Your Self is Deeper Than You Think: A Deep Self View of Moral Responsibility

YOUR SELF IS DEEPER THAN YOU THINK:
A DEEP SELF VIEW OF MORAL RESPONSIBILITY

by

Ke Zhang

_____

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF PHILOSOPHY

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2023

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by: Ke Zhang
titled:

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Michael McKenna
Michael McKenna (Mar 23, 2023 17:12 PDT)   Date: Mar 23, 2023
Michael McKenna

Carolina Sartorio
Carolina Sartorio (Mar 23, 2023 17:29 PDT)   Date: Mar 23, 2023
Carolina Sartorio

Mark Timmons
                          Date: Mar 23, 2023
Mark Timmons

                          Date: Mar 24, 2023
Jason Turner

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Michael McKenna
Michael McKenna (Mar 23, 2023 17:12 PDT)   Date: Mar 23, 2023
Michael McKenna
Philosophy

ACKNOWLEDGMENTS

First, I want to thank my parents, 孙亚玲 Yaling Sun and 张南 Nan Zhang. They never doubted my pursuits in higher education, especially philosophy, even though they know little to nothing about what I do. I wouldn't have been able to stay in the program for the first two and half years without their financial and emotional support. Their unconditional love, support, and belief in me is a privilege I have that a lot of women and gender minorities back home don't.

老爹和玲玲：我爱你们。

I'm especially grateful to my supervisor, Michael McKenna. I've said this to Michael before and I want to say it again here: Michael hand-walked me through the whole process of writing my dissertation and becoming the scholar I am today. I would have had a way tougher time, not only with my dissertation, but also as a graduate student, if it wasn't for his academic guidance and moral support. I thank Carolina Sartorio for all her critical comments, suggestions, and encouragements that helped improve my work and grow as a philosopher. I'm grateful to Jason Turner and Mark Timmons for joining my committee; the questions they raised for my work have since inspired me with new paper projects.

I learned a lot from all my teachers and mentors at Arizona, from whom I've taken classes and seminars over the years. For that I thank Julia Annas, Allen Buchanan, Stewart Cohen, Juan Comesaña, Terry Horgan, Michael McKenna, Mary Peterson, Marga Reimer, Carolina Sartorio, Mark Timmons, and Jason Turner. I'm grateful to Trevor Hedberg, Guido Pincione, Steven Wall, and Houston Smit for helping me prepare for the job market. I owe many thanks to Dave Schmidtz, who was the first to lend me a hand and lift me away from my financial hurdles, and continued to do so semester after semester.

Throughout the years I've received many helpful comments and suggestions on my work, advice and support on surviving grad school, from fellow graduate students. For that I'm grateful to Jacob Barrett, Jacob Blitz, Rhys Borchert, April Chan, Phoebe Chan, Caleb Dewey, Ding, Luke Golemon, Tim Kearl, Caroline King, Max F. Kramer, Robert Lazo, Elizabeth Levinson, Andrew Lichter, Joe Metz, Alex Motchoulski, Dave Poplar, Travis Quigley, Sarah Raskoff, Will

Schumacher, Lucia Schwarz, Anna Bella Sicilia, Lenin Vazquez-Toledo, Robert H. Wallace, Sean Whitton, Karolina Wisniewska, and 周奕李 Yili Zhou.

For more than a decade of friendship that is my emotional shelter and a source of my motivation, I thank 小路, 双双, 龙龙, 沈戴娟, 李晓, 小红:

我们在世界的不同角落生活，奋斗，一点点长成各自该长成的样子。这期间，我们没有忘记彼此，愿意与对方分享各自的生活，依旧能够欣赏、支持彼此，这是何等的人间美事。你们在的地方，就是我的根伸展去的地方。

My deepest thanks go to Sean Whitton and Yili Zhou.

Sean: You are my role model in so many ways. I would have done most of the things I did without you, but I wouldn't have achieved as much or done so as gracefully without you. It's strange that you are *that* much better at using chopsticks than I am, but I've given that up since I was ten. So……just let it be known that properly using chopsticks is not one of those ways in which I look up to you. Thank you for allowing me to be myself, and putting up with me for always having something to say about almost everything, however bold or unpopular my opinions can be. Many people shy away from it. It's rare that one is not only aware of their privileges but also open to be challenged. You are that, and for that I admire you. Oh, and thank you for not being too much of a reverse snob to my occasional snobbery because indeed I am that cool and artsy and all that jazz.

Yili: To me, we became really good friends the moment I shared with you a film I had recently watched and you responded with one of the standard criticisms to Kant's moral philosophy. It was the meeting of two great minds that are also extremely nerdy, still two great minds though, in my opinion. I was talking with you about a scene from Lars von Trier's *Nymphomaniac* (still one of my favorite films he has made), where the protagonist/antagonist publicly shames a man's inner struggle and desires with pedophilia that have never been manifested in real actions; her words shred him into pieces of ash that he wishes he had already been burned into. She makes him cry like a baby. And you said that she is like Kant and some people criticize him for making screening every little thought that goes through one's mind a moral requirement. That was nine years ago and we've been enjoying the freedom to talk about *every*thing since. Then we came here together, dreaming of a rosy future full of intellectual challenge and joy. Beyond what we had expected, we went through all the shit together. It's a

privilege for me to have been there with you and for you, to have fought together, grown together, cried together, and most importantly, to have been able to still laugh together. There's always laughter.

So, thank you both, for your unwavering love, the best company, and for being my moral backbone. Through both of you I'm reaffirmed over and over again that dignity, open-mindedness, a genuine care for justice, and faith in goodness and humanity, are some of the fundamental qualities that matter to us as moral agents, that matter to me—I wouldn't have written a dissertation that concerns moral matters if I had a corrupted heart. That'd be absurd.

They say what doesn't kill you makes you very, very, *very* weak and will probably get you next time (R.I.P. Norm Macdonald), but until next time, I'm proud to say that I've come out the other side with integrity, and of course, with this dissertation.

LAND ACKNOWLEDGMENT

We respectfully acknowledge the University of Arizona is on the land and territories of Indigenous peoples. Today, Arizona is home to 22 federally recognized tribes, with Tucson being home to the O'odham and the Yaqui. Committed to diversity and inclusion, the University strives to build sustainable relationships with sovereign Native Nations and Indigenous communities through education offerings, partnerships, and community service.

DEDICATION

To Sean and Yili

TABLE OF CONTENTS

ABSTRACT

This dissertation is a collection of standalone papers about a novel version of the deep self view of moral responsibility. Taken on its own, each chapter deals with a different thesis. But as the title of my dissertation reveals, taken together, the three chapters in it constitute the groundwork for my deep self view of moral responsibility. In Chapter 1, I develop and defend the thesis of responsibility for the deep self. In Chapter 2, I argue for a sufficient condition for responsibility for one's self that centers on the idea of aspiration. Drawing upon resources from the first two chapters I further develop and defend a thesis of responsibility for what one does and its downstream consequences in Chapter 3.

Here is a summary of my view. I argue that an agent acts freely and is morally responsible for what she does in the accountability sense only if she has a deep self for which she is responsible. How is one responsible for one's deep self? To be responsible for the deep self, one must have a history where one was afforded the unimpeded opportunity to develop and exercise the ability to shape one's own self. Exercise that ability in what way? I suggest that a critical way in which an agent shapes her own self is when she engages with various activities that I call *aspirational self-shaping*. Nevertheless, an agent need not exercise the ability to shape her self and thus engage with aspirational self-shaping every time she acts freely. Indeed, being responsible for what she does is consistent with her failing to exercise that ability when she acts freely and responsibly.

Standard deep self views in the literature say something much stronger. They contend that an agent acts freely and responsibly for what she does if and only if her actions or omissions issue from, and so express, her deep self. Counterexamples proliferate. By offering a necessary condition for accountability drawing upon resources from responsibility for the deep self, my view escapes counterexamples that standard views face, while retaining the core of the deep self view. Indeed, an agent may be blameworthy for her wrongdoing without it issuing from, and so expressing, her deep self. And yet, she must have a deep self for which she is responsible to be blameworthy for her wrongdoing. All of this is ultimately achieved by paying closer attention to the historical dimension of the deep self than other deep self proponents have.

INTRODUCTION

This dissertation is a collection of three papers that develop and defend a novel deep self view of moral responsibility. It aims to offer critical insight on the necessary condition under which our practical agents are responsible for their actions, omissions, and any downstream consequences. This necessary condition is developed in my thesis of responsibility for the deep self. So, in my view, an agent is responsible for what she does in the sense that she is blameworthy for her wrong-doing only if she is responsible for her deep self. Drawing upon resources from how our agents often shape themselves in an aspirational way, I further argue that both being responsible and taking responsibility for who they are require a fuller understanding of aspiration. Let me first offer some background of what I have established in this dissertation.

Deep self views remain among the most appealing contemporary theories of moral responsibility. In deep self theories, whether an agent acts freely and is morally responsible for what she does depends on whether her actions express who she *truly* is—namely, her deep self, in the sense that these views identify. This is insightful. Freedom understood in the traditional Hobbesian and Humean sense means that for an agent to be free is for her to be free of external constraints. However, this understanding invites a further question: is there a sense of constraint in which it arises from within her own psychological constitution? For example, without any external constraints, an agent who suffers from phobias or certain mental disorders may still not be free and responsible for what she does due to the internal constraints manifested in her psychological and mental state. Deep self views address this concern, and aim to account for those instances of internal constraints that are freedom and responsibility undermining. The key agenda has been to carve out a division within an agent's self to capture her deep self. Once this

work is done, they contend that an agent acts freely and is morally responsible for what she does if and only if her actions or omissions issue from, and so express, her deep self, however differently these deep self theorists identify it.

Continuing this endeavor, I defend a novel deep self view of moral responsibility in which an agent's deep self plays an essential role, albeit not the role it plays in other deep self theories. In my view, the deep self is partly constituted, and crucially expressed by one's deep valuing and values. To be responsible for what one does in the accountability sense, one must be responsible for one's deep self. This is the key feature of my view. To be responsible for one's deep self, in turn, requires that one has a history where one was afforded the unimpeded opportunity (e.g., a safe, stable, and healthy household to grow up in; access to proper education; access to affordable healthcare; just society, etc.) to develop and exercise the ability to shape their self. When exercising such an ability, one actively engages with activities such as evaluating their own motivation, examining their values and pursuing new ones, setting and pursuing ends and goals for themself. When one does these things and succeeds at achieving them, one is actively shaping one's current self to obtain a (slightly, moderately, or drastically) different self. I call this process *aspirational self-shaping*. Once one is responsible for one's deep self, one becomes a candidate for responsibility for what one does. And if one does something morally objectionable, one is blameworthy for it. But if one was never afforded an opportunity to become responsible for one's deep self, namely, to develop and exercise the ability to aspirational self-shaping, one is at least less responsible for what one does. If one does something morally objectionable, one's blameworthiness for it is at least diminished.

Although argued in a similar vein, standard deep self views say something much stronger. As I mentioned, according to any deep self view, when an agent acts freely, her action

issues from, and is an expression of who she truly is. And when she acts freely, she is then a candidate for being responsible for what she does, insofar as her action is an expression of her deep self. However, as we can see, this involves a very strong condition for responsibility, and counterexamples proliferate. For example, one can be reasonably motivated to question if acting from a deep self is sufficient for an agent's responsibility for her actions. Consider someone who developed their deep self in the most morally abhorrent environment—perhaps they went through powerful and thorough indoctrination. Regardless of how the deep self is understood in standard views, one may wonder if such an indoctrinated agent is, after all, responsible for actions that are indeed expressive of their deep self. Standard deep self views are committed to treating any such action as free and for which an agent is responsible. Second, one also has good reasons to question if acting from a deep self is necessary for an agent's responsibility for her actions. Consider someone who acts contrary to what they judge to be the most preferable—it is not expressive of who they are in the relevant regard. Let us just suppose that they are acting out of weakness of will. In a case like this, we want to say that they may still be responsible for their weak-willed action even if it does not issue from, and so express, who they truly are. Standard views are committed to treating any such action as not free and for which an agent is not responsible.

By weakening the condition for moral responsibility to offer only a necessary condition, my view provides a new way of advancing the deep self view while resisting these counterexamples instead. This is first achieved by developing two theses of moral responsibility. The first thesis I develop is a neglected topic about responsibility for one's deep self. The second thesis is about responsibility for what one does and any downstream consequences. My overarching goal in this dissertation is to argue for this following connection between these two

theses: responsibility for one's deep self is necessary for responsibility for what one does and all that one brings about.

