer.

The two strategies described below are two ways of framing my projects in terms of the larger whole. Strategy #1 involves enacting larger values in particular ways. The main question to ask myself here is, "How might I change the way I approach my current projects to enact general values I've neglected?" Strategy #2 involves harmonizing more particular values into a larger web of value. The main question to ask here is, "How do my particular limitations allow me to participate in forms of value/meaning which augment the complex whole of human value?"

3.2: Two Strategies for Incorporating Limitations

There are two ways we can incorporate our limitations. First, we can recognize that there are different ways of participating in the same value, each of which will add its own texture and richness to the value. We can therefore use and work within our limitations to engage with more general values in a unique way. In other words, we can participate in general, "big-picture" values in particular, concrete ways. In this case, it becomes (in some sense) less about *what* I am doing, and more about *how* I am doing it. To give the simplest example, one can work a number of different jobs, but still be committed to doing each job in the same way: with integrity, honesty, and compassion to those one must work with. So long as one mainly identifies with these larger values, one will largely be the same person

even if one changes professions. This first way works best with more general values or forms of meaning – that is, values which are not necessarily attached to a particular form, activity or context, but which can be realized in a wide variety of ways.

The second way we can incorporate limitations is by participating in a more specific kind of value while seeing how it fits into the larger web of meaning. Although I cannot participate in all the values and meanings the world has to offer, I can conceptualize the values I do engage with in terms of how they work within this larger whole. This goes beyond simply recognizing that there is a plurality of values in the world; mere recognition would not be sufficient for philosophical autonomy. I can recognize something as valuable without thereby appreciating it as valuable or thinking that I should be impacted by it in any way. To merely recognize that there are many values while remaining committed to the ones I happen to be most interested in is to fall back into letting my innate predispositions determine who I am; it is to fall back into governing *from* the self. In this case, I am giving my pre-given interests priority simply because I happen to like or care about them more. Once again, we must remember that we need to first and foremost conceptualize ourselves in terms of the broader perspective of value. In this case, this means I cannot simply recognize that other values exist; I have to appreciate them and see my more particular projects as in some way "harmonizing" with them. This second way works best with more particular forms of value or meaning – that is, values which are largely bound up with a particular material, activity, or context, and which are therefore not generalizable. Let us turn now to details, which will clarify what each strategy involves.

3.2.1: Strategy #1 – Enacting Larger Values in Particular Ways

It will be useful to draw an analogy. Consider different artistic mediums. Each medium presents its own limitations and its own possibilities. Music and sculpture are of course very different, but even similar mediums such as watercolor, oil, and acrylic paints all have their own challenges and opportunities. Many meanings (experiences, ideas, etc.) can be expressed with any number of mediums, but most mediums will have certain meanings which will be difficult for them to express (for instance, musical compositions are not capable of portraying the same complex meanings that a novel can). This means that while the medium one uses is in some sense secondary to what one makes of it, the medium also sets limits on what one can make of it. However, the techniques required to make a piece of art express something are going to be quite different depending on the medium. Successfully evoking a particular experience with a piece of art will require the artist to become intimately familiar with and to master all the particularities of the specific medium she has chosen. Evoking an experience or provoking thoughts that go beyond the particular medium require working within the particularities of the medium. And of course, the "content" of a piece of art – what is expressed/evoked – is undeniably colored and shaped by the particular medium which provokes it. Listening to a piece of music is manifestly different from gazing at a sculpture. Although each medium can express similar meanings, these meanings take on a texture bestowed by the medium itself.

Before I explain how artistic expression provides a model for incorporating limitations in philosophical autonomy, we must note one respect in which the two are *not* analogous. Art is agnostic about the meanings it conveys, expresses, or provokes. In

contrast, philosophical autonomy requires us to *not* be agnostic. While the artist may play with the rich fields of meanings and experiences, deciding to focus on only those they find interesting, the philosophically autonomous person is bound to take up the expanded realm of meanings which they have adopted as a result of perpetual openness[136]. Philosophical autonomy requires us to see the multiplicity of real values the world has, to synthesize them as best we can, and to live in a way that takes seriously this larger realm of meaning, not just focus on what we happen to find interesting to the exclusion of anything else[137]. This will be an important point to remember.

In this analogy, the *mediums* available for one to use are like the *limitations* one is faced with. If one is deaf, one only has access to certain kinds of "mediums" (and the point holds in reverse – if one is hearing, one will also only have access to certain kinds of mediums, since one will not be able to fully participate in the deaf community). Like mediums, the limitations one has can preclude certain kinds of meanings The *meanings* conveyed (expressed, evoked) by a piece of art are the various ways we can express the rich internality central to selfhood; that is, *the various values and meanings we can express*

[136] Two clarifications: 1) "Taking seriously" this larger world of meaning does not mean that I have to give expression to all valuable meanings. This is simply impossible for a finite mortal. The second strategy, discussed below, will be essential in addressing this point. 2) I am not requiring that the philosophically autonomous person should take all *possible* meanings seriously – some meanings are obviously silly and do not require being adopted/taken up in any way. For example, a child who puts goldfish crackers in water before eating them so they "get a chance to live like real fish" is not demonstrating a valuable new way to relate to our food. (Example from reddit user ZaPandaz: https://www.reddit.com/r/AskReddit/comments/hlij7r/comment/fx0ug4k/)

[137] It may be worth noting that art can be useful to P.A. in that it can help introduce us to new forms of meaning.

by relating to the world and ourselves in different ways. Finally, *meta-passivities* are represented by any *limiting ideologies* which truncate the scope of possible meanings we could express with the mediums we have: for instance, 1) the possible meanings we can *conceive* of expressing with a piece of art; 2) the meanings we think are *worthwhile* to express; and 3) whether each medium is considered more *valuable* than another. Notice that these limiting ideologies are *not* intrinsic to the mediums. Although each medium will pose its own challenges and potential limits, these challenges can typically be approached and even made use of in multiple ways. Similarly, meta-passivities limit the ways we can take up the internality of the self. They place constraints on how we develop and express the rich internality which is defining of our selfhood.

This analogy demonstrates a previous point that I made: viz., although limitations set the concrete options available to us, we still have extensive range for what we do with these options and the meaning we give them. The main point I wish to make now is that for one to be philosophically autonomous, one does not need to have access to all the possible mediums/materials. But this is closely connected to the first point. What matters is how one's meta-passivities cut one off from certain possibilities – how they make certain forms of value difficult to see, to appreciate, and to consider as genuine possibilities for ourselves. Put another way, what matters is not so much the reach of our hand, which is determined by our limitations, but the clarity and breadth of our eye, which is determined by our meta-passivities.

Just as a medium can be used to express a variety of meanings, a *limitation* is not *in itself* a "closing off" of meaning. We can work within and use different limitations to

express similar values and meanings. The worry is that our meta-passivities will limit the range of human expression we permit ourselves (as well as the expressions we permit others). Once we have expanded (and are committed to continuing to expand) the values and meanings we recognize, often we will be able to work within our limitations to express any number of these meanings. For example, a deaf person has all different kinds of meaning available to them; although they are denied certain "materials" or "mediums", they can "shape" the material they do have into an array of forms just as rich and expressive as the hearing person's[138]. The question thus becomes less a matter of the limitations, and more a matter of what is *done* with those limitations.

However, the particular material with which a meaning is expressed also colors the meaning, adds dimensions to it. Just as the particular medium used does in fact matter to the meaning an art piece conveys/evokes, adding further texture and layers of meaning, the fact that these forms of meaning and value are lived as a person working within certain limitations – like deafness – is not to be ignored. In other words, we should recognize that a person's deafness *in itself* has added a new kind of value that we should recognize even if we (assuming "we" are hearing people) cannot fully participate in it.

Let us look at an example. One meta-passivity which many westerners share is the inability to see value in engaging closely with one's local ecosystem. There is a long

[138] It's important to note that this example also involves important issues of social justice – for instance, is the deaf person given the same opportunities as a hearing person to participate and engage with meaning, especially in a community of similar persons? If they are denied this opportunity, this is an unjust societal limitation, and it is not the person's job to "incorporate" this limitation – rather, the societal limitation should be eliminated.

history, tied up with western industrialization and Christian doinionism, of seeing nature as merely raw material, which only takes on value when we form it into something we can exchange on the market; that is, it only takes on value insofar as it is essentially *interchangeable*. To pay close attention one's natural surroundings, to see it as *worth paying* attention to, to appreciate and be attuned to the cycles of one's local biosphere – all this requires us to overcome the meta-passivity which prevents us from seeing value in these activities, cuts us off from certain ways of engaging with the world, and thus limits the ways we can take up our selfhood. Now, a physically disabled person might not be able to engage in their local natural environment in all the ways that an abled person could; activities like hiking or planting a native garden might be unfeasible for them. This is a limitation. But they can still engage with their natural environment in a host of other ways: by learning about it, making art about it, becoming sensitive to its cycles and finding ways to incorporate these into human activities (for example, cooking with ingredients as they naturally become available). A disabled person who lives in a present-day western country will likely be subject to the meta-passivity described above, in which case they are cut off from this particular value of connecting to the local ecosystem. This is something which they (and all westerners) will ideally overcome. Although they are limited by their physical disability, this does not entail that their ways of meaningfully engaging with and expressing this value are any less rich than the ways of a non-disabled person.