I begin by defending my thesis regarding responsibility for one's deep self in Chapter 1 "Who You Are Matters to What You Do: Responsibility for the Deep Self." In it, I advance two novel claims regarding what it is for someone to be responsible for their deep self. First, an agent is responsible for her deep self if she has an actual deep self that is obtained through her aspiration to transform her previous deep self into a different deep self—an aspired deep self. Here, aspiration is understood as involving a positive conative state primarily towards acquiring, rejecting or retaining, and living up to one's values. Thus, an agent aspires to a different self primarily when she actively takes actions towards such value-engagements, and thus actively contributes to shaping her deep self.

But what if she does not aspire? Would she be off the hook? I then argue that a responsible agent must be able to aspire, even if she is not actually aspiring, given the time and opportunity to do so, and her not doing so is through her own making, not anything beyond her control. This leaves open the possibility that if further sufficient conditions for responsibility are not met, an agent who is able to aspire might still not be responsible for her deep self. For example, a severely brainwashed agent who was never afforded access to any moral knowledge would not be responsible for her self, even if she possesses the ability to aspire in the way I identify it.

As it is apparent, the idea of aspiration does a lot of the heaving lifting in both the sufficiency and the necessity claims I develop in Chapter 1. In Chapter 2 "Coming Out the Other Side, Responsible: Taking Inspiration from Aspiration," I further elaborate on the relationship

between aspiration and responsibility. Specifically, I argue for the claim that aspiration is sufficient for responsibility for the deep self. The motivation for developing and defending this claim stems from an intuitive idea about responsibility shared by different responsibility theorists: if an agent is actively engaged with making something, she is then responsible for it, assuming this active engagement is not a product of external manipulative influences. Particular to my argument in this chapter, I suggest that this intuitive idea holds true when the thing an agent is actively engaged with making is her own self.

I start out by inviting my readers to think about a process familiar to many of us, in which after we evaluate our current values, ends or goals, or elements in our motivation, we come to retain or reject extant ones, or acquire, and thus fulfill new ones. As a result, something about us in the relevant regard changes. I call this process *aspirational self-shaping*, and argue that when properly and realistically understood, aspirational self-shaping makes us responsible for who we are. In my view, to aspire is to form a higher-order intention that aims at something one judges or deems worth acquiring, retaining or rejecting, or reevaluating, and then acts from it. If one experiences a change of heart, one can go back and revise that intention. I call such a higher-order intention a *personal policy*. Thus, to aspire is to form, act from, or later revise a personal policy about that to which one aspires. Once one succeeds at it, in and after the first instance of doing so, one actively engages with the shaping of one's self, and becomes responsible for what one is.

Drawing upon resources from my first thesis regarding responsibility for the deep self, developed in Chapter 1, and the claim about aspiration and responsibility for the self in Chapter 2, I last offer my deep self view of moral responsibility in Chapter 3, "Bringing the Deep Self Back to the Racecourse: Rethinking Accountability and the Deep Self." I argue that an agent acts

freely, and is morally responsible for what she does, only if she has a deep self for which she is responsible. Here, what helps to explain her responsibility in both senses is the ability to aspire to a deep self primarily by way of various value-engagements. Do our responsible agents get this ability for free? No. A historical condition must be met. Specifically, to be responsible for the deep self one possesses, one must have a personal history where one was afforded the opportunity to develop and exercise the ability to aspire, in the way I identify in Chapter 2. Without this ability ever being developed or exercised, an agent is not responsible for her deep self. As a result, she is not responsible for what she does, whether or not her actions or omissions issue from, and so are expressive of, her deep self. And if she did get to develop and exercise that ability, and became responsible for her deep self, then when she acts freely and is morally responsible for what she does, she must retain this ability, and be able to draw upon resources from her deep self to modulate her behaviors. This, however, is consistent with her failing to exercise the ability to aspire when she acts freely and responsibly. So, different from standard deep self views, whether an agent acts freely and is morally responsibly is not determined by whether her actions issue from, and so express, her deep self. Rather, it depends on whether she has a deep self for which she is responsible, even if at the time of acting, her actions or omissions do not issue from, and thus are not expressive of that self.

Hopefully, by the end of this dissertation, I will have shown that my view offers the important starting points for a better alternative to extant deep self views.

CHAPTER 1

Who You are Matters to What You Do: Responsibility for the Deep Self

Abstract

The idea that *to be responsible for what you do, you must first be responsible for who you are* has long been shared by moral responsibility theorists from different camps. That said, no one in the literature has so far explicitly offered a sufficient or necessary condition for what makes us responsible for our selves. I take up this task in this paper, and provide two novel claims on responsibility for one's self—one is a sufficiency claim, another a distinct necessity claim—to answer the important question of what makes us responsible for our selves. The two are closely connected: the idea of aspiration plays the sufficient role, which further reveals the ability to aspire as the necessary ability for responsibility for the deep self.

Introduction

When we think about moral responsibility, we usually think about responsibility for our actions, omissions, and any downstream consequences. Beyond that, we are led to consider the freedom condition that determines whether we are responsible for these things. However, besides the focus on responsibility for what we do (or do not) and what we bring about, philosophers such as Sartre, Aristotle, Charles Taylor (1976), and Robert Kane (1999) have also taken up the issue of responsibility for our selves. But why should we care about this kind of responsibility? It is because the idea that *to be responsible for what you do, you must first be responsible for who you are* has long been shared by moral responsibility theorists from different camps (Wolf, 1990; Kane, 1999; Fischer and Ravizza, 1998; and Ishtiyaque Haji, 1998), though no one in the

literature has so far explicitly offered a sufficient or necessary condition for what makes us responsible for our selves. I take up this task.

In this paper, I advance two novel claims about responsibility for the deep self, one is a sufficiency claim, another a distinct necessity claim, to answer the important question of what makes us responsible for our selves. In my view, the deep self consists of a set of psychological features that develop and persist *over time*, and are expressed crucially through, albeit not exhausted by, an agent's *deep valuing* and *deep values*.[1] So, there is both a historical dimension and a special kind of evaluative element in my understanding of the deep self.

Although many philosophers have appealed to the idea of the deep self, they do so for purposes different from the one in this paper. Some of them appeal to the deep self to identify the conditions of moral responsibility for what one does and its consequences, either in the accountability or attributability sense (Frankfurt, 1971; Watson, 1975; Wolf, 1990; Bratman, 1997; Shoemaker, 2011, 2015; Sripada, 2016; Gorman, 2019, 2022). Others aim to theorize personal autonomy (Dworkin, 1981; Bratman, 2005). My focus in this paper is not on applying the deep self to these issues but instead on explaining conditions for responsibility for one's deep self. Indeed, my view on responsibility for one's deep self on its own is not an alternative to any deep self views of moral responsibility. Nevertheless, the historical dimension of the deep self that is unique to my account, but overlooked by familiar views, will have implications on how a deep self view can better explain cases of indoctrination, especially those in which the development of an agent's deep self does not meet the historical requirement. Defending this

---

1   Throughout the paper, I will discuss *value* both as noun and as active verb. This is because I believe that 1) an agent with a deep self is one who has things about her self and in the world that she deeply values, and 2) that which she values deeply are her deep values. The second claim concerns the relationship between *valuing* and *having what one values as one's values*. To further explore and defend this claim deserves a full project of its own, and I will have to leave that task to another paper. In the following discussions, I will sometimes drop either value as noun or value as active verb, depending on my emphasis on the discussion at hand.

latter thesis deserves its own philosophical project, and is beyond the scope of this paper. But by developing the thesis of responsibility for the deep self in this paper, hopefully, I can establish important starting points to offer critical insight in future work on how we can rethink accountability and the deep self by paying closer attention to an agent's history.

In the first two sections, I will lay the groundwork for my two claims regarding responsibility for one's deep self by (i) introducing my understanding of the deep self, and (ii) how it is different from familiar accounts in the deep self literature. Then, I will offer a novel distinction between two kinds of deep self. Once that is done, I will articulate the sufficiency claim and the necessity claim for responsibility for the deep self. After that, I will further motivate my two claims, and the conceptual space generated by those two claims.

1. Laying the Groundwork Pt.1: What is the Deep Self?

1.1 The deep self in my view: A new version

I take the deep self to be consisting of a set of psychological features that develop and persist *over time*, and they are expressed crucially through, albeit not exhausted by, an agent's *deep valuing* and *deep values*. So, a historical dimension of the deep self and a special kind of evaluative element are crucial to my understanding of the deep self. In addition, the focus on the historical dimension is relevant to this extra level of evaluation and reflection in understanding the deep self. Let me start with the evaluative element.

By deeply valuing something, I mean that an agent judges it to be good, and desires it primarily for those reasons for which she judges it to be good. In addition, she is susceptible to a range of emotions responsive to it. Here, these three aspects of deep valuing do not merely co-occur but relate to each other in a non-additive way. Specifically, they are connected by the

reasons to which an agent is sensitive, and such a sensitivity need not be conscious, or in line with what is objectively or uncontroversially good. More importantly, that which she values in this way partly constitutes and crucially expresses her practical identity in the realms of morality, society, aesthetics, and physical and mental well-being, albeit, possibly, in a disparate manner.[2]

Given how an agent may pursue and fulfill values in different ways in these different realms, an agent with a deep self does not have to have that deep self *as a whole*, and as a matter of fact, many of us do not. Instead, our deep valuings in different realms of human life express different parts of our practical identities and our deep selves. As a result, we may have a deep self in the realm of morality, but not one in the realm of aesthetics. Or, we may later develop a deep self in the realm of aesthetics but only with regard to, say, the aesthetics of food, but not regarding the aesthetics of fine art.

How deep is deep valuing? To compare, consider first Alfred Mele's characterization of valuing something in the *thin* sense, which involves a conjunction of a positive motivational element of desiring it and an evaluative element of judging it to be good (1995: 116).[3] For instance, consider Peta who desires to eat an ice cream sandwich because her friends bet that she would do it, or because it is the only thing left in the freezer and she craves snacks in the

---

2 It is a vexing matter and so I mean to leave it as an unfinished philosophical project to fully state all the elements that bear on the constitution of one's practical identity. I thank Michael McKenna's suggestion to make this clarification.

3 Mele further distinguishes between thinly valuing something that is of importance to an agent and her personal values as follows:

"We can say that *S* at least *thinly values X* at a time if and only if at that time *S* both has a positive motivational attitude toward *X* and believes *X* to be good. Unfortunately, accepting this analysis does not settle what it is for something to be *among one's values*…Can we properly say that *X* is among a person's values if *X* is both valued by the person and of special importance to the person? No…[T]he range of personal values under consideration can be limited to things that are *valued by* valuers and are clear cases of the valuers' values" (Mele, 1995: 116).

I accept this distinction between thinly valuing something that is of importance to an agent and having it among her values. My view on deep valuing is an extension of what it is to have something among one's values. But as I stated in the previous footnote, to further develop and defend this idea is beyond the scope of this paper.

moment. Independently, she judges ice cream sandwiches to be good. However, she would rarely be motivated to go and get an ice cream sandwich without further enticement or under exceptional circumstances—like winning a bet or craving snacks with no other options besides an ice cream sandwich. She would count as thinly valuing ice cream sandwiches in Mele's sense because there is both positive motivational and evaluative components in her valuing ice cream sandwiches. However, ice cream sandwiches do not matter much to Peta and her practical identity in the realms of aesthetics of food or cuisine culture—it does not speak to who she is in those realms in life; she is no true gourmand and connoisseur of ice cream. This, of course, does not mean that she must not be a true gourmand and connoisseur of, say, spices. Rather, what is important to note here is that we would not know about this aspect of her practical identity with regard to the aesthetics of food and cuisine culture by her thinly valuing ice cream sandwiches. Deep valuing is deeper than Peta's valuing ice cream sandwiches, because the deeply valued item constitutes and expresses the valuer's practical identity in relevant realms of human life.

Or consider Peter who also desires ice cream sandwiches; but unlike Peta, he desires to eat them on a regular basis primarily for those reasons for which he judges them to be good. Peter would also count as thinly valuing ice cream sandwiches in Mele's sense, and in this case, ice cream sandwiches mean more to Peter than to Peta. But deep valuing is still deeper than that. For Peter, ice cream sandwiches do not constitute what he is in the realms of aesthetics of food or cuisine culture: like Peta, he is no true gourmand and connoisseur of ice cream, either. Indeed, to deeply value ice cream sandwiches, among other things, an agent could be a true gourmand and connoisseur of ice cream sandwiches, who judges ice cream sandwiches to be good, desires to eat and learn about them primarily for those reasons for which she judges them to be good, and is susceptible to a range of emotions in the prospect of not having access to sustain her

pursuit, for example. Her deeply valuing ice cream sandwiches makes her who she is in the realms of aesthetics of food and cuisine culture.

Though deep valuing is important to the deep self, I suggest that it does not exhaust the deep self. Indeed, there might be other explicit or implicit attitudes that an agent holds that are constitutive of who she is, but are not parts of her deep valuings and deep values. So, characterizing the deep self as crucially expressed by an agent's deep valuings does not mean that there is a privileged set of psychological features that *just is* one's deep self, like many deep self theorists would have us believe.[4] Nevertheless, I suggest that it is important to focus on an agent with a deep self as someone who has something in the world that she deeply values. This is because the way in which she deeply values something and acquires her values is critical to our assessment of her responsibility for her deep self: does she have a personal history in which she was afforded the unimpeded opportunities (i.e., a stable and healthy household growing up, access to education and affordable health care, and just society, and so on) to develop and exercise the ability necessary for responsibility? If she does not have such a personal history, as I will soon argue, then she is not responsible for who she is. This leads to the historical dimension of the deep self, a dimension unique to my view.

There are two parts of the historical dimension of the deep self. First, no one is born with a deep self. The true gourmand and connoisseur of ice cream sandwiches, for example, does not deeply value ice cream sandwiches from age one. It takes time for an agent to obtain characteristics of a deep self, and to develop and exercise the ability to acquire, examine, reject or retain, and fulfill her values, among other things that she does to and about her self. Second, not only does this process take time, it also requires that such an agent have a personal history in

---

4 For an insightful identification of this problem for a lot of deep deep self views, see Gorman (2022).

which she was afforded the unimpeded opportunities to develop and exercise the necessary ability to do so. As I will show momentarily, this feature of my view sets it apart from standard views in the literature.

1.2 The deep self in standard deep self views: Others vs. me

Different deep self theorists of moral responsibility define the deep self differently. For instance, Harry Frankfurt (1971, 1977, 1987) presents the deep self as captured by an agent's identification with her first-order desire through forming a second-order volition regarding that first-order desire. The aspects of herself that are identified in this way are thus incorporated into her deep self, whereas aspects that are repudiated are those with which she does not identify, and thus do not constitute her deep self. Gary Watson (1975, 1987) argues instead that an agent's deep self lies in her valuational system. This is because one's valuational system is an important source of one's will and gives one the authority to one's own behaviors. So, an agent's valuings constitute her practical standpoint about how to be, and what ought to be done.

In comparison to these two views, Michael Bratman (1997, 2000, 2003, 2004) proposes that an agent's deep self consists of her self-governing policies over time. This draws upon resources from his theory of planning agency, according to which an agent's practical standpoint and identity is instantiated through the governing of her intentions and activities over time by her planning through self-governing policies.

With renewed interest in this approach, philosophers such as David Shoemaker (2015) and Chandra Sripada (2016) argue that an agent's deep self is a disjunction of her caring and her evaluative commitments. Although she distinguishes her own view from the deep self view, Jada Twedt Strabbing (2016b) suggests that the self to whom an action can be attributed is her

practical identity understood in terms of her judgments for normative reasons. Benjamin Matheson (2019) proposes an ideal narrator that connects an agent's moral identity in different person-stages, and August Gorman (2019, 2022) argues that an agent's deep self is that which an agent identifies in terms of her partial and hypothetical approval for her behaviors were she to reflect on them.

The emphasis on the evaluative element in an agent's deep self is present in all these accounts, although in different ways. Some focus on an agent's actual psychological structure with which she identifies that explains a free action for which she is responsible. Some, instead, focus on psychological elements with which she would identify, either by an ideal narration or partial and hypothetical consideration. I suggest that these accounts either make the deep self too narrow or too wide. Traditional views like Frankfurt's, Watson's, or a combination of both, are too narrow—indeed, an agent needs not act from what she desires to desire, or values, or a disjunctive of these two things, to act freely and responsibly. More recent developments of the deep self view render the deep self too wide—ideal narration and partial, hypothetical identification may include elements in an agent's psychological construction that are either too ideal, partial, or hypothetical, and thus too external to her deep self, to capture the deep self.

In contrast, my view aims to be neither too narrow nor too wide. Joining philosophers like Charles Taylor (1976) and Christine Korsgaard (2009: 20, 22, 24, 42), I take us to be beings who not only evaluate ourselves and the world, but more importantly, we are also beings who are able to shape who we are and who—at least sometimes—do shape who we are. What express us as evaluative and self-shaping beings are our deep valuings, values, and our doing something to and about our selves primarily by means of various value-engagements. This, again, does not mean to restrict the deep self only to an agent's deep valuings and values. But the focus on the

special kind of evaluative elements is further motivated by the historical requirement I propose. So, my view is wider than traditional views. Furthermore, the historical dimension of the deep self I mentioned above expresses a developmental feature of the deep self—as we acquire and fulfill our deep valuings and values through the passage of time, in response to our experiences and interactions with the world, the deep self is acquired, pursued, realized, or later revised through the passage of time, and in response to those experiences and interactions. If we put it in terms of identification, then in my view, an agent's identification with her deep self is obtained, sustained, or revised in the very process of her self-shaping primarily by means of her various value-engagements. As we focus on the psychological elements employed in that process, namely, elements in aspiration (more on this in section 4), we discover who an agent is. This makes my understanding of the deep self just right—neither too wide nor too narrow.

2. Laying the Groundwork Pt. 2: Two Kinds of Deep Self

Now that I have established my understanding of the deep self and its evaluative and historical components, I offer a novel distinction between two kinds of deep self.[5] These two kinds of deep self will feature in both the sufficient and the necessary conditions for responsibility for the deep self that I will soon propose.

The distinction between the two kinds of the deep self concerns an agent's relationship with her deep valuings and values, especially with respect to the status of her value-acquisition and value fulfillment—namely, the stages she is at in the process of acquiring and living up to her values. Here, I take acquiring a value to be not only involving seeing something to be good

---

5   This distinction is significantly inspired by Agnes Callard's discussions on aspiration and responsibility (2008: Chapter 5, 6).

or valuable, but also seeing it to be good or valuable *to the agent.*[6] By seeing something as good to her, an agent may desire it, judge it to be good either consciously or subconsciously, feel certain emotions towards it, or all of these things together. In addition, to acquire a value is to be disposed to live up to it. Given the time and opportunities she is afforded, she acts on what she values and fulfills her values. Depending on what her deep valuings and values are, they will then be incorporated to who she is as a practical agent in the relevant realms of human life.

There are two ways to look at the stages an agent might be in with her value-acquisition and fulfillment. First, looking back, she may have already acquired and fulfilled values through various processes. Looking forward, she may be aiming at acquiring and fulfilling values she does not yet possess. Second, how values are acquired varies for different agents. It also varies in different stages in an agent's development. An agent may acquire values in a passive, unreflective, or superficial manner. For instance, people who undergo powerful indoctrination typically acquire values in a passive and unreflective way due to the nature of powerful indoctrination.[7] In comparison, an agent may also acquire values in an actively engaged manner. While adopting those values, she also reflects on them. Moreover, she may also reflect on the ways in which she acquires her values, consciously or subconsciously, directly or indirectly, through introspection or interpersonal inquiry.[8]

---

6   For an alternative way of understanding "seeing something as good to an agent herself," see Jeffrey Seidman (2009: 284-288). According to Seidman, when an agent sees something as important, she is disposed to believe that relevant considerations regarding it are practical reasons *for* her.

7   Although, indoctrination is not the only way of passively acquiring values. For example, people may simply feel content with values that they acquire and would feel discontent at the idea of not having them. But such feelings of content or discontent need not be the result of reflection.

8   Daniela Dover suggests that our conception of ourselves and conception of the world at large are comprehensively intertwined. An example to consider is through Dover's conversational model of interpersonal inquiry: when an agent talks about herself, she is also talking about the world; and when she talks about the world, she is offering a sense of who she is (Dover, 2022: 119-120). I think this is insightful and right.

To capture these differences, let me introduce my distinction between an *actual* deep self and an *aspired* deep self, the latter of which is a self different from one's actual self, and one that one aspires to become, for better or for worse.

The first way to understand the difference between an actual deep self and an aspired deep self is in terms of the stages of an agent's value-acquisition and fulfillment. An agent's actual deep self is crucially expressed by values that she has already acquired and lived up to; they constitute who she is *now*. An aspired deep self is crucially expressed by values that an agent is yet to acquire, that she judges or deems worth acquiring. Here, by deeming something as worth acquiring, I mean that an agent has not yet formed a judgment (consciously or not) either that it is good, or that it is good for her. But she may notice that she has a desire to learn about it and to take it as her own, or she may have a vague impression that it is good from other people's testimony (cf. Callard, 2018), among other things. This means that with values that express an aspired deep self, an agent has yet to possess them; they would constitute what she wants to become *in the future.*

A second way to understand the difference between an actual deep self and aspired deep self is in terms of how the relevant values are acquired. An aspired deep self is obtained by aspiration. And the process of aspiring involves both a cognitive state of judging or deeming it as good, and a conative state of taking actions to obtain the self that to which an agent aspires. To be sure, her actions towards obtaining it needs not result from her judging or deeming it to be worth acquiring. An actual deep self, in contrast, can be obtained either actively or passively in the manner mentioned above.

3. Responsibility for the Deep Self: Two Claims

Now that I have laid the groundwork for what I take to be the deep self and my distinction between an actual and an aspired deep self, let me propose two claims regarding responsibility for one's deep self. The first is a sufficiency claim, the second, a distinct necessity claim. Both trade in the two kinds of deep self discussed above.

Sufficient An agent is responsible for her actual deep self if it is obtained through her aspiration to transform her previous deep self to an aspired deep self, in which case the previously aspired deep self is now her actual deep self.[9]

Necessary An agent is responsible for her actual deep self only if she possesses the ability to aspire to transform her actual deep self to an aspired deep self, given the time and opportunity to do so. In addition, her not exercising this ability is by her own making, not something beyond her control.

To illustrate *Sufficient*, consider Gena.

Aspirational Gena Gena grows up in a community that emphasizes value system X, and acquires and fulfills values from X in a passive and unreflective manner. As she does this,

---

9   Here, I acknowledge that there is a challenge from manipulation and brain engineering of the sort such that an agent might come to aspire as a result of such manipulation and brain engineering. In that case, it is argued that such manipulative causes can be responsibility-defeating. See Mele (1995) for relevant discussions. These challenges will not be addressed in this paper, and for simplicity, I will leave the condition of non-manipulation implicit throughout.

I also acknowledge that there is another challenge from the regress problem, in that responsibility for an actual deep self must be consequences of previous actions for which an agent is responsible. But if this is the case, who is responsible for the self that issued those actions for which the agent is responsible? This could go on and on. For relevant proposals in response to this problem, see Kane (1999: Chapter 5) and Callard (2008: Chapter 5, 6). Although significant, I will not explore this topic in this paper.

she comes to obtain characteristics of an actual deep self. She then develops an ability to reflect on herself and, in particular, on the acquired values. In addition, she also develops and employs a sensitivity to the internal structure of her current values and becomes aware of values outside of X as she explores and interacts with the world. After noticing an inconsistency among her values, or conflict between hers and others', she may come to suspect that some of her acquired values should be deepened and others rejected in order to become the kind of person she wants to become. She then makes and carries out plans and policies for deepening the values she judges or deems worth deepening and rejecting values she judges or deems obstacles to her becoming the kind of person she wants to become. As she does this, she aspires to a different self that is crucially expressed by a set of values slightly, moderately, or drastically different from her current values.

Gena, and any agent who undergoes a similar process with passive and unreflective value-acquisition before obtaining a slightly, moderately, or drastically new deep self through aspiration for the first time, would become responsible for their deep self *in virtue of* having their actual deep self obtained through aspiration. And if they further aspire, they would be responsible for who they are in virtue of those further aspirations. Many of us have gone through processes like this.

As I have mentioned and will further motivate in the next section, we as practical agents can and do shape who we are. And this is an essential aspect of us as rational and responsible beings (cf., Taylor, 1976; Korsgaard, 2009). But how are we responsible for who we are given that we shape ourselves? This is further motivated by the intuitive idea that *we are responsible*

for something if we actively engage with the making of it.[10] Here, I suggest that if the very things that we are actively engaged with making are our very selves, we are responsible our selves. Indeed, aspiration in *Sufficient* indicates an active way of self-shaping.