In sum, while limitations can set the concrete options available to one (the "medium") – and this is not insignificant – what matters is how one takes up the concrete options available. Meta-passivities limit how we conceive of our limitations – the possible

meanings we think we can express with them, what we think is appropriate to express with them, and so on – and hence limit what we can *do* with them. To become philosophically autonomous, we do not need to overcome our limitations: we need become less meta-passive in how we *conceive of* and *take up* these limitations. To be perpetually open means learning about all the ways we have been meta-passive, learning from new perspectives that show us new forms of meaning and value. We then adjust our perspectives and how we engage with our limitations based on what we have learned. Often, this will require the additional element that we learn about the unique possibilities and challenges that the limitation presents, so that we know how to work with it and how to leverage it to "live" it in the ways we have determined appropriate. Finally, the particular limitation (the particular "medium") we have used to express these values is significant; since these add a layer of meaning, we should recognize the value of these particular expression of meaning *qua* the *particular* expression it is.

3.2.2: Strategy #2 – Enacting Particular Values

So far we have discussed how we can incorporate limitations by focusing on what we express with them and making sure that this choice is informed by the perspective of perpetual openness (that is, the perspective of governing the self). But as noted before, this mainly works for more general values. There are many particular values which are not similarly generalizable. For example, the value of being a programmer is quite different from the value of being a novelist. The programmer and the novelist might both be committed to humanitarian projects, but both *how* these projects will help and the *aspects* or *faculties* of the actor's humanity they engage (i.e., the kinds of meaning they enact) will

likely be quite different. Programming enables a certain type of abstract thinking: one that emphasizes breaking down a problem into smaller pieces and seeing the logical relations between these pieces. In contrast, writing fiction requires a sensitivity for the connotations of different words and the ambiguities of human experience. These mindsets are both valuable: both of them allow the person to develop one of their human capabilities, thus expanding their ability to meaningfully engage with the world. But given that we are finite and mortal beings, we simply cannot participate in all the particular forms of value available to us. I cannot be a programmer *and* a writer (*and* a baker *and* a scientist *and* a philosopher. . .)

This brings us to the second way we can incorporate our limitations: focusing on how more specific forms of value fit within the larger whole of meaning. What does this mean in practice? Most obviously, this means pursuing these projects in ways that do not cut off the possibility of others pursuing different kinds of legitimate values. This includes not demeaning these other kinds of meaning actively or inadvertently (perpetual openness should obviously preclude this). Instead, we must be *curious* about other people's projects, even if we are not actively pursuing them ourselves. Simply learning to appreciate other particular values is a way to expand myself, for it means I conceive of the world I live in more richly than I would if I was simply engrossed in my own projects.

Secondly, where possible we should be concerned with how we are contributing to the whole of value and meaning. Are we buttressing or keeping alive meanings which are in danger of being dismissed or lost? Or are we perhaps contributing to other people's ability to flourish and thus engage with various kinds of meanings? For example, I could

be mainly focused on honing my own abilities as a philosopher, but if I am too narrowly focused on myself, I will not be prioritizing the broader perspective. But if I also focus on how by honing my own abilities I can better contribute to projects bigger than myself – for instance, I can become a better teacher, and pass down the skills and knowledge I have acquired – then I am understanding my project in terms of how I connect up with the broader whole. This particular strategy is roughly analogous to being a single piece in a larger puzzle. I know I am only one piece with a particular shape, and I can only personally manifest part of the larger picture, but I nevertheless see and appreciate the larger picture I am contributing to. Furthermore, in cases where I pursue my particular projects in the context of contributing to this larger whole, these projects become shaped and colored by this whole. If I pursue a teaching career in the context of being committed to social justice, then this will impact the way I teach and what I teach. In this way, I participate in the whole by engaging with a particular part.

Most obscurely, we should have a sense of *gratitude*. Since we appreciate the value of these other projects, and since we cannot personally engage with and enact all of them, we should be grateful that *others* are enacting them and keeping these forms of meaning alive. This is essential to appreciating these other values as opposed to simply recognizing that they exist. When I am grateful these other expressions of the human spirit exist, or that parts of the world which make possible kinds of human meaning exist, I acknowledge that although I may never be directly involved with these forms of meaning, they still have "something to do with me" *qua* self. They are expressions of the same kind of expansive selfhood I have.

In all these ways, we conceptualize our pursuit of specific values, sometimes modifying *how* we pursue these values, in the context of the broader perspective we uncover via perpetual openness. We therefore ensure that this broader perspective has pre-eminence; in short, we incorporate our more particular projects into the attitude of governing the self.

3.3: A Final Case – Transforming Meta-Passivities into Limitations

We have now touched on the two keys ways we can deal with limitations. But there is one final possibility we have not yet discussed: in some cases, we might "transform" a meta-passivity into a limitation. This happens when we become aware of a meta-passivity which has been impacting our agential decision-making process and realize that we can't truly overcome it. In this case, the proper response is twofold: 1) to synthesize the original, meta-passive perspective into a larger understanding and 2) to realize that it presents a genuine *limitation* for me. One example of this (perhaps a bit silly) is found in the phenomenon of love languages[139]. A "love language" is the particular way a person is naturally inclined to both receive and show love. For instance, one may find that the verbal expression of love and care is the most potent way one feels loved, in which case this is also the way one will be most likely to show love; or one may find that physical closeness is how one best feels loved, and so *this* will be how one expresses love, and so on. The point is that our own love language may feel so natural to us that we have a hard time recognizing other ways love is expressed (and we may not be conveying our love in ways

[139] The idea of a "love language" was first introduced by Gary Chapman. Chapman claimed there were five languages, but I think that we need not be beholden to his initial categories for the basic idea to retain its usefulness.

which are most accessible to our loved ones). This means that love languages are a meta-passivity: they significantly impact how I interpret and emotionally respond to others, and how I act.

As with all meta-passivities, I will need to overcome my love language – I will need to become aware of how it is limiting my perspective and my ability to appreciate the different ways of expressing love, and my ability to express love in different ways. But while it may be relatively easy to expand the ways I express love, it might be substantially harder to change what I immediately emotionally respond to. In this case, I may need to accept that this is a limitation for me. The goal will therefore be to grow beyond this where I can: growing my ability to express love in different ways; appreciating the meanings embodied in each unique way of showing love (including the way that comes naturally to me); and seeing and appreciating when others are showing love to me in a love language other than my own. However, since this is a limitation, I may also want to inform my loved ones what my natural love language is and ask them to tailor their expressions of love to this. In this way, I can accept my limitation *and* work with it within the boarder understanding afforded to me by perpetual openness.

We now have a sense of how philosophical autonomy can deal with both kinds of limits. We must strive to become aware of and overcome our meta-passivities: the guidelines along which we have been tacitly exercising our agency (and living our subjectivity). In some cases, this awareness will reveal that our meta-passivities are negative – that they relate us to the world and to ourselves in a way that is incompatible with selfhood. In these cases, we will need to abandon our old perspectives. In other cases,

we will discover that our meta-passivities were like a lens: a partial perspective which simultaneously brought certain valuable forms of meaning into view and obfuscated others. In these cases, the goal will not be to abandon our previous perspectives entirely, but to synthesize them with new perspectives and the new insights these bring into view. In both instances, it makes sense to say we *overcome our meta-passivities*, since by either abandoning them or integrating them with new perspectives we are no longer being "taken in" by them, but are instead actively taking them up and deciding what role they will play in our selves.

While we must overcome our meta-passivities, we should *incorporate our limitations*. The goal here is to ask how our limitations can be made to fit and harmonize with the larger whole of value. We can do this either by using our limitations to engage with larger meanings in unique ways, *or* by engaging with particular meanings in a way that complements the larger whole. It is important to remember that these two movements of overcoming meta-passivities and incorporating limitations are essentially related to each other: it is only by overcoming our meta-passivities that we continually develop the larger understanding under which we can appropriately incorporate our limitations. In this way, I accept that I am a particular self while still taking the question of who I will become seriously.

With this conceptual apparatus in place, we can now turn our attention to addressing the series of concerns laid out in section 1.

Section 4: Solving the Core Problems

In section 1, I used the term "limit" when articulating each of the five problems. We can now more accurately call the limits with which these problems are concerned *limitations*. It will make sense to address these problems slightly out of order.

Problem #2 was about how we can incorporate *hard* individual limitations. In many ways, my account of incorporating limitations speaks most directly to this problem. The goal with hard limitations is work within them, either to participate in general values in a unique way or to participate in more specific values in a way that harmonizes with the whole of human meaning. This will involve fully accepting our limitations such that we don't just accept them, but also appreciate the forms of value they make available to us. This may acquire an extra layer of difficulty and significance if our limitations are socially devalued and marginalized – but this has to do with problem #4. Accepting our limitations will also involve learning about our limitations: the ways we can use them and the possibilities they open up as well as the obstacles they present.

Because *soft* individual limitations (problem #3) are more permeable than hard limitations, they require a more flexible response involving a dual perspective. On the one hand, we should not see our soft limitations as absolute, and we should try to stretch beyond them when this seems especially valuable. The case of introversion and extroversion is a good example of this. An introvert may lose out on opportunities to interact with diverse kinds of people, while an extrovert may lose out on opportunities for self-reflection. Both the introvert and the extrovert should be willing to go outside of their

comfort zone for purposes of growth which allows them to engage more richly with the world and with their own selfhood.

On the other hand, it will always take some effort to push past our soft limitations, and we are finite beings with a limited reserve of energy. The introvert may never be as comfortable interacting with large groups of people as the extrovert, and although he may learn to do so it might exhaust him to do it too frequently. Given this, it will sometimes make sense to treat these limitations as *hard* and to apply the two strategies outlined above. The introvert and the extrovert will be able to participate in many of the same kinds of values, often in ways particular to them (strategy #1). Both will engage in valuable social relationships, but the introvert may only have a few close friends while the extrovert has a wider array of friends and acquaintances. Furthermore, there may be particular values which each person can more easily participate in given their introversion and extroversion (strategy #2). For instance, the extrovert may be more keen to engage in group sports and get bored with reading for hours, while the introvert may be just the opposite. To be sure, the introvert and the extrovert should both try to understand and appreciate the value of the activities they are not naturally inclined towards; they must remember to cultivate the broader perspective.

Striking a proper balance between trying to push beyond our soft limitations and accepting that it will sometimes be more feasible to treat them as hard limitations will be an ongoing project. Since soft limitations are permeable, it would be disingenuous to let ourselves entirely off the hook and not take advantage of opportunities to grow when this would be valuable; but since we have limited reserves of energy, we should be sensitive to

when our energy might be better spent elsewhere. This connects to the next problem we will discuss.

Problem #1 deals with the most non-negotiable limitation of all: that we are finite and mortal. We simply won't be able to participate in a meaningful way with all the different forms of meaning the world has to offer. Given this, it often makes sense to fall back on our contingent partialities (i.e., what we just happened to be interested in). How is this compatible with philosophical autonomy?

Another way of putting this point is that we will not be able to fully express the broader perspective we carefully cultivate via perpetual openness. But since philosophical autonomy requires us to first and foremost identify with this broader perspective, we must try to express this broader perspective in other ways. The second way of incorporating limitations will be most applicable here. Since we can only take up a small number of projects, we will only be able to participate in a small percentage of the particular values the world has to offer. The key, therefore, will be to pay careful attention to the way we conceptualize and pursue our particular projects: to make sure that they harmonize with and don't demean other forms of value; that we try to connect them up to the larger whole of value; that we adjust as necessary as our perspective grows; and that we appreciate this larger whole. Since most particular projects will allow us to express a wide variety of more general values, we should be on the lookout for ways we can integrate new forms of value we come across. Being aware of how we are *contributing* to the whole of value and meaning is perhaps most important for conceptualizing our particular projects in terms of the broader perspective we uncover through perpetual openness.

For example, I have spent the past ten years learning about philosophy. I did this because I happened to find philosophy especially compelling; however, I pursue this project within the context of my broader understanding in two ways. First, I try to contribute to bigger forms of value projects, especially with the courses I teach. I choose to teach content and skills which are not only philosophically rigorous but relevant to contemporary issues, hoping that this will empower students to engage with the world more critically and autonomously. (Perhaps they will even choose to pursue justice-promoting projects based on this understanding.) Secondly, I try to learn about perspectives outside of philosophy and incorporate new insights from these, both in the way I teach (and what I teach), and in the philosophy I do. Finally, I remain aware of all the forms of value which are not philosophy, and I respect and appreciate these. In this way, I am guided by a contingent interest of mine, but I am not completely sequestered in this interest.

This suggests that we should provisionally treat our natural interests and proclivities as *soft* limits: trying to expand them where we can, but recognizing that in most cases our energy will be most effectively spent trying to work within them. I would not be able to contribute to the whole of value as effectively if I spread myself among many small projects, and I would not be able to engage as deeply with the value each one presents. Just as with soft limits, it will be important to strike a proper balance: we do not want to rest too complacently in the limits of our natural interests.

Such cases where it makes sense to follow our natural interests reveal how it is possible for philosophical autonomy to incorporate the value of both political autonomy and authenticity. Being free to pursue one's natural interests (political autonomy) is not the

defining goal of philosophical autonomy, but what comes "naturally" to us can be of use. If we take the perspective of the larger whole of value, we can see our natural interests and proclivities as making us uniquely fit to contribute to the whole in certain ways. Since not everyone will have our natural interests and proclivities, not everyone will be able to see the forms of meaning which we are able to. If we engage with a form of meaning which we find uniquely compelling (the idea behind authenticity), we can introduce this kind of value to others who have been missing it. I have emphasized being open to new perspectives, but this relationship goes both ways: often *our* perspectives will be important for enhancing the philosophical autonomy of others. It will not be enough to pursue what we are interested in simply because we are interested in it; we will need to think critically about what we find meaningful about these particular interests and be able to articulate these insights[140]. In other words, we need to not just be interested in X, but we need to understand why X is valuable. Such articulation will enable us to open up this perspective for other people. This presents a final way we can conceptualize our particular projects within the larger whole of value: we can see ourselves as preserving and making available forms of value which can potentially enrich others.

Problem #4 asked how we can deal with societally imposed limitations (remembering that this has to do with societal limitations in their *external* aspect, and not insofar as we have *internalized* them; in this second case they would be meta-passivities,

[140] Charles Taylor talks about articulation: I think he's on to something important. By being able to articulate a way we have been relating to/living in the world/ourselves, even if we've already been doing this implicitly, we bring this into the sphere over which we can successfully incorporate it.

not limitations). Let us start by focusing on unjust societal limitations. The answer here has two parts: how we respond as individuals and how we respond as part of a larger group. As individuals, these limitations are closer to being hard limits than soft ones: frequently, they present an almost unyielding obstacle. Since these limitations are societally imposed via systems, institutions, and so forth, the only way we can genuinely destroy these limitations is by collective action. (We will return to this point in a moment.) But this need not mean we should just accept them in our individual lives. It would be unhelpful to treat them as hard limits, since this would require us to simply accept them; and if we simply accepted them, they would never change.

A roughly Foucauldian understanding of power is useful here. Power is not simply a top-down matter: it is also inherent in more mundane interpersonal interactions. This means that every interaction where power is present is also an interaction where power can be resisted. If a person from a traditionally upper-caste talks over a person from a traditionally lower-caste during a meeting, this is an instance of power. But this means that if the person from the lower-caste (or someone else on their behalf) refuses to have their point be dismissed and forgotten, this power dynamic has been challenged. In other words, societal limitation as enforced via individual persons can be challenged in a multitude of ways. This example is undoubtably a small victory, but it is a victory nonetheless. Of course, this is not to say that a person should challenge power wherever they see it; in many instances, this would lead to harm (or even death) for the person breaking out of their socially-sanctioned role. My main point is simply that although societal barriers are often more or less “hard”, we should not be resigned to treating them as completely set in stone.

Therefore, our *individual* response to these societal limitations should involve us resisting them where we can, while working within them where we must. It is significant that intractable social barriers need not preclude philosophical autonomy: this allows us to recognize that even people in extremely disadvantaged circumstances are able to govern themselves, and thus able to develop their humanity in the unique way presented by philosophical autonomy. Nelson Mandela presents an excellent historical example. During his 18 years of imprisonment on Robben Island, he gradually fought for and won increased rights for himself and other prisoners: they were allowed to wear long pants instead of shorts in winter (a right traditionally denied to black prisoners); to read, study and have desks in their cells; to play sports and music. Mandela himself described these struggles as a continuation of his core values: "We regarded the struggle in prison as a microcosm of the struggle as a whole. . . The racism and repression were the same: I would simply have to fight on different terms". Mandela's circumstances were extremely disadvantaged: prison is a clear case of a societal institution which is unquestionably "hard" for the prisoners. Despite working within extreme hard limitations, it would be inaccurate to claim that Mandela was not autonomous in a deep and important sense.

(This example also serves to emphasize a core point: the possibility for a wide range of external expression for one's values is not the core issue for philosophical autonomy. This is important to remember while living in an unjust and unfair world, where not everyone is given the same opportunities. Even if one is not given all the chances one should be, one can still practice the art of self-governance, and one can still enact authentic values in ways which demonstrate genuine autonomy. My account allows us to recognize

and admire someone who is deeply autonomous while being in extremely perverse circumstances[141].)

Of course, the fact that not everyone has all the opportunities they ideally should is a core ethical issue; my points do not speak against this. This is the obvious way unjust societally imposed limitations are *not* like hard individual limitations: they are unjust! By cutting people off from participating in certain kinds of meaning, they cut people off from developing their abilities to the fullest extent and from expressing the full extent of their rich interiority. (In many historical cases, this was precisely their point.) Although my account philosophical autonomy has been largely individualistic, it does have something to say about this.