But what if an agent never aspires? Would she be off the hook? Not necessarily, because some people shape who they are in a passive way; their self-shaping involves not an active effort, but rather, only the possession of an ability to do so. *Sufficient* reveals that important ability in *Necessary*: the ability to aspire to a different deep self. So, for an agent who never engages with active self-shaping, she may still be responsible for who she is on the condition that she has the ability to do so, does not exercise that ability despite having the time and opportunities to do so, and not doing it results from her own making, rather than something beyond her control.

To illustrate, consider John.

Zen Master John John grows up in the same community as Gena, and acquires and fulfills values from system X in a passive and unreflective manner. Just like Gena, he later obtains an actual deep self. He also gets the chance to develop the ability to reflect on and evaluate himself and the values acquired, and to aspire to a self different from his actual deep self. However, unlike Gena, given the time and opportunity, John does not exercise that ability to aspire to a deep self of any kind. Consciously or subconsciously, he chooses not to—he thinks to himself: "I am fine and content the way I am. Change is for other people, not for me." He remains to be the same person he ever was since 18. People call him a real Zen master.

---

10 See Kane (1999) for an explicit development of this idea.

Many of us have probably met a John at some point in our lives. John could even be one of our friends, relatives, or colleagues, who is the most easy-going person and laid-back about almost everything (though, they could be a death metal fanatic, for instance). Different from the kind of active self-shaping involved in *Sufficient*, this is another kind of self-shaping that involves only passive engagement with shaping one's self by merely staying idle. I suggest that someone like John is responsible for who they are on the condition that they possess the ability to actively shaping who they are, but stay idle with it. More importantly, not exercising that ability is through their own making, rather than anything beyond their control.[11]

In offering a sufficiency claim and a necessity claim as two separate principles, I mean to leave room for the possibility that if further sufficient conditions for responsibility are not met, an agent who is able to aspire might still not be responsible for her self. As I will illustrate in the last section, for example, an agent who was never afforded access to any moral knowledge would not be responsible for her self, despite possessing the ability to aspire. What would fill in the gap in this case? A complete theory of moral responsibility for one's deep self would answer this question. For the purpose of this paper, I leave it as an unfinished philosophical project to fully state all the sufficient conditions that bear on responsibility for one's deep self, and only argue for a more modest theory, one that advances one sufficient condition for responsibility for one's deep self, and one that advances a distinct necessary condition with important details informed by the sufficient condition.

4. Motivating the Sufficiency Claim: Aspiration and Responsibility

---

11 Could not exercising this ability by one's own making be beyond one's control? It could; there may be cases in which an agent's staying idle as an internal constraint results from something beyond their control as an external constraint. But I take these cases to borderline cases, not the central cases, of passive self-shaping. And I take only the central cases to be relevant to my discussion at hand. I thank Yili Zhou for discussion here.

So far, I have not fully motivated *Sufficient*. As I mentioned in the last section, *Sufficient* is further motivated by the intuitive idea that *we are responsible for something if we actively engage with the making of it*. In this section, I will provide two considerations for which aspiration, as a kind of active engagement with one's value-acquisition and fulfillment, is sufficient for responsibility for one's deep self.

First, aspiration involves employing psychological features such as an internal sensitivity and an external awareness, and exercising certain evaluative abilities. When we employ our internal sensitivity and external awareness to values we have and those we yet to have, and evaluate ourselves by evaluating those values, shifts in our internal motivational features tend to occur, which lead us to further value-engagements. When we successfully come to retain or reject, or acquire new values, we become a slightly, moderately, or drastically different version of ourselves. Sometimes it is for the better, sometimes it is for the worse. As a result, we actively shaped our selves.

Second, when we aspire to a different self in the sense that I will identify, there involves both a cognitive component of judging or deeming something to be worth aspiring to—obtaining a conception of what one wants to become, and, importantly, a conative component of acting towards realizing that conception. This indicates another way in which aspiration is the kind of active engagement with self-shaping one's deep self that is sufficient for responsibility for one's deep self. Let me start with the first consideration.

4.1 Internal sensitivity, external awareness, and a special kind of evaluation

Aspiration to a different self requires a sensitivity to features internal to an agent's current set of values. For instance, suppose there is an inconsistency between her two values, one is the value

of prioritizing self-preservation over everything else in face of a communal crisis, another is the value of prioritizing the preservation of her community over everything else in face of the same communal crisis. She may be able to recognize that in a situation like that, these two values may be impossible to realize at the same time. After employing her sensitivity to such an inconsistency, it may in turn induce a change in her internal motivational features to retain one value and reject another, revise both, or even reject both.

Aspiration to a different self also requires an awareness to values external to an agent's current values, which may help bring to her attention any conflicts between her own values and other values. For example, from having access to and learning about other values different from hers, our agent might come to realize that there are considerations other than preservation either of her community as a whole or of herself in face of a communal crisis, and thus come to further reject or revise some of her current values, or, instead, deepen her endorsement of some or all of her current values.

Employing an internal sensitivity and an external awareness leaves an opening for an agent to retain, revise or reject her current values, or acquire new values. Then, she may come to retain, revise or reject, or acquire certain values. As a result, aspects of her deep self changes. Employing her sensitivity and awareness thus is a first step for her active engagement with shaping her deep self.

In employing one's internal sensitivity and external awareness, not only can an aspirational agent evaluate her values and other values, she can also evaluate the ways in which she acquires her values, and other ways in which values can be acquired. Specifically—and important to aspiration—the evaluation involved concerns the following question: would those values and the ways in which they are acquired yield a way of living and being that an

aspirational agent herself judges or deems worth living and being? This is the evaluative component key to aspiration.

To further illustrate the kind of evaluation involved in aspiration, consider first Charles Taylor's (1976) distinction between qualitative evaluation and non-qualitative evaluation. When making a *non-qualitative* evaluation of one's motivation, an agent weighs her desires based on facts about herself regarding those desires, regardless of whether those facts about her are contingent on how she is and the nature of those desires. For instance, Peter may weigh his desires non-qualitatively for and against eating an ice cream sandwich in accordance with the intensity of those desires he feels at the moment. If the desire for an ice cream sandwich is stronger at the moment of evaluation, then he may act accordingly.

In comparison, in making a *qualitative* evaluation, an agent evaluates her desires *in accordance with a conception of a way of living and a way of being*. This conception supports certain options of living and being while repudiating others (Taylor, 1976: 283, 286-297). By aligning the target of her evaluation with a conception like this, in which case she may have that conception either consciously or subconsciously, she may come to endorse certain ways of living and being, and rejects others. By doing so, she would come to gain a sense of control over her way of living and being.

For example, suppose besides always loving ice cream sandwiches, Peter later obtains a conception of himself of one day becoming a true gourmand and connoisseur of ice cream sandwiches. He then could evaluate qualitatively his desires for ice cream sandwiches in accordance with that conception of himself. After this evaluation, he may deem eating ice cream sandwiches as an opportunity to learn the culture and history behind different ice cream sandwiches as an acceptable pursuit and end, in addition to enjoying them. Concomitantly, he

may deem eating ice cream sandwiches always and only for the sake of satisfying his sweet tooth as unacceptable. By evaluating qualitatively, Peter would come to gain a sense of control over his way of living and being regarding food aesthetics and cuisine culture by aligning his deep self in the relevant regard with his conception of living and being as a true gourmand and connoisseur of ice cream sandwiches.

This distinction between non-qualitative and qualitative evaluation is insightful for my discussion here. Namely, the kind of evaluation involved in aspiration sufficient for responsibility can be understood as qualitative evaluation. More specifically, in aspiring to a different deep self, together with employing her sensitivity and awareness, an agent evaluates her actual deep self in the attempt to align it with a conception of her self that she wants to realize. It thus involves qualitative evaluation in Taylor's sense. After such an evaluation, she judges or deems certain values worth retaining or revising, some worth acquiring, and others worth rejecting. She would then gain a sense of control over what she is by having aligned herself with what she wants to become. And this is an important step towards actively engaging in the making of one's self.

4.2 Realizing a conception of one's self of one's own

As I have suggested, having a conception of what one is to become is an important cognitive element in the sort of evaluation implicated in aspiration, which partly explains aspiration as the kind of active engagement sufficient for responsibility. To engage in courses of action to realize such a conception is, in turn, another important part of actively shaping one's self.

But first, let me clarify what I mean by having a conception of what one is to become. There are cases in which one has a conception of what one is to become and has it *as one's own,*

and other cases in which one has such a conception without having it *as one's own.* Here, I want to say that it is having the conception as one's own that is central to aspiration as active self-shaping sufficient for responsibility. The distinction between merely having a conception of one's self and having it as one's own is connected to how one acquires such a conception.[12] Specifically, under my account, if one's value-acquisition and fulfillment inhibits the development and employment of one's sensitivity and awareness to both their values and other values, their evaluation of those values, and the ways in which they acquire those values, one's conception of one's self—crucially expressed by her deep values—is not one's own.

To illustrate, consider the following two cases.

Oppressed Lucrecia Lucrecia acquires values in a community through systematically oppressive indoctrination, and only through this way. The indoctrination is so thorough and powerful that there is no control left in her to further develop and employ her sensitivity and awareness to evaluate values both within and outside of that value framework, and her ways of acquiring her values through indoctrination. Suppose after acquiring values through indoctrination, Lucrecia has developed characteristics of a deep self. In that case, her conception of her actual deep self and a different deep self to which she aspires would shift as the content and guidance of the indoctrination shift.

Liberated Lars Lars acquires values in the same community as Lucrecia and in a similar manner; namely, through systematically oppressive indoctrination. However, unlike Lucrecia, Lars somehow manages to develop and employ his sensitivity and awareness to

---

12 See Haji (1998) and Mele (1995) for understanding the distinction in this way.

values within and outside of that value framework despite how powerful and thorough the indoctrination is.[13] As a result, after employing his sensitivity and awareness, Lars actively endorses values he acquires through indoctrination, and either explicitly or implicitly endorses indoctrination as a good means to acquire values.[14] Upon acquiring and endorsing those values, he too, has developed characteristics of a deep self. In that case, his conception of his deep self and a self to which he aspired would shift, too, as the content and guidance of the indoctrination shift.

Despite both acquiring their values through the same powerful indoctrination, and despite both of their conceptions of their deep selves shifting as the content and guidance of the indoctrination shift, it strikes me that Lucrecia does *not* have a conception of her self as her own, whereas Lars does. Indeed, one can acquire a conception of one's deep self without being sensitive to the internal structure of the content of that conception or being aware of the challenges from values outside of that conception. If their ability to develop and employ such a sensitivity and an awareness is hindered by the very process of value-acquisition, like in *Oppressed Lucrecia*, but not in *Liberated Lars*, then one's conception of their self—expressed by the acquired values and valued to be acquired—would not be their own.

In other words, when the change of an agent's conception of her self only follows the change in the broader framework of values where she acquires her values, and the asymmetric dependency does not involve the employment of her sensitivity and awareness to her values and values different from hers, then the change is beyond her own making. She thus does not have a

---

13 How does on develop and employ an internal sensitivity and an external awareness in an environment like that? I leave it as an unsettled issue that bears on further development and defense. Here, I take it to be a result of empirical observation that some people do manage to do that despite their environment.

14 I thank Max F. Kramer for this point.

conception of herself as her own. So, not having the development and exercising of a sensitivity and an awareness inhibited is one negative constraint for obtaining one's own conception of one's self.

So far, I have suggested that what is required by aspiration is also required in order for one to form one's own conception of their self; namely, uninhibited development and employment of one's sensitivity and awareness. But apart from that, aspiration plays a more prominent role in possessing and realizing one's own conception of one's self. One helpful way to make sense of what it involves is to first look to Taylor's discussion of *articulating* one's self. To articulate one's self is to be equipped with a vocabulary of certain modes of life and modes of being (Taylor, 1976: 287-288). To articulate one's self in terms of a vocabulary of this kind is not only to *express* one's understanding of oneself. Crucially, it *shapes* one's self through shaping one's understanding (Taylor, 1976: 285-286).

To further illustrate, compare articulation and description. Mere *description* of an object is fulfilled when the description matches with what the object is like in itself. *Articulation* of an object is different from mere description in that it brings about a change both in the object articulated and the articulator's understanding of that object. In this sense, articulation is both *inventive* and *creative*. So, to the extent that a change of understanding in what an agent wants to become is brought about by an agent through her articulation of an aspired deep self, the agent is actively engaged with inventing and creating aspects of her actual deep self on a cognitive level. Possessing one's own conception of one's self involves articulation of that self, understood in Taylor's sense.