Since perpetual openness requires us to care about the world and other people such that we want to do right by them, a *prima facie* reason to fight against injustice is baked in. The goal of philosophical autonomy is to get our selfhood "correct" – that is, to expand it to be able to appreciate as wide and accurate a range of values as possible, and to allow this perspective to shape not just our understanding, but our affective and motivational being. If we practice philosophical autonomy, we should come to care about other people and the injustices they face; and care enough to want to put in the hard work to learn about the complex causes and potential solutions[142]. We can expand upon this prima facie reason

[141] It is important to not use the strength of the human spirit against people's inherent vulnerabilities: that is, we should be careful to not use the fact that people can develop core aspects of their humanity as an excuse to not support human flourishing in all the ways we can.

[142] This points towards something I have not been able to adequately address in this dissertation: the importance of learning about the complex context one lives in.

with a further consideration. Each person's perspective gives us an opportunity to expand our own – i.e., to become more philosophically autonomous, to develop more fully all the rich possibilities for expressing our selfhood. The silencing and marginalizing of another person's perspectives (or "ways of being"); the limiting of another person's ability to fully develop and express all their potentialities; these injustices cut off opportunities for my own development. An unjust limitation on someone else which curtails the full blossoming of their selfhood therefore harms *me*, and curtails the full blossoming of *my* selfhood[143]. What precisely each individual person can and should do to fight against injustice is quite a thorny question. Presumably it will involve the limitations one has, and both the limits and unique opportunities these present.

Problem #5 dealt with what can be considered an upshot of unjust societal limitations: since we live in an unjust hierarchy, this means that there are ethical limits to how precisely one can adopt the perspectives of others. People in oppressed groups should be allowed to be partial to their original perspectives, and people in privileged groups should not simply take on the perspectives of those in oppressed groups as if they were their own. Let's start with the latter.

Recall that particular manifestations of values and meanings can vary. In cases of unjust hierarchies, those in privileged groups should try to integrate insights they have learned from oppressed groups without appropriating the specific forms and ways these

---

[143] One may be reminded of John Donne: "Each man's death diminishes me,/ For I am involved in mankind./ Therefore, send not to know/ For whom the bell tolls,/ It tolls for thee."

values have manifested. Consider the relationship of modern-day white Americans to the cultures of aboriginal Americans. As many Native authors themselves have pointed out, the dominant culture in America has much of value to learn from traditional indigenous cultures: the idea of humans as simply one natural species among many; a deep respect and love for nature; non-punitive ideas of justice; and emphasis on interconnectedness among people and among species, to name just a few. These broad values, which are common throughout indigenous peoples, have a wide variety of manifestations; for instance, the specifics of the ceremonies which express and enact these values are unique to each tribe. (This is closely connected to the fact that each tribe has an intimate relationship with the *specific* region they historically lived in – one of the reasons why Indian relocation was particularly devasting. Being attentive to and building a relationship with the specifics of one's home environment is another value we could learn from, as discussed above.) Since white persons do not share the history of indigenous groups, they cannot lay claim to these more specific (and more intimate) manifestations of meaning. But they *should* strive to learn from and adopt the general values which underlie these specific cultural practices, modifying them in ways which make sense in their own context.

The other side of the problem is that oppressed groups have a special reason to be uniquely partial to their initial perspectives. This is in fact entirely compatible with philosophical autonomy. There is value in engaging with particular perspectives and meanings which are in danger of being lost. This is the flip side of gratitude discussed on section 3.2.2 above. Just as we should be grateful that others are enacting forms of meaning which we cannot, we should recognize that lost forms of human meaning are a loss for

everyone. If a particular kind of meaning (in this case, entire cultures of meanings) is in danger of being lost, this diminishes the human whole. By re-affirming the value of their particular culture and doing what they can to ensure that this form of meaning is not lost as a result of forced assimilation, Indigenous peoples are in fact contributing to the larger whole[144].

Of course, in order to be philosophically autonomous one must still be perpetually open in all the ways I have described. That is, they must still be committed to recognizing other forms of value and meaning, and where applicable, they must be committed to incorporating insights from other perspectives into their particular projects[145]. One will not be able to incorporate every new value one acknowledges: and since in this case one will be uniquely committed to the particular projects of preserving the core of one's initial perspectives, this point will be especially true. But for the purposes of being philosophically autonomous, the person must inhabit the larger perspective in addition to their particular perspective. In other words, they must both appreciate that their culture is valuable in itself, *and* understand that their culture is a particular form of value which is part of a larger whole. These perspectives are not fundamentally at odds: the broader

[144] One might worry that I am saying that these cultures should only be preserved because they contribute to the whole of value, or because they help other people become philosophically autonomous. But this misunderstands how value works. Particular cultures contribute to the whole because they, as the particular forms of meaning they embody, are valuable. This is why to lose them is to impoverish the whole.

[145] Most obviously, this will include useful ethical critiques: viz., ways that one's traditional culture itself marginalizes or devalues certain groups of people. This sort of critique and adjustment does not itself devalue one's traditional culture. In fact, treating cultures as static is what devalues, since it sees people as incapable of change.

perspective of philosophical autonomy gives an additional reason to value the particular one[146].

To be sure, many traditional indigenous cultures do *not* think from a cosmopolitan perspective. Their world views are rooted in a particular locale and typically place their tribal group at the center of the world. In fact, this both demonstrates the limits *and* proves the general theory of philosophical autonomy. To a large degree, philosophical autonomy is only possible when two *significantly different* groups are forced into close, continuous contact with each other. In other words, it only becomes a concrete issue in the context of cosmopolitanism. But problem #5 – which of course is a problem which assumes the project of philosophical autonomy – applies to indigenous cultures who are oppressed, i.e., who have already been drawn into a cosmopolitan context.

This leads to an essential clarification: to speak of a person's "traditional culture" as if this is simply passed to them intact is, of course, a gross oversimplification. Typically, to even make the distinction between a "traditional" culture and a hegemonic culture, one would need to live *in* a society where their "traditional" culture has been marginalized: in other words, the person straddles *both* cultures. (Many people have written about this;

---

[146] Are there some forms of traditional values which are simply incompatible with philosophical autonomy? Philosophical autonomy precludes anything which denies selfhood (understood as I described it in Chapter 4) to other people or myself. So the question becomes, are there forms of value are predicated on this very denial? Put another way: are there forms of value which could not be realized in any society *except* one which had social/cultural/economic forms defined by denying selfhood to certain people? It seems difficult to rule out this possibility completely. Would there be a principled way to decide between such values and philosophical autonomy? I cannot adequately address these questions here.

Anzaldua calls this the “borderlands”; bell hooks calls it “living on the margins”.) In speaking of “traditional cultures”, it might seem like I am saying that we can evenly separate the individual into a “traditional” and “assimilated part”, and that the only authentic way to take up traditional cultures is to recover the culture in some pristine past historical form. It may also seem like I am saying that someone who is living “on the margins” in this way is no longer a genuine part of their culture, and that in order to be a “real” X, they have to reject the parts of themselves that have been impacted by assimilation. These would be very problematic implications of my account, and they would only increase the alienation experienced by those in this position.

The point of re-affirming traditional values and perspectives is *not* to preserve the culture as some pristine and static artifact. The point is rather for current members of the culture and group to validate their own perspectives, to affirm the values and meanings they see in their culture which are generally dismissed by the dominant culture. This includes not just what their culture was like in the past but what it is like now (and this itself will be diverse). Cultures are not stagnant things. In fact, to think of a culture as essentially static would be to deny the selfhood and activity of the people who are a part of that culture. Therefore, to preserve and affirm a culture one does not need to “go back” to the culture exactly as it was in some idyllic historical form. Instead, core perspectives and values can be preserved in new forms adapted to new contexts. (In this way, the point I made about how general values can be expressed in a variety of different particular ways applies *within* a culture.) The culture lives on and evolves with successive generations of people: indeed, it evolves as each individual adapts it to their own unique needs and

circumstances. In short, I am *not* saying that there is only one right way for someone who lives in the borderlands to affirm their traditional culture. How each person reconciles their conflicting identities – and how they do this in a way that is philosophically autonomous – is a deeply specific and personal question.[147]

Finally, I do not mean to decree that people who count themselves members of marginalized (sub) cultures *should* take up the project of preserving more traditional values. I simply claim that *if* they choose to do so, this kind of partiality can be totally compatible with philosophical autonomy (assuming, of course, they have taken up the perspective of perpetual openness as described above).

Conclusion: Taking Responsibility for Oneself

Chapter 4 articulated the foundational attitude at the heart of governing the self: perpetual openness motivated by taking both my self and the world seriously. In this chapter, I have tried to explain what this attitude actually looks like in practice, in all the messiness and confusion of the real world. The overarching concern has been how I can reconcile the claim implicit in philosophical autonomy that I can expand beyond my current limits indefinitely with the banal fact that I *am* limited in innumerable ways. I have argued that there is a difference between meta-passivities and limitations, and that while we must overcome our meta-passivities as we strive to broader our perspective in accordance with perpetual openness, we should incorporate our limitations so that they fit within this

[147] It's worth noting that living with this kind of double consciousness, while often deeply alienating, can in fact be a great boon for philosophical autonomy. In these cases, one is already practiced at viewing and critiquing oneself from a different perspective. Of course, this does not automatically mean one is perpetually open in the way required for philosophical autonomy.

broader perspective. The goal is not to ignore or downplay our limitations, but to accept them and see what we can make of them. In short, I must take up both aspects of my existence. I must recognize that my essential selfhood is not essentially limited, but admits of an expansive understanding which corresponds to the richness of the world and allows me to see, appreciate, and even participate in this richness. In this way, my selfhood is the same as everyone else's. But I must also recognize that I am still always a particular self: because I have certain hard and soft internal limitations, exist in a world with various societal limitations, and because I have the ultimate limitations of being finite and mortal, the ways I will take up this expansive internality will necessarily have a specific shape. These two aspects of my existence as a self need not be incompatible. By conceptualizing and shaping the ways I go about my particular projects in terms of the broader understanding I develop through perpetual openness, I am able to be a specific self while still being philosophically autonomous. We must not deny or downplay the particularities of our limitations, but take them up and make something meaningful out of them. In practice, governing the self means *taking responsibility for* one's self: *both* in the sense that I take up the project of perpetual openness and try to overcome my meta-passivities, *and* in the sense that I take up my specific limitations and live within them in a way that is sensitive to my ever-growing understanding of the world.

Conclusion

Many accounts found within the literature have elements of governing the self. Some were explicitly concerned with something this kind of fullest autonomy. However, none were able to provide the thorough-going control that governing the self requires. If my account works, it fills in this theoretical space, albeit in an unexpected way. Being philosophically autonomous requires consciously committing myself to pursuing philosophical autonomy. Since there is no pure space from which I can decide the self I am going to have, no Archimedean point from which to be fully in control of myself, I have to take up the ongoing project of learning about the ways I have been outsourcing this decision to external influences. This requires open-ended engagement with the world. I can only govern myself by going beyond myself.

I have been inspired by insights and themes from others in the literature: Frankfurt's basic notion of caring about the self I have; Meyers' idea of autonomy as based on skills and ongoing practice; Seidman's and feminists' ideas that the self is not just an agent, but a subject. In terms of the taxonomy of autonomy views, the account of governing the self I have described has affinity with both independent procedure accounts and substantive accounts. Regarding the former, it clearly demands active reflection from the agent and requires this reflection meet certain standards. It does not give a specific procedure to follow, but emphasizes a set of skills to be exercised in pursuit of governing the self: chief among them, openness to new perspectives, sensitivity to what these perspectives may have to teach, willingness to face the discomfort of learning and changing. This makes my account most akin to Meyers' competency view.

But since my view asks us to modify ourselves in ways that are responsive to ethical truth (in the broad sense I have been using), it also seems to be a strongly substantive view – more specifically, a reality-tracking view. As discussed in chapter 3, such accounts face a serious problem: they require that a person be able to track reality and therefore take it for granted that *we* already know what reality is. (Put differently: how could we apply the standard to determine if someone is autonomous if we do not know the standard?) My view of governing the self essentially sidesteps this problem (or tries to) by making *discovering this standard* (and *then* living in accordance with it) the exact goal of the person who strives for philosophical autonomy. This puts my account in an odd liminal space. It does not directly limit the content which is acceptable for philosophical autonomy, as a separate requirement that must be met in addition to following a certain procedure; rather, it incorporates this limit into the very procedure itself. By undergoing the procedure (i.e., practicing perpetual openness for a period of time), we come to fill out more and more of this content.

My account is primarily focused on shaping the self, or what I take to be *self-governance* proper; I have said little about acting in the world, or *self-direction*, except that we should always view our current selves as provisional, as the best current answer we have to the question of the self. Does this mean we are not autonomous when we act based on the current, provisional self? I said in chapter 4 that we should properly speak not of *being* (philosophically) autonomous, but of *practicing* (philosophical) autonomy. We can therefore think in terms of degrees. As our provisional self becomes more accurate, nuanced, and full, we are able to govern ourselves more, and therefore we can act in more

philosophically autonomous ways (insofar as we have successfully translated our self-governance into self-direction).

But I think that the need to determine how philosophically autonomous people are from a third-person perspective will be rather rare. Philosophical autonomy is a rather intimate project since it involves doing deep work on yourself. It is also a rather relational project, since it requires interacting with others and allowing them to impact you. Finally, it is an *ongoing* project; you can never complacently assume you have become fully philosophically autonomous. This means that insofar as other people do provide feedback on how philosophically autonomous you are, the point is less to judge you for the purposes of praise and blame (attitudes which correspond to moral responsibility, and not autonomy). Rather, the point is to give you feedback to help you get better at being autonomous. Because it is rather intimate, the ideal space for such pointed feedback is a close friendship.

Philosophical autonomy is a particularly exacting kind of self-governance. It likely seems far removed from what most people are concerned with when they speak of autonomy. It's undeniable that this fullest kind of autonomy takes us beyond more common conceptions of autonomy: the minimal capacity for self-governance presumed of all typical adults, the political autonomy which is meant to protect the expression of this minimal autonomy, and the fuller kind of (typically authenticity-centered) autonomy which articulates what it means to be not just externally free of coercion, but internally free to pursue the kind of life one finds most fulfilling. But while philosophical autonomy *is* quite different, I believe we should care about it because it has potential to be an important tool

modern liberal society. I would like to end by sketching out these hopes, which need further investigation to be properly evaluated.

Despite obvious differences, philosophical autonomy is on a continuum with more common notions of autonomy. Political autonomy and authenticity-centered autonomy are both primarily concerned with freedom – the first with external freedom, the second with internal freedom. Philosophical autonomy is concerned with the deepest freedom: it wants to ensure that who I am is ultimately controlled by me, and not outsourced to something pre-given and beyond my control. Likely, anxiety about this level of freedom is rare; nonetheless, it is the logical extension of the kind of freedom that ordinary notions of autonomy are concerned with. If we could convince people to care about philosophical autonomy – already a big "if" – and to start the project of governing their selves, then we may be partway to resolving a big difficulty we face today: the viability of liberal democracy.

I will limit the following comments to America, since this is the society I know most about. First, a note on the terms. Liberalism is political philosophy that takes ensuring individual freedom to be paramount. Democracy is rule by the people. There are clear connections between the two: individual freedom seems to require that I not be subject to an arbitrary sovereign and can only legitimately be governed by something I give consent to. This, combined with liberal ideas of equality, naturally leads to democracy. Furthermore, liberalism's emphasis on individual freedom pushes us to tolerate differences: I let you live your life, and you let me live mine. This allows for a diverse group of people to live together in peace – or so the hope is. In this way, liberalism provides

a foundation for democracy, since it makes space for public life amongst different people. But the two can come apart. A democracy could potentially violate the rights of certain individuals if the majority decided to do so and there were no laws in place to protect such rights; an autocrat could rule such that individual rights were protected, even though the people had no say in governing. Nonetheless, the two have close ideological roots – chiefly, equality and freedom from domination – which makes them natural and desirable to combine.

Liberalism has come under scrutiny. The trouble comes from various directions, all of which touch on autonomy in some way. Some theorists point to the shallowness of a society whose public life lacks the shared values and projects of traditional societies[148]. Some conservative or reactive groups believe that personal freedom has gone too far and is destroying other, traditional values. Some progressive groups see liberalism as too individualistic to be able to cope with structural problems, such as systemic racism and other forms of oppression which operate on group identities.

To these concerns I would add the observation that many people have become focused on their rights and freedoms almost to the exclusion of a sense of responsibility. This tension between freedom and responsibility has always existed within liberalism. Insofar as liberalism is connected to democracy, and each citizen has (ideally) the ability to influence how society is governed, each person has a responsibility to participate in a well-informed, well-considered way. (Of course, these points relate to many issues I cannot address here: voter suppression, the vast and vastly influential sums of money poured into

[148] E.g., Macintyre's *After Virtue* and Taylor's *Hegel and Modern Society*

politics by corporate interests and the wealthy, under-funded schools, rampant disinformation, the disillusionment of American voters, and consistently low voter turnout – to name a few.) But insofar as we are concerned primarily with external coercion of the individual, we will focus on ensuring each person's rights. This places only a negative requirement on how people relate to each other, since it makes non-interference and non-harm primary. With nothing to replace traditional values liberalism can encourage relating to one another in a utilitarian way, cooperating with each only when it is mutually beneficial[149].

All these worries reveal a common theme: the possibility that liberalism cannot provide a solid foundation for democracy because it cannot provide common ground for the demos. Each criticism shows a division in the demos which undermines the ability to talk and work together for common purpose, and each division is something liberalism either cannot help with or makes worse. With no shared values or projects, each person is on their own to pursue "the good life" (or frankly, simply the means to survive). The focus on individual rights to the exclusion of responsibility is thus exacerbated, meaning people are less motivated to find common cause and purpose. There *are* systemic issues which work on group identities, and liberalism with its focus on individual rights does not give us the resources to either conceptualize or make progress on these issues. This leaves "social justice activists" at odds with people who think individualistically, as liberalism encourages us to do. Those who long for traditional values to the point of wishing to impose

[149] This is especially true as citizens come to be disillusioned with their ability to meaningfully influence government.

this vision on others go against core liberal ideas. This is not itself a shortcoming of liberalism: what *is* a problem is that liberalism, with its merely negative value of protecting individual freedom, does not have a substantial counter-proposal such people might find compelling. Liberal democracy, in short, has the potential to cause its own unraveling.