Nevertheless, I suggest that articulation of one's self, though important to having one's own conception of one's self, is *not* all there is to shaping and realizing that self. For example, an

agent may not articulate her self at a certain time due to a lack of vocabulary in reference to the kind of person she wants to be, and the kind of life she wants to lead. But does the lack of articulation mean that the agent may never get to realize a self she wants to become? No. Often we gain tools for self-articulation only after we have taken some action without fully understanding its implication on our conception of our selves. That is, one's understanding of one's self is obtained only after the performance of certain courses of action. In that case, articulation comes retrospectively, and only retrospectively.[15]

So, I suggest that having and realizing one's own conception of one's aspired deep self involves not only a cognitive state, as Taylor suggests, but requires also a conative state. Aspiration plays such a role because aspiration here involves a conative state that leads an agent to perform actions that will be instrumental to her future understanding and realizing of that self. Articulation of that self might initiate such a process, but it does not have to be. It may occur along the way, or only after such a process. Aspiration therefore features not only in securing an understanding of one's own conception of one's self, but also in realizing that conception.

Centering aspiration in *Sufficient* thus makes my view a further development of Taylor's view on articulation and responsibility for the self. Articulation of one's self involves only the cognitive aspect of how one can shape one's self through one's understanding. To be responsible for one's self, one's active engagement should involve more than a cognitive shake up. Aspiration plays a crucial role in this because an agent's aspiring to a different deep self contributes to both the understanding and the realizing of her own conception of that self. And thus, through aspiration, an agent actively engages with the making of her self. To the extent that

---

15 See Matheson (2019: 464) for an alternative expression of retrospective articulation I'm discussing here, which he called the "connectedness narrative explanation."

she contributes to the making of her deep self by way of aspiring to a different deep self, and obtains that self through aspiration, she is responsible for her actual deep self.

5. Motivating the Necessity Claim: Being Able to Aspire

After motivating *Sufficient*, let us return, in closing, to *Necessary*, which holds that an agent is responsible only if she has the ability to aspire to transform her actual deep self to an aspired deep self, given the time and opportunity to do so, and her not doing so is through her own making. Two further details will help to flesh out the view. The first detail is an emphasis on the *possession* of an ability to aspire. As I have suggested, we as human beings at least sometimes shape who we are, and this is because we are able to. *Sufficient* appeals to an active way of self-shaping, whereas *Necessary* appeals to a passive way of self-shaping. This passivity is expressed through an agent's possession of an ability to active self-shaping. The second detail is to explore the conceptual space left unsurveyed, specifically cases in which the possession of that ability is in place, but the agent still lacks responsibility for her deep self.

5.1 The possession of an ability to aspire

By being able to aspire to a deep self, I mean the possession of an ability that an agent exercises on relevant occasions.[16] Such a possession makes it possible that an agent can retain it while not exercising it when relevant circumstances obtain. In cases like this, an agent may be responsible because she possesses the ability, while not exercising it.[17]

---

16 My understanding of abilities is permissive between general abilities and specific abilities. I take the ability pertinent to responsibility for one's deep self to be an ability developed and retained by the agent over an extended period of time. When circumstances and opportunities obtain, the exercise of that ability would render an agent's relevant specific ability. But this is consistent with an agent's responsibility for her deep self when she does not exercise that ability. I thank Michael McKenna for pushing me to clarify this point.

17 See Strabbing (2016a) for a similar point on weak-willed actions, where she rightly points out the importance of possessing the relevant ability, rather than merely exercising it, to account for weak-willed actions.

To illustrate, consider again our aspirational Gena and Zen master John who grow up in the same community, acquire their values from the same value system, and both have acquired characteristics of a deep self upon such value-acquisition and fulfillment. Recall that Gena then aspires to a deep self after developing and exercising her sensitivity and awareness to values within and outside her current values, and her ability to evaluate her actual deep self. In comparison, John does not aspire to a deep self of any kind. Rather, he remains idle after he has acquired his current deep self, despite the time and opportunity available to him to exercise his ability to aspire.

It is easy to see that Gena possesses such an ability to aspire, given that she exercises it as she actually aspires to a different deep self. And let us assume that John possesses this ability too, without ever exercising it. For purposes of illustration, I have suggested that both are responsible for their deep selves. With *Necessary* in mind, we could say that both are responsible for their deep selves on the condition that they possess the ability to aspire, and by either exercising it or not exercising it, they shape themselves, actively or passively. Nevertheless, given that this difference in exercising their ability to aspire, the complete explanation for their responsibility differs. In Gena's case, she is responsible for her deep self by virtue of successfully exercising her ability to aspire, and thus actively and successfully shapes who she is. In John's case, he is responsible for his deep self because he possesses this ability, though he does not exercise it, and his not exercising it is through his own making.

But things could have turned out different for someone like John, who possesses the necessary ability in *Necessary*, and is given time and opportunities to exercise it, but not the sufficient condition of exercising it, and yet unlike John, they might not be responsible for their self. And this is because they do not satisfy some further sufficient conditions for responsibility

for the deep self which renders them not exercise that ability despite the time and opportunity afforded. Let me turn to a final sample case to complete the picture, one in which the necessary ability *Necessary* is satisfied, but an agent still lacks responsibility for her deep self.

5.2 Failing to aspire not through one's own making: A sample case

The reason John is responsible for not exercising his ability to aspire, but still is responsible for his Zen self is that his not aspiring is through his own making. He idles with his passively obtained actual deep self despite retaining the ability to do something about it, when having the time and opportunities to do so. Therefore, it is through his own making that does not aspire, which results in passive self-shaping. However, retaining an ability to aspire but never exercising that ability can also be principally caused by obstacles beyond an agent's own making. In this latter case, even if an agent possesses the ability to aspire, she may still not be responsible for her self.

To illustrate, consider Chantal.

Misled Chantal Chantal develops and retains the ability to aspire. She is able to employ an internal sensitivity and an awareness, and to evaluate her self and the world. She would exercise this ability were she to have the right knowledge. But sadly, Chantal's knowledge about the world and herself is so distorted that what she deems right from the moral education she gets is the complete opposite to what we deem right in the most uncontroversial cases. Given this lack of sufficient knowledge to act for any reasons relevant to proper moral evaluation, when Chantal is given time and presented with access to opportunities to exercise her ability to aspire, she could not recognize that those

opportunities are opportunities for her. As a result, she does not employ her executive

abilities to take advantage of those opportunities.[18]

Chantal can surely have a deep self that is developed through her morally abhorrent upbringing. Although, given the distorted moral knowledge that she was afforded, she would not have developed a deep self other than an evildoer's deep self. Is she responsible for the kind of evildoer she has become? I suggest not, even if she gets to develop and possess the necessary ability in *Necessary*. This is because when circumstances for her to exercise that ability obtain, she would not recognize them nor take advantage of them; her lack of moral knowledge prevents it. This case presses on the idea that circumstances that involve access to relevant moral knowledge matter to our assessment of an agent's responsibility for her deep self, arguably it could be one of the further sufficient conditions for responsibility for the deep self. And to the extent that circumstantial constraints of the sort prevent the agent from exercising her ability to aspire by misleading the agent with the wrong set of knowledge, she is not responsible for her deep self, despite possessing the ability to aspire.

Therefore, *Sufficient* and *Necessary* leave open and are consistent with the case discussed above. Filling out the remaining details would require nothing less than a full account of responsibility for one's deep self. Though I have not offered an account like this, I have shown and acknowledged the importance of certain circumstantial constraints in assessing one's responsibility for her deep self.

In conclusion, I have developed and defended two novel claims about responsibility for the deep self, to answer the important, but neglected question about what makes us responsible

---

18 Chantal can be compared to Susan Wolf's well-known case of Jojo (Wolf, 1990).

for who we are. In it, I offered a sufficiency claim and a distinct necessity claim. Both claims employ the idea of aspiration. There are, of course, further inquiries worth pursuing to further motivate my thesis and show its implications for work on responsibility. For example, the historical dimension of the deep self in my view can help better explain those cases of indoctrination in which we may want to see the indoctrinated agent as less blameworthy for their wrongdoing, where familiar deep self views cannot do so due to the lack of a historical dimension in their views. This could be further motivated by a potential connection between my thesis of responsibility for the deep self and a deep self view of moral responsibility for what we do and bring about—responsibility for who we are may be required for responsibility for what we do. Although these are important topics to further explore on their own, I must leave these projects to other papers. But hopefully, in this paper, I have established the important starting points that help us move towards those directions.[19]

CHAPTER 2

Coming out the Other Side, Responsible: Taking Inspiration From Aspiration

Abstract

Sometimes after we evaluate our current values, ends or elements in our motivation, or recognize new ones as worth obtaining, we come to retain or reject extant ones, or acquire new ones. As a result, we may learn a new skill, change our perspective on a political matter, deepen our appreciation of an aesthetic value, and so on. Something about us in the relevant domain changes. I call this process *aspirational self-shaping*. Philosophers have characterized these typically inward-looking and future-directed aspects of self-shaping in different ways. To continue their endeavor and bring it to new life, I suggest that understanding aspiration more fully sheds new light on a long-neglected topic on responsibility for our selves. Indeed, when we aspire in the way I identify it, we become responsible for what we are. To develop and defend this claim, I offer a model for aspiration in terms of a hierarchy of personal policy that fits with a realistic understanding of our practical agency.

Introduction

Suppose I was raised with the perfect immersive experience with death metal. It was the background music at home, while I was doing homework, on my way to and back from school, together with extensive discussions about it at the dinner table every night. Indeed, suppose I was raised by two death metal fanatics. I acquired the aesthetic value of appreciating music through my experience with death metal. Entering middle school, I made friends with music tastes different from my parents' and gained access to classic rock. "Oh my goodness," I said to myself, "I never knew music could make me feel the same Utopian dream collapsing without

inducing any headaches." With my friends' encouragement, I challenged my parents' musical education for me, and began a new chapter of value-pursuit in music by exploring a variety of genres other than death metal. I went to concerts with my friends, engaged in debates with my parents on the possibility of a genre-bending future in music, gained new perspectives on death metal, and deepened my appreciation of classic rock.

In a similar fashion, we evaluate our motivation, acquire values, set up ends and goals in different realms of human life, and reevaluate them given the opportunity, after gaining the intellectual, affective and practical tool to do so. As a result, we may reject or retain ones that we acquired in the past, pursue new ones, and try to live up to them. These engagements with our own motivation, values, ends and goals shape what we are as practical agents, and we often come out the other side a different person (slightly or drastically). When we actively embark on the endeavor with these engagements, we actively shape our selves. And if an agent actively engages with shaping her self, she is then responsible for it (Kane, 1996; cf., Taylor, 1976). Philosophers have discussed these aspects of self-shaping in different ways, either in terms of an agent's evaluation of her own motivation, her future-directed value-pursuit, or her end-setting (Frankfurt, 1971; Taylor, 1976; Schmidtz, 1994; Kane, 1996; Korsgaard, 2009; Callard, 2018). These aspects of self-shaping can be characteristically understood through the lens of *aspiration* (cf., Callard, 2018). So, aspirations of these kinds contribute to our understanding of responsibility for our selves.

Nevertheless, the idea that aspiration is *sufficient* for responsibility for our selves has not been explicitly developed in the literature.[20] I will take up this task. In this paper, I offer a novel

---

20 Here, let us assume that aspiration is not a product of external interferences such as indoctrination, covert or overt manipulation, brainwashing or brain engineering of that sort. I acknowledge that cases with these kinds of external interferences pose a challenge to the idea that aspiration is sufficient for responsibility for one's self, as these interferences can be argued to be responsibility-defeating. See Mele (1995) for relevant and extensive

model for aspiration that bears on our understanding of responsible agency. Indeed, engaging with her own motivational make-up, values, ends and goals in an aspirational way specifies an important way in which an agent actively contributes to the shaping of her self, and thus makes her responsible for her self. To actively engage with things in life in this way, I argue that an agent forms a higher-order intention about her own motivational make-up, values, ends and goals she deems worth acquiring, rejecting or retaining, and fulfilling. This higher-order intention will help her govern her motivation, acquire and fulfill her valuings, and achieve her ends and goals in accordance with what she deems valuable. And if she experiences a change of heart, she will later revisit and revise that intention. I call such a higher-order intention a *personal policy*. Aspiration is thus understood in terms of a hierarchy of personal policy in my view.