The hope is that philosophical autonomy could be fertile ground for working together to resolve these issues. First of all, by emphasizing that the fullest form of freedom is not simply the freedom to do what I want, or even the freedom to do what I "really, authentically" want, but is the freedom to shape the self that I have even beyond what I was originally given, philosophical autonomy starts to bridge the gap between freedom and responsibility. Since being responsible for myself requires me to be responsive to other perspectives, it forces me to start looking outside of myself, and to see this engagement as an essential part of my freedom. If the third aspect of my view – taking the world seriously such that I want to do right by it – works, then I will take on the full project of responsible citizenship, and do so with the goal of recognizing the value of other forms of life, not simply tolerating them.

Since philosophical autonomy is an expansive project which requires me to learn to appreciate as many forms of value and meaning as possible, it may also lay the groundwork for a common basis for public life which is neither empty nor composed of isolated factions. It does not ask us to abstract from substantive values or to adopt a value-neutral standpoint; rather, it asks us to learn to appreciate as many different kinds of meaning as possible, to incorporate these insights into our own lives where possible, and to harmonize with those we cannot directly engage with. Even if there are different

communities, each engaged with their own "form of life" – a possibility discussed in chapter 5, problem #5 – there can still be a common basis of mutual respect which allows for genuine dialogue and support, not just competing interests.

And since philosophical autonomy does not want us to get rid of original forms of meaning, but to preserve and better appreciate what is valuable about them, there is a conservative *aspect* to it. This does not mean that those who want to return wholescale to an earlier society with stricter, more defined societal roles are entirely correct. They, too, in working on being philosophically autonomous should strive to articulate what is valuable about the "forms of life" they are concerned about losing, while being willing to critique aspects which are problematic because they deny full selfhood or freedom to others. (It would probably help if more "liberal" people took seriously the idea that there is something of value in traditional life forms – something which a person striving for philosophical autonomy would need to do.)

Finally, philosophical autonomy has affinity with critical theory insofar as it takes seriously the ways people are deeply influenced by ideology and social systems. This puts it in a better position to understand systemic issues, a progressive complaint against classic forms of liberalism. It also recognizes the existence of societally imposed, structural limitations (problem #4 of Chapter 5), and makes room for the possibility of affirming one's particular culture or contingent identity (problem #5 of Chapter 5). If the third aspect of my view – taking the world seriously and wanting to do right by it – holds, it even gives us a direct reason to work for social justice.

Simply convincing individuals to take up the project of governing their selves – which, as I have argued, is equivalent to taking responsibility for their selves – will not be enough to solve all the troubles of liberalism. The nature of today's biggest problems, from polarization of wealth to climate change, is systemic and therefore requires systemic solutions. If more individuals build up a broader consciousness, they may be more motivated to collective action to effect systemic change, but this does not change the fact that the solutions can only be changes to the system; changing individuals is not enough by itself.

Nonetheless, the individual is still significant, especially if we continue to believe that democracy is the best political system. The core idea of liberalism remains worth fighting for: the individual person matters, not just because she can feel pain and pleasure like any other creature, but precisely *because* she has the unique freedom which gives humans depth and makes them proper subjects of dignity – because, in short, she has basic autonomy. None of what I have said above is meant to undermine the protection of basic autonomy. Such rights must continue to be upheld as part of respecting the individual. My point is rather that if we could get more people to see that the fullest expression of this autonomy is philosophical autonomy, this might help with the challenges liberalism faces.

The argument I have sketched out can be summed up thusly: Liberal democracy is based on solid principles of equality and basic autonomy. But these principles in themselves are rather empty, and do not provide a substantial enough common ground for the demos. This means that the demos becomes fragmented into competing factions and disinterested (or disillusioned) individuals. Liberal democracy thus has the seeds of its own

unraveling. But the fullest expression of basic autonomy is philosophical autonomy. And philosophical autonomy has the potential to bring citizens into dialogue with each other, and to come to mutual respect and understanding of the different kinds of meaning the demos can support, and to see themselves as working, if not together, then at least in harmony, and as having the "meta-goal" of preserving this harmony (giving them at least one shared project). Philosophical autonomy can thus be the fruition and the redemption of liberal democracy.

Something like this argument is the hope. All this dissertation has attempted is to articulate one piece of the puzzle: what philosophical autonomy involves. Even at this initial step, questions remain. For instance, how do we actually implement changes into ourselves? Once we have "governed" in the sense of legislating who we will be, how do we execute these changes – especially if they are big changes and hard to implement? There is still work to be done, but the humbler goal of this project was to demonstrate that philosophical autonomy is promising ground for further investigation.

Bibliography

Anderson, Joel & Honneth, Axel. “Autonomy, Vulnerability, Recognition, and Justice”. *Autonomy and the Challenges to Liberalism: New Essays*, ed. by John Christman and Joel Anderson. Cambridge University Press, 2005.

Anderson, Joel. “Autonomy and the Authority of Personal Commitments: From Internal Coherence to Social Normativity” *Philosophical Explorations,* vol 6.2, 2003, pp 90-108.

Anderson, Joel. “Disputing Autonomy: Second-Order Desires and the Dynamics of Ascribing Autonomy”. *SATS,* vol 9.1, 2008, pp 7-26.

Arpaly, N., 2004, “Which Autonomy?” in J. Campbell, M. O’Rourke, and D. Shier (eds.), *Freedom and Determinism*, Cambridge, MA: MIT Press, 173–88.

Arpaly, Nomy and Schroeder, Timothy (1999) “ Praise, Blame and the Whole Self,” *Philosophical Studies* 93 : 161–88 .

Arpaly, Nomy and Schroeder, Timothy (1999). “Alienation and Externality”. *Canadian Journal of Philosophy* 29 (3):371-387.

Barclay, Linda, 2000, “Autonomy and the Social Self,” from *Relational Autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self*, New York: Oxford University Press. 2000.

Benson, Paul, 2005, “Taking Ownership: Authority and Voice in Autonomous Agency,” *Autonomy and the Challenges to Liberalism: New Essays*, ed. by John Christman and Joel Anderson. Cambridge University Press, 2005.

Benson, Paul. “Autonomy and Oppressive Socialization.” *The Journal of Philosophy* , Dec., 1994, Vol. 91, No. 12. JSTOR, https://www.jstor.org/stable/2940760.

Benson, Paul. “Free Agency and Self-Worth.” *Social Theory and Practice*, vol. 17, no. 3, 1991, pp. 385–408. JSTOR, www.jstor.org/stable/23557430. Accessed 10 June 2021.

Berofsky, Bernard. “Identification, the Self, and Autonomy”. *Social Philosophy and Policy,* vol 20.2, 2003, pp 199-220.

Bratman, Michael E. “Autonomy and Hierarchy”. *Social Philosophy and Policy*, vol 20.2, 2003, pp 156-176.

Bratman, Michael E. “Identification, Decision, and Treating as a Reason”. *Philosophical Topics,* vol. 24.2, 1996, pp 1-18.

Bratman, Michael E. “Planning Agency, Autonomous Agency.” *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, ed. by James Stacey Taylor, Cambridge University Press, Cambridge, 2005, pp. 33–57.

Bratman, Michael E. “Reflection, Planning, and Temporally Extended Agency”. *Philosophical Review,* vol 109.1, 2000, pp 35-61.

Bratman, Michael E. (2004). “Three Theories of Self-Governance”. *Philosophical Topics* 32 (1/2):21-46.

Buss, Sarah. “Autonomy Reconsidered”. *Midwest Studies in Philosophy,* vol 19.1, 1994, pp 95-121.

Carter, Alan. “Morality and Freedom”. *Philosophical Quarterly, vol* 53 (211), 2003, pp 161-180.

Christman, John. “Autonomy and Personal History”. *Canadian Journal of Philosophy*, vol 21.1, 1991, pp 1-24.

Christman, John. “Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves”. *Philosophical Studies*, vol 117 (1-2), 2004, pp 143-164.

Code, Lorraine, 1987, “Second Persons.” *Canadian Journal of Philosophy*, Supplementary Volume 13:357

Cooke, Maeve. “Questioning Autonomy: The Feminist Challenge and the Challenge for Feminism”. *Questioning Ethics: Contemporary Debates in Philosophy*, ed. by Richard Kearney and Mark Dooley. Routledge, 1999, pp 258-282.

Darwall, Stephen (2006). “The value of autonomy and autonomy of the will”. *Ethics* 116 (2):263-284.

Dillon, Robin S. “Toward a Feminist Conception of Self-Respect.” *Hypatia*, vol. 7, no. 1, 1992, pp. 52–69. JSTOR, www.jstor.org/stable/3810133. Accessed 21 July 2021.

Dworkin, Gerald. “The Concept of Autonomy”. *Grazer Philosophische Studien,* vol 12.1, 1981, pp 203-213.

Ekstrom, Laura Waddell. “A Coherence Theory of Autonomy”. *Philosophy and Phenomenological Research,* vol 53.3, 1993, pp 599-616.

Feinberg, Joel. "Autonomy." *The Inner Citadel: Essays on Individual Autonomy*, ed. by John Christman. Oxford University Press, 1989, pp 27-53.

Feinberg, Joel. “Abortion”. *Matters of Life and Death*, 2d ed., ed. by Tom Regan. McGraw-Hill, 1986.

Ferrero, Luca. “Constitutivism and the Inescapability of Agency”. *Oxford Studies in Metaethics,* vol 4, 2009, pp 303-333.

Fischer, J. and Ravizza, M., editors. *Perspectives on Moral Responsibility*. 1993.

Fischer, John Martin & Ravizza, Mark. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press, 1998.

Fischer, John Martin. “Responsibility and Autonomy: The Problem of Mission Creep”. *Philosophical Issues,* vol 22.1, 2012, pp 165-184.

Fischer, John Martin. “The Cards that are Dealt You”. *The Journal of Ethics,* vol 10 (1-2), 2005, pp 107-129.

Frankfurt, Harry G. “Autonomy, Necessity, and Love”. *Necessity, Volition, and Love*. Cambridge University Press, 2009, pp 129-141.

Frankfurt, Harry G. “Caring and Necessity”. *Necessity, Volition, and Love*. Cambridge University Press, 2009, pp 155-167.

Frankfurt, Harry G. “Freedom of the Will and the Concept of a Person”. *Journal of Philosophy*, vol. 68.1, 1971, pp 5-20.

Frankfurt, Harry G. “Identification and Externality”. *The Identities of Persons*, ed. by Amelie Rorty. University of California Press, 1977, pp 239-253.

Frankfurt, Harry G. “The Faintest Passion”. *Necessity, Volition, and Love*. Cambridge University Press, 2009, pp 95-107.

Frankfurt, Harry G. *Taking Ourselves Seriously & Getting It Right*. Stanford University Press, 2006.

Frankfurt, Harry. “Identification and Wholeheartedness”. *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, ed. by Ferdinand David Schoeman. Cambridge University Press, 1987, pp 27-45.

Friedman, Marilyn A. “Autonomy and the Split-Level Self”. *Southern Journal of Philosophy*, vol 24.1, 1986, pp 19-35.

Friedman, Marilyn. “Autonomy and Social Relationships: Rethinking the Feminist Critique”. *Feminists Rethink the Self*, ed. by Diana T. Meyers. Westview Press, 1997, pp. 40--61.

Gorman, August. “The Minimal Approval View of Attributability”. *Oxford Studies in Agency and Responsibility 6,* ed. by David Shoemaker. Oxford University Press, 2019, pp 140-164.

Govier, Trudy. “Self-Trust, Autonomy, and Self-Esteem.” *Hypatia*, vol. 8, no. 1, 1993, pp. 99–120. JSTOR, www.jstor.org/stable/3810303. Accessed 21 July 2021.

Grimshaw, Jean, 1988, “Autonomy and Identity in Feminist Thinking”. *Feminist Perspectives in Philosophy*, ed. Griffiths and Whitford. Palgrave Macmillan, London.

Helm, Bennett W. “Love, Identification, and the Emotions”. *American Philosophical Quarterly,* vol 46.1, 2009, pp 39-59.

Hill, Thomas E. “The Kantian Conception of Autonomy”. *The Inner Citadel: Essays on Individual Autonomy*, ed. by John Christman. Oxford University Press, 1989, pp 91-105.

Hyun, I., 2001, “Authentic Values and Individual Autonomy,” Journal of Value Inquiry, 35(2): 195–208.

Jaworska, Agnieszka. “Caring and Internality”. *Philosophy and Phenomenological Research,* vol 74.3, 2007, pp 529-568.

Jaworska, A. 2009, “Caring, Minimal Autonomy, and the Limits of Liberalism,” in H. Lindemann M. Verkerk, and M. Walker (eds.), Naturalized Bioethics: Toward Responsible Knowing and Practice, Cambridge: Cambridge University Press, 80–105.

Jaworska, Agnieszka. “Identificationist Views”. *The Routledge Companion to Free Will*, ed. by Kevin Timpe, Meghan Griffith, and Neil Levy. Routledge, 2016.

Jaworska, Agnieszka. “Respecting the Margins of Agency: Alzheimer's Patients and the Capacity to Value”. *Philosophy and Public Affairs*, vol. 28.2, 1999, pp 105-138.

Katsafanas, Paul. “Constitutivism about Practical Reasons”. *The Oxford Handbook of Reasons and Normativity,* ed. by Daniel Star. Oxford University Press, 2018, pp. 367-394.

Katsafanas, Paul. *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford University Press UK, 2013.

Kittay, Eva Feder. “At the Margins of Moral Personhood”. *Ethics*, vol 116, No. 1, Symposium on Disability, 2005, pp. 100-131

Korsgaard, Christine M. "Morality as freedom." *Kant’s Practical Philosophy Reconsidered*, ed. by Yirmiyahu Yovel. Springer Science and Business Media Dordrecht, 1989, pp 23-48.

Korsgaard, Christine M. *Sources of Normativity.* Cambridge University Press, 1996.

LeBar, M., 2005, “Eudaimonist Autonomy,” *American Philosophical Quarterly*, 42(3): 171–83.

Mackenzie, Catriona. "Three Dimensions of Autonomy: A Relational Analysis". *Autonomy, Oppression and Gender*, ed. Andrea Veltman and Mark Piper. Oxford University Press, 2014, pp 15-41.

Mackenzie, C. and N. Stoljar “Introduction: Refiguring Autonomy,” from *Relational Autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self*, New York: Oxford University Press. 2000.

McKenna, M., 2013, “Reasons-Responsiveness, Agents, and Mechanisms.” *Oxford Studies in Agency and Responsibility*, 1: 151–183.

McMahan, Jeff. “Identity”. *The Ethics of Killing: Problems at the Margins of Life,* Oxford University Press, 2002

Mele, Alfred. “History and Personal Autonomy”. *Canadian Journal of Philosophy,* vol 23.2, 1993, pp 271-280.

Meyers, D., 1987, “Personal Autonomy and the Paradox of Feminine Socialization,” *Journal of Philosophy*, 84: 619–28.

Nedelsky, Jennifer. "Reconceiving autonomy: Sources, thoughts and possibilities." *Yale JL & Feminism* 1 (1989): 7.

Nehamas, Alexander. “How One Becomes What One Is”. *Nietzsche: Life as Literature*. Harvard University Press. 1985.

Nelkin, Dana K. “Two Standpoints and the Belief in Freedom”. *Journal of Philosophy*, vol 97 (10), 2000, 564-576.

Oshana, Marina A. L. “Personal Autonomy and Society”. *Journal of Social Philosophy*, vol 29.1, 1998, pp 81-102.

Oshana, Marina A. L. “The Misguided Marriage of Responsibility and Autonomy”. *The Journal of Ethics,* vol 6.3, 2002, pp 261-280.

Roughley, Neil. “The Uses of Hierarchy: Autonomy and Valuing”. *Philosophical Explorations*, vol 5.3, 2002, pp 67-185.

Seidman, Jeffrey. “Valuing and caring”. *Theoria,* vol 75.4, 2009, pp 272-303.

Shoemaker , D. (2003) “ Caring, Identification, and Agency ,” *Ethics* 114 : 88–118

Smith, Michael. “Constitutivism”. *The Routledge Handbook of Metaethics,* ed. by Tristram McPherson and David Plunkett. Routledge, 2017, pp. 371-384.

Smith , A. (2004) “ Conflicting Attitudes, Moral Agency, and Conceptions of the Self ,” *Philosophical Topics* 32 : 331–52 .

Sripada, Chandra. “Moral Responsibility, Reasons, and the Self”. *Oxford Studies in Agency and Responsibility: Volume 3,* ed. by David Shoemaker. Oxford University Press, 2015, pp 242-264.

Stoljar, N., 2000, “Autonomy and the Feminist Intuition,” from *Relational Autonomy Feminist Perspectives on Autonomy, Agency and the Social Self*, New York: Oxford University Press. 2000.

Strawson, Galen. “Against Narrativity”. *Ratio,* vol 17.4, 2004, pp 428-452.

Strawson, Peter. “Freedom and Resentment”. *Proceedings of the British Academy, vol 48, 1962,* pp. 1-25.

Taylor, Charles. “Atomism”. *Powers, Possessions and Freedom: Essays in Honour of C.B. Macpherson,* ed. by Alkis Kontos. University of Toronto Press, 1979.

Taylor, Charles. “Responsibility for Self”. *The Identities of Persons,* ed. by Amelie Oksenberg Rorty. University of California Press, 1976, pp. 281-99.

Taylor, James Stacey. “Introduction”. *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy,* ed. by James Stacey Taylor. Cambridge University Press, 2005, pp. 1-29.