By having an account of aspiration like the one I am offering, we are connecting dots between value-pursuit, end-setting, and self-shaping in a way that will shed new light on a long-neglected topic on responsibility for what we are[21]—self-shaping in an aspirational way is sufficient for responsibility for our selves in that it *makes* us responsible for what we are. What is more, developing this sufficient condition reveals an important way in which we *take* responsibility for what we are and what we have become, if we deem worth doing so given the time and opportunity.

---

discussions on this point. I will not address this challenge in this paper. For simplicity, I will leave the qualification of *absence of responsibility-defeating interference* implicit throughout.

21 The idea that *to be responsible for what one does, one must first be responsible for what one is* has long been shared by moral responsibility theorists from different camps. See Susan Wolf (1990), Robert Kane (1996), Fischer and Ravizza (1998), and Ishtiyaque Haji (1998), among others. Though, no one in the literature has so far explicitly offered a sufficient or necessary condition for what makes us responsible for our selves. To be sure, this idea is by no means assumed to be correct. Others have explicitly argued against it. For example, Galen Strawson's skepticism about moral responsibility for actions results from what he calls a paradox of moral responsibility for one's self (Strawson, 1994: 6-7; 18-19). See a relevant reply from Agnes Callard that centers on her understanding of aspiration (Callard, 2018: 182-184; 205-206).

In what follows, I will first offer an overview of my account on aspiration and responsibility. I will then specify the sense in which aspiration is sufficient for responsibility for one's self; namely, we must understand aspiration to involve a conative state in which an agent actively takes actions towards achieving that to which she aspires. After that, I further show how my understanding of aspiration is realistic of what we do and what we can do as practical agents. I do so by offering a model of personal policy for it. It will require an understanding of a personal policy and its two functions. Last, I will illustrate my account with three promising young musicians, and show how differences in their aspirations play an important role in understanding their responsibility for their selves.

1. Aspiration and Responsibility: An Overview

My goal in this paper is to develop a sufficient condition for responsibility for what we are as practical agents, which centers on the idea of aspiration.[22] When we engage with values like I did with death metal and classic rock, for example, we experience things such as—but not necessarily in the particular order of—a sense of uncertainty or inner conflict about what we used to value or were taught to value, a curiosity and wonder in values that are new to us, a conscious or subconscious judgment that something among those new values is worth learning about and appreciating, and that values we acquired in the past deserve reevaluating. We then set up goals, make plans and policies to reevaluate what we knew, and learn and appreciate what is new. To

---

22 But what if an agent never aspires in the way I identify it? Would she be off the hook for being responsible for what she is? No. What we could say is that the sufficiency claim reveals the necessary ability for responsibility for one's self. That is, responsibility for one's self requires that one possesses *the ability to aspire*, and her failure to exercise that ability must be through her own faults, rather than something beyond her control. See Chapter 1 for an explicit development for this claim.

achieve those goals, we take action. During this process, elements in our motivational make-up, old goals and values maybe retained or rejected, and new ones may be acquired.

If we succeed at achieving that, aspects of our selves in the relevant domain will be shaped—our conception of our selves changes, and our conception of the world will shift, too.[23] Things we obtained in that process will be incorporated into our core identities. We are not the same people we were before this endeavor, and it would be difficult to shift back to what we were, because that endeavor will have left a set of stable traits in us. So, by actively engaging with our own motivation, values, ends and goals in this way, we shape what we are in relevant realms of human life.

Philosophers have discussed these aspects of shaping one's self in length. For example, Charles Taylor argues that what is important to responsibility for an agent's self is her *qualitative evaluation* of her own motivation, according to which an agent evaluates her motivation in accordance with a conception of modes of life she wants to lead, and the kind of person she wants to become (Taylor, 1976). David Schmidtz proposes that not only do agents rationally *pursue* ends for their own sake (as final ends), but they can also be justified in rationally *choosing* those ends as their final ends by means of having *maieutic ends*, the latter of which are achieved by an agent coming to choose and realize certain final ends for herself (Schmidtz, 1993). Christine Korsgaard suggests that as rational beings, we self-constitute by choosing to act, and by actually acting, in accordance with conceptions of particular practical identities. This makes us the authors of our actions and makers of our own identities (Korsgaard, 2009). Agnes Callard has recently argued that rational valuational transformation marks an agent's own

---

23 Daniela Dover suggests that our conception of ourselves and conception of the world at large are comprehensively intertwined. An example to consider is through Dover's conversational model of interpersonal inquiry: when an agent talks about herself, she is also talking about the world; and when she talks about the world, she is offering a sense of who she is (Dover, 2022: 119-120). I think this is insightful and right.

making of becoming what she wants to become, and she characterizes this transformation as aspirational. During that aspirational process, an agent acts for what Callard calls *proleptic reasons* to obtain values that she wants to obtain and has yet to fully understand. As a result, she will understand those values more fully by having acquired them (Callard, 2018).

As we can see, these aspects of self-shaping are characteristically both inward-looking and future-directed. Borrowing partly Callard's characterization of "aspiration," I suggest that we shape aspects of our selves in relevant domains of life *if we aspire by evaluating our motivation, ends and values, retaining or rejecting extant ones, and acquiring new ones,* among other things that we aspire to obtain that the possession of which effectively results in shaping our selves.

This idea about aspiration and self-shaping, in turn, complements an intuitive idea about responsibility—namely, when we actively engage with making something, we are responsible for it.[24] In those cases where the objects of our active engagements are our very selves, if we succeed at obtaining what we aspire to at least in the first instance of aspiring,[25] we are responsible for the relevant aspects of our selves in virtue of our active engagements with shaping them.

"But what is a self?" one might ask. Different views on responsibility give different answers. On those views that an agent's self plays the grounding role in virtue of which she is morally responsible, the key agenda has been to find the right agential features that stand in the right relation with her actions or omissions. Once this relation is properly captured, it will render

---

24 See Kane (1996) for an explicit development of this idea. Although, by endorsing his development of this idea, I do not mean to endorse or commit to his overarching goal in defending incompatibilism between determinism and freedom.

25 Indeed, an agent would not become responsible for her self immediately after she starts aspiring to shape her self.

her responsible for those actions or omissions that express her self, and excludes those that result from external or internal constraints such as compulsion and certain mental disorders. And yet, philosophers disagree on what exactly the right agential features are and what the right relation between them and an agent's actions is.[26]

My proposal on aspiration and responsibility is permissive of different understandings of an agent's self, and the purpose of this paper is beyond confirming or proposing any specific understanding of the self. Rather, my goal is to advance the importance of understanding an agent doing something *to* her self and *about* her self, however the self is understood. Indeed, in my view, what I call aspiration is a critical way in which an agent does something to and about her self. In particular, when she evaluates and modifies her motivational make-up, ends and values, she is actively engaged with doing something to and about her self. And if she succeeds at actively shaping her self in the relevant regard at least in the first instance of these engagements, she becomes responsible for her self. In addition, understanding aspiration and responsibility in this way helps show what an agent can do to take responsibility for what she is, if she ever deems worth doing so.

2. Aspiration in What Sense?

---

26 For instance, traditional deep self theorist views such as in Harry Frankfurt's puts forward an understanding of the self crucial to responsibility in terms of a hierarchy of desires, in which an agent's self is identified through her second-order desires about her first-order desires (Frankfurt, 1971) ; Gary Watson argues, instead, that an agent's self lies in her valuational system, where her valuings constitute who she truly is as a practical agent with practical stances in the world (Watson, 1975, 1987) . Michael Bratman's suggests that an agent's self is a self extended and sustained and cross-temporally by means of her planning and self-governing policies over time. New developments in this area offer new insight. To list a few, David Shoemaker and Chandra Sripada take an agent's self as constituted by her cares and commitments (Shoemaker, 2015; Sripada, 2016); Jada Twedt Strabbing proposes that an agent's practical identity is constituted by her judgments for normative reasons (Strabbing, 2016b); Benjamin Matheson argues that an agent's moral identity through different person-stages is connected by an ideal narrator who can provide a narrative that explains her behaviors as attributable to her (Matheson, 2019); and August Gorman understands an agent's self in terms of her partial and hypothetical approval of her behaviors were she to reflect on them (Gorman, 2019, 2022).

The sense in which aspiration is sufficient for responsibility for one's self bears specification. Although it might be intuitive from how we normally use the word "aspiration" to think that it merely suggests a future-directed attitudinal change in the cognitive sense—that is, we aspire to obtain a musical value, for example, just when we aspire to obtain a further understanding of it. Nevertheless, for aspiration to indicate an active engagement with the shaping of one's self, it cannot merely involve a cognitive state. More importantly, it must involve a conative state where an agent actively takes courses of action to obtain and fulfill what she aspires to.

To explain, if aspiration involves merely a cognitive state of grasping what one recognizes as good or valuable, consciously or not, then one aspires to obtain a value, end or goal just when one aspires to obtain an understanding of it. But in that case, one may or may not have done anything to and about one's self, nor does one have to do anything about it after obtaining that understanding. How so?

First, an agent may obtain an understanding of, say, different political values, without having done anything to and about her self in the political domain. She may have acquired that understanding from a friend who is a political activist. She appreciates her friendship with them and is inquisitive about what they are and what they value. And yet, she does not need to do anything to or about her self to gain an interest in understanding different political values. When learning *from* her friend, she might simply be interested in learning *about* her friend. In that case, an understanding of a value is independent of whether someone has done anything to and about one's self. Thus, aspiration understood as merely involving a cognitive state about what one recognizes as good or valuable can be independent of whether one has done anything to and about one's self. This understanding, in turn, does not help show the sense in which aspiration is

sufficient for responsibility for one's self in virtue of one's active engagements with shaping one's self.

Second, an understanding of a value is not sufficient for inducing a shift in one's internal motivational features that involve actions. That means that no shaping of one's self needs to occur after one obtains that understanding. For example, suppose our protagonist's appreciation for her friend took a romantic turn and she is now motivated to spend all her time to be with them. She took four semesters of extensive courses in political theory together with them, and as a result, she acquired an understanding of every political theory and ideal there is. And yet, despite her determination to be with her friend, to participate in political affairs to acquire and live up to certain political values is not part of her life—she does not find herself subscribing to any political values after finishing the courses, and being political is in no way constitutive of what she is. And let us suppose that our protagonist does not live in a society where no opportunities are afforded to her to acquire and live up to political values that she may subscribe to. Similar to the previous case, obtaining an understanding of a value is independent of whether someone will do something to and about her self in the relevant domain of life. Therefore, aspiration understood as merely involving a cognitive state about what one recognizes as good or valuable can be independent of whether someone will do something to and about their self. This too will not help show the sense in which aspiration is sufficient for responsibility for one's self in virtue of one's active engagements with shaping one's self.

To be clear, neither of these cases shows that a pertinent understanding is *not* relevant to aspiration. Rather, one has to have at least some understanding of what one wants to obtain to aspire to it, however vague, incomplete, or incorrect that understanding is. So, an understanding *of the target of aspiration* is necessary for aspiration.

As I have suggested, the process of actively engaging with shaping one's self typically involves, albeit not exhausted by, an evaluation of what one already knew or understood, and a recognition of something new that is worth understanding and obtaining. The evaluation and recognition should induce a shift in one's internal motivational features so that one forms plans and policies and then acts from these plans and policies to acquire what one deems worth acquiring, and reject or retain what one deems worth doing so. Understood in this way, aspiration in the sense that is sufficient for responsibility for one's self must involve a conative state for an agent to perform courses of action to obtain that to which she aspires, in addition to a cognitive state of recognizing it as good or valuable. In other words, one does not aspire in the sense identified without taking actions to obtain that to which one aspires. Let us call aspiration in this sense *robust* aspiration. In this paper, I use and understand aspiration only in the robust sense.

3. A Model for the Aspirational: A Hierarchy of Personal Policy

How is robust aspiration a realist understanding of what we do to be responsible for our selves and what we can do to take responsibility for our selves? To complete the picture, in what follows, I will offer a model for aspiration in terms of a hierarchy of personal policy.

A personal policy is an executive attitude that an agent has towards things she aspires to obtain. It is a higher-order intention that guides and structures her other intentions about that to which she aspires. My understanding of a personal policy is broadly in line with Michael Bratman's understanding of a self-governing policy, according to which a policy is a general commitment to an agent's cross-temporal plans and planning (Bratman, 1987; 2007). However, I

should clarify two things about my personal policy that make it diverge from Bratman's self-governing policy.

First, my personal policies are only a subset of Bratman's self-governing policies. Although in a crude sense it is true that everything we do in accordance with our plans and policies will inevitably shape what we are, even just a little bit, I focus on those policies from which when an agent acts, her self is shaped in a more purposefully self-shaping way than by her acting from her self-governing policy to put seat belt on every time she sits behind the wheel.

Second, Bratman emphasizes the normative constraints on plan consistency and means-end coherence. For him, policy formation and conformity are subject to instrumental rationality, and primarily so. My focus on personal policy, instead, is not exhausted by this instrumental model of rationality. It is important to my view that pursuing courses of action to obtain that to which she aspires is not merely characterized by an agent following through her instrumental reasoning. In addition to that, by forming and acting from or later revising her personal policy about the target of aspiration, an aspiring agent is more than governing her self: she governs her self in the pursuit of actively shaping her self so that she becomes what she wants to become, whether she does this consciously or subconsciously. Korsgaard puts forward a similar idea when she argues that our actions are governed by our conceptions of our practical identities. We pursue certain actions that we deem as worthwhile for the sake of certain ends, the more general form of which she characterizes as "one of the inescapable tasks of human life." In doing so, we constitute ourselves and make our own identities (Korsgaard, 2009: 20, 22, 24, 42). To push this idea even further, pursuing that to which she aspires, whatever it is for the agent, is rationally chosen and justified by means of having what Schmidtz calls a *maieutic end*—namely, a further end to an agent's final ends, the latter of which are what she pursues for their own sake. A

maieutic end is achieved by means of her choosing and achieving those final ends (Schmidtz, 1994: 226; 231). In the case of aspiration, the maieutic end for all aspirational agents is *to aspire to something that they deem worth doing so*.

Callard makes a similar but more elaborative point to Schmidtz's by making a distinction between two faces of aspiration—"a proximate face that reflects the kinds of things that appeal to the person she is now and a distal one that reflects the character and motivation of the person she is trying to be. (Callard, 2018: 73)" So, an agent employs her instrumental reasoning in the process of aspiring, but to aspire is more than that—she is at the same time aiming at achieving something she is yet to understand fully or possess, but deems worth achieving, and the soon-to-be achieved justifies her instrumental reasoning. To borrow this distinction, in aspiring to something, I suggest that an agent employs her instrumental reasoning in forming and acting from or later revising her personal policies, and this is the proximate face of aspiration. Beyond that, her personal policy formation, conformity, and revision is justified by achieving the target of her aspiration, whatever it is for the agent. This is the distal face of aspiration.[27] So, when a personal policy is about rejecting or retaining an agent's ends, values, or elements in her

---

27 I take my model of aspiration in terms of personal policy to be a further development of Callard's view. Callard proposes the idea of *a proleptic reason* for which an agent acts to aspire. The two faces of aspiration are the two faces of a proleptic reason. For Callard, only the proximate face is motivating, because it results from an agent employing her instrumental reasoning. The distal face is not motivating, because according Callard, when an agent aspires, the distal face is connected to the intra-value of the target of aspiration that an agent is *yet* to fully understand or acquire. A partial and incomplete understanding of the intra-value of the target of aspiration cannot be motivating. But I suggest that without further elaboration on how these two faces connect to each other, this proposal runs the risk of reducing a proleptic reason to just any motivating reason—it seems that any agent, aspirational or not, can act from a proleptic reason so long as she employs her instrumental reasoning and possesses an incomplete understanding of the pursued, in which the incomplete understanding can indeed justify her motivation. To resolve this, I suggest that to make a proleptic reason characteristically about aspiration, we should incorporate my idea on personal policy. Namely, to act for a proleptic reason to aspire, an aspiring agent forms, acts from, or later revises her personal policy about the aspired. As I will argue momentarily, a personal policy is implicitly motivating, and it motivates in accordance with what an agent deems worthy aspiring to. The two faces of aspiration are integrated in an agent's personal policy formation, conformity, and revision. So, an agent who forms and acts from a personal policy about an aspired object is more characteristically aspirational than one who does not.

motivational make-up, or acquiring new ones, it is rationally supported by her general value judgment about aspiring to something she deems worthy doing so, and more specific value judgments about the soon-to-be acquired, retained, or rejected.[28] To be sure, an agent needs not hold those judgments consciously, nor do these judgments have to be correct.[29] For example, consider my personal policy of going to classic rock concerts more often. It is a higher-order intention to do what I already intend to do or may intend to do—exploring and appreciating musical genres other than death metal. It is rationally supported by my general judgment that music appreciation is worth pursuing, and my more specific judgment that classic rock has great musical value, thus is worth appreciating. By forming this executive attitude and acting in accordance with it, I pursue and hopefully acquire what I deem as good or valuable through my actions.

To further explain a personal policy's executive role in governing and realizing an agent's pursuit in self-shaping, let me introduce a personal policy's two functions—a personal policy as a form of desiring, and as a form of valuing and value-fulfillment.

I start with the first. I suggest that a personal policy can function as a form of desiring either by helping an agent to select among and favor certain occurrent desires she has or generate new desires in her.[30] A personal policy can help govern an agent's motivation in different ways.

---

28 The way I understand a personal policy being rationally supported by one's value judgments is similar to Angela Smith's understanding of the rational relation between an agent's attributable attitudes and her value judgments. That is, an attributable attitude is rationally supported by an agent's value judgments if it is a direct reflection of her value judgments, or it is susceptible to be governed by it (Smith, 2005). Similarly, a personal policy is rationally supported by one's value judgments if it is a direct reflection of the latter, or it is susceptible to be governed by it. Although note that the agenda of this paper is different from the one in Smith's paper.

29 For a similar characterizations of responsibility relevant judgment, see Strabbing (2016b: 744-745).

30 Regarding this latter point, Bratman suggests that he departs from Frankfurt's (1971) treatment of higher-order desires to first-order desires in that a higher-order policy about, say, helping others, can do more than just involving an extant first-order desire to help. When such a desire is absent, a policy can "involve a commitment to coming to have the desire" (Bratman, 2000: 260). I accept this idea and take "the commitment to coming to have a desire" to be a way of generating a new desire. Bratman then suggests that a policy involving a commitment to coming to have a desire that is currently absent would count as a peripheral case of valuing (261). I diverge from this point. As I will argue, by generating a new desire, a personal policy manifests its key

When selecting among an agent's occurrent desires, a personal policy helps an agent to adjudicate between competing desires. For example, suppose I struggle to come to a decision about whether to go to a classic rock concert this weekend, or to work on my paper at home. I desire to do both, but it cannot happen at once. If I have a personal policy about self-education in music, it can help me favor the relevant desire to go to the concert, but not the other desire to work on my paper.

Although, in playing an executive role to govern an agent's motivation, a personal policy does not need to trump every occurrent but irrelevant desire every time a conflict of this kind arises. Indeed, I may give in to the desire of working on my paper, even though I have a personal policy of music education. But if a perceived personal policy never helps an agent to select and favor the relevant and occurrent desire of hers over her other occurrent but irrelevant desires, then it is reasonable for us to question whether she has formed that personal policy. Still, one might wonder if she indeed has formed that personal policy, but because it remains the lowest priority among her personal policies, it never effectively governs her motivation. In this latter case, I suggest that it is then reasonable for us to question if a permanently inert personal policy like that is, after all, still a personal policy that concerns the discussion at hand[31]

Now suppose, more plausibly, that together with a personal policy of music education, I also have a personal policy of building a future career in academia. My desire to work on my paper is relevant to this policy. In cases like this where a conflict arises between competing desires that are governed by different personal policies, one can appeal to a higher-order personal policy to resolve it. A higher-order personal policy in this case could be about an idiosyncratic

---

function of not only helping an agent to value the aspired, but also to fulfill it, as a personal policy helps effectively translate her value judgments to her actions.

31 I thank Luke Goleman for pointing out this latter case to me.

ranking among different values. For example, I may judge career building more valuable to me than self-education in music. Or it could be about consulting a friend whom one sees as one's rational counterpart. For example, I could talk with my friend who knows me well and is usually not a fallacious reasoner, and ask their opinion.

Consider another case where there are two equally desirable courses of action, and an agent has an occurrent desire for one but not for the other, and both are in line with the same personal policy. For example, going to a classic rock concert and watching a documentary about classic rock are equally desirable and relevant to my self-education in music. But suppose I only have the desire to go to the concert, and not a desire to watch the documentary. In that case, my personal policy can help me favor the relevant desire I have, rather than the one I do not have.

But what if I come to desire to watch a documentary of classic rock the next week, and the conflict arises, but this time between two occurrent desires of mine, both in line with the same personal policy. And suppose further that the last chance to go to the concert and to watch the documentary are on the same day. In a case like this, I suggest that by way of acting from my personal policy of music education, I can either follow my desire for the concert or the documentary. Doing either would help manifest my active engagement with what I deem valuable. And settling on one course of action in a way that renders not settling on another leaves a trail of regret or disappointment, but this regret does not compromise a personal policy's executive role with governing an agent's motivation and realizing her valuings.

Beyond helping an agent to select among and favor her occurrent desires, a personal policy can also govern her motivation by generating new desires in her. Suppose when I first form a personal policy of self-education in music, I have no specific desire for any particular

course of action because I have no idea where to start my pursuit. By forming that personal policy, it may help generate a relevant desire in me by first directing me to collect information on where to start, for example, I could ask my musician friend for suggestions, or enroll myself to an extracurricular course in contemporary music, and so on. So, by raising an agent's sensitivity and awareness towards relevant matters, a personal policy can help generate new desires in an agent.

The treatment of a personal policy towards an agent's motivation is about providing reason-giving attitudes in her. Here, I understand a personal policy's reason giving feature as broadly in line with Bratman's understanding of it. That is, to treat a desire as reason-giving is to treat it as a potentially justifying end in one's practical reasoning and action (Bratman, 1996: 8-9). So, when a personal policy helps an agent to select among and favor her occurrent desires, it helps her to treat those desires as potentially justifying ends in her practical reasoning and actions. Similarly, a personal policy generates a new desire for an action after an agent recognizes a reason for that action.[32] The desire generated is then treated as a potentially justifying end in her practical reasoning and action.

Given this understanding of a personal policy, it is implicitly motivating. Aspiration in terms of forming and acting from a personal policy is then implicitly motivating, because it helps shift an agent's internal motivational features by treating certain desires as potentially justifying ends in her practical reasoning and actions. As I have suggested, this is achieved by means of helping an agent to select among and favor her occurrent desires or generate new desires in her.

A personal policy's treatment of an agent's motivation makes a model of personal policy inherently hierarchical. This is because a personal policy's function of treating an agent's first-

---

32 I thank Michael McKenna for pushing me to clarify this latter point.

order desires as reason-giving is itself a policy-like attitude (cf., Bratman, 2000: 263). One important component of that function is to bring about what an agent deems valuable to bring about. So, when a personal policy helps an agent to treat a desire as reason-giving, the treatment itself results from a judgment that what she deems valuable to bring about is valuable to be brought about. Understanding it in this way, a model for aspiration in terms of personal policy is inherently hierarchical.

This model may seem reminiscent of a familiar model of hierarchical desires, the latter of which is most famously employed to identify conditions for free actions for which an agent is morally responsible (Frankfurt, 1971).[33] With a different goal in mind, I argue that my model of a hierarchy of personal policy fares better than a model of hierarchical desires in explaining an agent's active engagements with shaping her self.

To explain, consider the case where a relevant first-order desire is absent in an agent. When a personal policy succeeds at shifting an agent's internal motivational features by generating a relevant first-order desire in her, it can succeed in more ways than one. For example, recall little old me in the very first stage of self-education in classic rock, without having any idea where to start my pursuit. Among the many things I could try, I could ask my musician friend for suggestions, or enroll myself to a course in contemporary music. In comparison, when a second-order desire succeeds at shifting an agent's internal motivational features, it succeeds only in one way—by desiring to desire. But a well-known criticism of this

---

33 According to Frankfurt's account, for example, an agent acts freely and is responsible for what she does if and only if she acts from a first-order desire with which she identifies through a second-order desire about it. Her identification marks a division within her self so that actions that issue from those first-order desires with which she identifies via her second-order desires are performed *by her* in a representative and authoritative way, rather than external or internal forces alien to her, and for which she is thus responsible. This model has the advantage of creating a boundary within an agent' self where we can distinguish actions performed by the agent that are free, and those other actions, although performed by her, are brought about by external or internal constraints, and thus are not free actions for which she is responsible.

model says it best: a second-order desire is not so different in nature from a first-order desire; it is after all still a desire (Watson, 1975: 218; Velleman, 1992: 471). So, when a first-order desire is absent in an agent's motivational set, we do not have a good reason to think that a second-order desire could generate it.