Thalberg, Irving. “Hierarchical Analyses of Unfree Action”. *Canadian Journal of Philosophy*, vol. 8.2, 1978, pp 211-226.

Velleman, J. David. “What Happens When Someone Acts?”. *Mind,* vol. 101 (403), 1992, pp 461-481.

Velleman, J. David. “Introduction”. *The Possibility of Practical Reason*. Oxford University Press, 2000, pp 1-31.

Watson, Gary. “Free Agency”. *Journal of Philosophy*, vol. 72, 1975, pp 205-20.

Watson, Gary. “Two Faces of Responsibility”. *Philosophical Topics,* vol 24.2, 1996, pp 227-248.

Watson, Gary. “Volitional Necessities”. Agency and Answerability: Selected Essays. Oxford University Press. Oxford: 2004.

Wolf, Susan. “Sanity and the Metaphysics of Responsibility”. *The Inner Citadel: Essays on Individual Autonomy,* ed. by John Christman. Oxford University Press, 1989, pp 123-136.

Wolf, Susan. *Freedom within Reason*. Oxford University Press, 1993.

Young, Robert. “Autonomy and the Inner Self”. *American Philosophical Quarterly,* vol 17.1, 1980, pp 35-43.

Appendix A

Why were moral responsibility and autonomy easily conflated?

(Note: Here I use "autonomy" in a broad sense, and do not mean "philosophical autonomy".)

Moral responsibility, autonomy, internality, and identification are commonly intertwined in the literature. Both internality and identification have been thought to be plausible candidates for grounding both moral responsibility and autonomy. This has meant that someone who is working on an account of autonomy (for instance) who wants to consider the role either internality or identification could play in autonomy will run into people using these concepts in accounts of moral responsibility. In other words, these two branches have both been hypothesized as having the same roots, and we sometimes forget that they are nevertheless two different branches.

Why was identification hypothesized as central to both moral responsibility and autonomy such that these two ostensibly different concepts became blurred? Such intelligent and dedicated people as write about these topics did not simply get confused about what the word applied to; the deeper reason for the confusion is that identification was originally defined in a way that made it sound remarkably close to a conception of autonomy, even though it was presented in an account that was explicitly about moral responsibility. I am referring once again to Frankfurt's "Freedom of the Will and the Concept of a Person".

Frankfurt's initial notion of identification implied conscious endorsement of a desire – not necessarily deliberate and reflective endorsement, although he moved on to a

view like this in his paper "Identification and Wholeheartedness". This was unsurprising since identification naturally lends itself to reflection and deliberation. The notion of second order volitions which are about first order desires indicates reflection on those first order desires. Once this space is opened up, there is room for deliberation. Whether there is deliberation or not, the idea of deciding to endorse a first order desire indicates a kind of self-directed activity (a point Frankfurt explicitly makes in "Identification and Wholeheartedness".

From the beginning, therefore, identification strongly suggested a process of consciously reflecting on desires that you have, reasoning about which one you want to act on, and then deciding to commit yourself to this desire. This sounds a lot like self-governance: you do not simply accept your motivations, but step back from them and ask yourself "what do I *really* want?", and then you take a stand – a stand which thus seems to be grounded in and defining of your self. It sounds so close to autonomy, in fact, that one might wonder why Frankfurt originally intended it to ground an account of moral responsibility. The answer points us towards the deepest reason why moral responsibility and autonomy were confused.

Frankfurt's concern in this initial paper was with what makes us *persons*. He suggested a hierarchical account of the will as an answer to this specific question. His intuition for making such hierarchy essential to personhood appears to be twofold. First of all, hierarchy emphasizes what seems to be the central ability of persons to reflect on themselves, or at least to relate to their desires and motivations in a complex way. This is intuitively at least part of what separates persons from most other animals: not only is it a

rather unique ability, but it is an ability which radically alters everything else about us. But Frankfurt is explicit that reason is not sufficient for personhood, though it is almost certainly necessary, and his main concern is more interesting. He believes that what makes us persons is that we *care* about the shape and character of our will. This is why persons form second order volitions at all: a person wants it to be the case that a certain desire (or set of desires, or kind of desire) is what constitutes his will because he cares that his will be a certain way.

The way I read Frankfurt, what this comes down to is the idea that what is defining of a person is that *she cares about the self she has*[150]. Why else would she care that her will be one way or another? A wanton, which Frankfurt contrasts with a person, *is* such a creature who does not care. She may be a complex reasoner, capable of planning how to get what she desires, or how to fulfill the maximum number of her desires as possible; but when it comes down to it, she simply does not care what the shape and character of her will is. She never asks the question "*Which* of my desires do I actually want to act on?", but takes them all as equal. She may end up acting on the desire to do cocaine, or her desire to buy out the entire stock of Casey's Cupcakes, or her desire to stay home and do crafts with her young child, but ultimately, she simply doesn't care which one of these desires wins out. It's all the same to her. A person is therefore a creature that cares about the sort

[150] To clarify: the way I understand Frankfurt, it is indeed the person's *actual state* of caring, and not just the capacity to care. A wanton is fundamentally someone who simply does not care about the will he has. If he were to start caring, then he would become a person; but he must *actually* care, and not simply have the capacity to care.

of *self* she has[151]. This is why I care about my will: I want to have a certain kind of will because this is what gives me a particular character.

To be a person is to care about my will; and to care about my will is to care about my *the kind of person I am*. This means that being a person just *is* caring about the kind of person I am. Notice that this definition of personhood has reflexivity built into it such that it sounds remarkably like self-governance: to be a person is to care about the self or person I am such that I *make decisions* about the kind of self I want to be. In other words, Frankfurt made the very definition of personhood reflexive such that it implies autonomy. The point I wish to make is that this was not just a coincidence: it *made sense* for Frankfurt to build such reflexivity into personhood, and he was onto something deep about the relationship between personhood, agency, moral responsibility, and autonomy – something which explains why these concepts can be so easily entangled.

[151] It might be tempting to read the contrast between the person and the wanton as rooted in the fact that for the person, some things have importance in a way they don't for the wanton, and that this is why the person cares that her will be constituted by only certain desires – she wants to act on the things which are important to her. In other words, one might be skeptical that we need to understand Frankfurt's conception of personhood as essentially self-focused; it could be world-focused. This might seem like a plausible way to understand the examples, but it does not capture Frankfurt's main intuition. The idea of importance does not necessarily lead to the hierarchy of desires which he emphasizes. If certain things are important to me, my deciding what to do – what to will – will not primarily be put in terms of a question of which of my desires I will want to act on. It will mainly be a ground level question about the world, and not one which is primarily focused on my inner states and inclinations. (This is, of course, similar to a familiar point that Watson made in his essay "Free Agency".) In contrast, Frankfurt's focus on hierarchy and second order volitions indicates that the person is in some way concerned with her self, and not just unreflexively concerned with things that are important to her. Persons care about the kind of will they have, and this is a kind of concern which cannot be captured without reference to the self.

Roughly, we think that *persons* have a special kind of *agency* such that they are *morally responsible* for their actions and are at least capable of becoming *autonomous*. It is intuitively plausible that this special kind of agency is the ability to distance themselves from things like instincts and inclinations, and to deliberate about what to do. To act on instinct is to be simply and directly determined by inner processes and external stimuli. It is therefore to a large degree passive, despite the fact that the individual organism is acting. But when persons deliberate and chose, they are no longer simply and directly determined; they more actively participate in producing their actions.

Once again, this sounds like some form of self-governance. A person has a multitude of desires and inclinations floating around inside of her, which are in some undeniable sense *hers*, but when she imposes a structure on them such that some are given priority and others are completely discounted, this makes is the case that when she acts she is no longer simply determined by external stimuli. The stimuli work through her internal structure such that the actions she takes is also a result of her. It may be that most of these structural decisions do not happen consciously and actively, but nevertheless it seems that in all persons there *is* such a structure. This structure shapes to our interpretive, judgmental, emotional/affective, and volitional patterns – in other words, our internality. Assuming that most of the time, this structure is not consciously decided on, this is "self-governance" in only a very attenuated form. Nevertheless, it is still *like* self-governance in that we are not simply acting on the desires or instincts of the moment, but acting in accordance with a deeper structure. We may not give ourselves this structure, but we nevertheless govern what we do based on this structure, and so are more active.

In sum, persons seem to have a more robust form of control over their actions than other animals, and this is intuitively what makes them persons. But this form of control is plausibly thought of as a minimal form of self-governance. Given that we think it is some unique form of agency which humans have which makes them morally responsible, and given that this unique kind of agency seems to be a form of self-governance, and, finally, given that autonomy is thought of as self-governance, it makes sense that moral responsibility and autonomy would be easily confused and conflated.

I believe that this idea that personhood necessarily involves a minimal form of self-governance is quite a plausible picture, and I think that Frankfurt was not completely misguided in in making self-governance defining of personhood. It is not a coincidence that both personhood and autonomy involve self-governance. What we need is to be careful to separate the minimal form of self-governance involved in personhood from the more full-fledged form involved in autonomy.