On a relevant note, when both a personal policy and a second-order desire fail to generate a new first-order desire in an agent, the number of ways they fail reveals a difference in our assessment of her self-shaping and responsibility for her self. If an agent tries all that she can by modifying her personal policy over and over again, and yet still fails, we might want to say that the relevant aspect of her self is not something for which she is responsible, since she simply cannot actively shape her self in that respect. In comparison, when an agent fails to desire something by means of possessing or forming a second-order desire about it, we cannot conclude that she cannot shape her self in that respect. Because her failure to desire through her second-order desire is not conclusive of her inability to actively shape her self. And thus, we cannot draw the same conclusion from her failure in desiring to desire that she is not responsible for the relevant aspect of her self.

Not only does aspiration understood as a hierarchy of personal policy helps govern and manage an agent's motivation, it also does so *in accordance with what an agent deems valuable.* So, a personal policy helps an agent to procure her valuings and fulfill her values. For example, *through* my forming and acting from the relevant personal policy, the value judgment that classic rock is worth learning about and appreciating motivates me to perform courses of action. By forming and acting from it, I may be settled on a particular course of action of, say, going to the concert. Once I do that, I am a step closer to further valuing it, and potentially living up to that value.

But someone might find this all too easy for us to circle back to the folk understanding that aspiration involves just a cognitive state where an agent judges the target of aspiration to be valuable. Because given what I just said, value judgments seem to be at the bottom of an agent's active engagements with her motivation, ends and values. But to think of aspiration in this way is not right. I suggest that a distinction between valuing something and judging it to be valuable is key.

Valuing something, as many philosophers have argued, involves a desire for it, and thus is implicitly motivating, whereas merely judging something to be valuable is not (Watson, 1987: 150; Herman, 1993: 129-132; Mele, 1995: 115). The difference between valuing something and judging it to be valuable is relevant here. Although an agent's value judgment about the target of her aspiration is an important component of her relevant personal policy, the personal policy is more than an expression of that judgment. As I have suggested, a personal policy as a form of desiring is implicitly motivating. So, understanding aspiration in terms of a hierarchy of personal policy does not reduce it to mere value judgments.

But does this mean that a model for aspiration in terms of personal policy can be, instead, reduced to a mere instance of valuing as a combination of judging the target of aspiration to be valuable and desiring it?[34] This would not be the case with aspiration, either. Here is a counterexample to consider. One may desire to go to a classic rock concert only because they want to impress their friends. Independently, they may judge classic rock to have musical value, despite never getting motivated to listen to it for reasons that they judge it to be valuable. By going to the concert due to external enticement, are they aspirational about classic rock? It does not seem that way. Rather, one may be thought to value classic rock only in a particular fashion :

---

34 I thank Michael McKenna for raising this concern.

they value it one-off.[35] And their one-off valuing needs not extend from, nor realized by their personal policy forming and conformity—it is reasonable to assume that they did not form any personal policy about it.

So, a model of personal policy, as a form of valuing and value-fulfillment, cannot be reduced to mere value judgments, nor is it an instance of one-off valuing. Rather, a hierarchy of personal policy as a model for aspiration helps explain how an agent shapes her self by means of governing and managing her own motivation in accordance with what she deems valuable. When she aspires to acquire a new end or value, reevaluate her current motivational make-up, ends or values so that she either ends up retaining or rejecting them, she forms a personal policy about obtaining that which she aspires and acts from it, in accordance with her judgment about them. If she experiences a change of heart, she may revisit and revise that personal policy and act from it. In the process of doing this, she shapes aspects of her self. And thus, she is responsible for those aspects of her self in virtue of her active engagements with them by her personal policy formation, conformity and revision.

4. Lily, Billy, and Milly: Two Aspirational; Another, Interrupted

To illustrate the model I offered, consider three promising musicians:

---

35 Gary Watson's "perverse cases" (1987: 150) can be understood in terms of this *one-off valuing*. According to Watson, perverse cases are cases where an agent is motivated to do things that she judges as good or valuable in a way that is incongruous with what she values that is expressive of who she truly is—her deep self. For example, she may do something she thinks is fun or thrilling, and in that sense judges it to be good or valuable. Nevertheless, in doing that, she is not acting in accordance with her valuational system in a typical way that renders her free and responsible. And yet, she is also not acting out of internal disturbances that would render her unfree and non-responsible. Watson raises these perverse cases to express a concern for his own account of free agency and identification, where one's identification comes through one's valuational system. But if there is a principled way to differentiate perverse cases and cases that express free agency, the idea of identification should be explained in a way beyond appealing to one's valuational system.

Lily Lily plays the piano from a young age and she plays it well. She loves classical music, and is deeply fascinated by a performer's ability to transform herself into a vessel through which a composer's intention, vision and story-telling are brought to our ears. With this love for classical music and piano performance, she got admitted to a music program and starts taking courses to deepen her knowledge, and refine her skills. As she moves forward with her pursuit, her appreciation for these things is deepened after gaining more knowledge, obtaining new skills, and further developing her musicality. She now values classical music and piano performance on a different level than she did before.

Billy Lily has a classmate in the music school, Billy. Similar to Lily, he plays the piano from a young age and he plays it well. The prevailing values for a man's self-fulfillment in his community is to have a decent but not too ambitious career in the arts, which will then help him to marry well. He subscribes to these values, and recognizes being a classical pianist as a promising way to live up to them. So, he applied for the same program as Lily and got admitted at the same time. He takes the same courses, and practices just as diligently. As he moves forward with his pursuit, he learns more about classical music and gains advanced skills in piano performance. Nevertheless, during this process, his appreciation for music and musicianship is boosted only in the sense that he recognizes each step he takes as further contributing to obtaining a profession that will hopefully bring him a good marriage.

Milly Lily and Billy have a classmate Milly, who got into the same program at the same time. She, too, plays the piano from a young age, and appreciates classical music and piano performance for the same reasons Lily appreciates them. She is determined to become a better musician and a deeper lover of classical music just like Lily. She takes the same courses, and makes practice plans together with both Lily and Billy. However, once in a while Milly would miss classes for extended weeks and not show up in her practice room. Whenever she comes back, she would try her best to catch up and adjust her plans and policies accordingly. But she struggles to follow them through. As it turns out, due to an undiagnosed and untreated mood disorder, she makes and revises her plans and policies in hypomanic episodes when she experiences an enhanced level of energy and an elevated sense of self-government, and falls into extended depressive episodes that follow hypomania, and can never stick to her plans and policies. A semester has passed; she ends up not making any progress.

All three of them are perceived promising young musicians when they first got into the program. They made up their mind to pursue music in the music school, and make plans and policies to live up to what they deem valuable in life with regard to music. Nevertheless, it strikes me that they are not exactly the same with their pursuit, and I suggest that we can explain their differences in terms of differences in their aspirations.

Let me start with Lily. Lily's goal is to become a better musician and a deeper lover of classical music. She judges classical music to be worth appreciating, musicianship as worth pursuing, and has a personal history where she was afforded the opportunity to gain a rough

understanding about those things. She makes policies to obtain her goal for reasons that she judges classical music and musicianship to be worth appreciating and pursuing. In acting from those policies, she gets closer and closer to her goal. During this process, her conception of her self with regard to music gradually changes, and her conception of the world shifts too. She is not the same person she was before this pursuit, and it would be difficult or almost impossible for her to shift back to what she was, given that the pursuit will have left a set of stable traits in her. With my understanding of aspiration, we can say that she is aspirational about music and musicianship. Her aspiration shapes aspects of her self in that regard. She is thus responsible for those aspects of her self in virtue of her active engagements with shaping them.

Billy may seem exactly the same as Lily on the surface. He, too, judges classical music to be worth appreciating, musicianship as worth pursuing, and has a personal history where he was afforded the opportunity to gain an understanding about those things. However, unlike Lily, his primary goal is to conform with the communal values of having a decent but not too ambitious career in the arts so that eventually that will bring him a decent marriage. So, his pursuit in music and musicianship is not merely supported by his judgment about music and musicianship in their own right. Rather, it is further and primarily supported by his judgment about their instrumental value in living up to the communal value he endorses. If having refined skills in piano performance and an extensive knowledge in classical music are not required by entering into the profession as a means to live up to his communal values, Billy would not have made those policies in music school that he has in fact made, and would have been less engaged with this pursuit in music and musicianship than he is now. But given the actual requirements to enter the profession, he has made the same policies as Lily, although only coincidentally so, and with a different goal at which his policies aim. In acting from those policies, he gets closer to his goal.

During this process, his conception of himself with regard to music changes, but those changes matter to him only as a means to help him measure and evaluate where he is at in living up to his communal values.

Given the model of aspiration I offered, Billy is aspirational about music and musicianship, but less wholeheartedly so than Lily. However, this does not diminish his responsibility for those aspects of his self shaped by his pursuit in music, because his aspiration lies primarily in something else—he is wholeheartedly aspirational about living up to his communal values. So, he is responsible for his self in regard to music largely in virtue of his active engagements with living up to his communal values.[36]

Indeed, both Lily and Billy are aspirational, and their aspirations explain their responsibility for their selves in the relevant regard. What sets them apart is that one aims at achieving accomplishments in music primarily for music's sake, and another aims at the same things but for other things that they would further bring about. This difference in their aspirations explains how their selves are shaped differently.

Why is it important to explain that difference in how they shape themselves? In music and performance art more generally, you can hear and see on stage or in front of the camera what that difference may bring out. This is because a performer's job to bring to reality a composer's or a character's intention, emotions,and stories is comprehensively intertwined with the complexity of the performer's inner world. Their self is embedded in their work, and how their

---

36 Another way to understand their difference is to see it through counterfactual scenarios I mentioned I above—if the ways in which Billy acquires his communal values happened not to coincide with the ways in which the valuable in classical music and musicianship are acquired, he would not have formed the same personal policies as Lily. And this is not only because the goals embedded in their aspirations are different, but also because their different aspirations would have prescribed different methods for achieving the goals. I thank Luke Goleman for pointing out this alternative way of phrasing the difference. I also owe thanks to Caleb Dewey and Joe Metz for pushing me to clarify the difference.

self is shaped can be expressed through their work, sometimes inevitably so. But regardless of the profession, understanding an agent's aspiration helps us to give a fuller explanation of her responsibility for her self. Additionally, understanding the differences in aspirations among different people helps us to better understand the diversity among the selves and practical identities there are.

Last, let us consider Milly. Milly sets out to pursue music and musicianship in a similar way to Lily. But given her undiagnosed and untreated mental disorder, despite being able to aspire and trying to aspire, she is interrupted in a way that prevents her from acting from her policies, and thus is unable to actually aspire under the circumstances. Her self regarding music has not been shaped as much because of a lack of active engagements due to factors beyond her control. We can then say that she is not responsible for her self in regard to music and musicianship on the condition that she cannot aspire, at least for now. This assessment of Milly shows that the sufficient condition I offered further reveals a necessary condition for responsibility for one's self. Namely, to be responsible for aspects of one's self, it is required that one is able to aspire, and one's failure to do so is not through one's own faults. Although this necessary condition bears further elaboration, it deserves another paper to fully address it.

To conclude, I have argued that robust aspiration is sufficient for responsibility for one's self because it indicates an active engagement with an agent's own contribution to shaping her self. To actively engage with doing that, an agent governs her own motivation, reevaluates her current ends and values in accordance with what she deems valuable, so that she retains or rejects some of them, and acquires new ones. This is achieved by her forming, acting from or later revising her personal policies about that to which she aspires. Understanding the relation between aspiration and responsibility for the self helps us better understand when we are

responsible for what we are and what we have become. Further, it also informs us a of critical

way in which we can take responsibility for what we are and what we have become. Hopefully I

have shown that this is something we should all aspire to understand and appreciate.[37]

CHAPTER 3

Bringing the Deep Self Back to the Racecourse: Rethinking Accountability and the

Deep Self

Abstract

Deep self views of moral responsibility suggest that an agent fully satisfies the freedom condition for responsibility if and only if her actions or omissions issue from, and so express, her deep self. This involves a very strong condition for responsibility, and counterexamples proliferate. I defend a novel version of the deep self view by weakening its condition for responsibility in the accountability sense, while retaining the familiar core of deep self views. Indeed, an agent may be blameworthy for her wrongdoing without it issuing from, and so expressing, her deep self. And yet, I argue that she must have a deep self for which she is responsible to be blameworthy for her wrongdoing. This is achieved by paying closer attention to the historical dimension of the deep self than standard deep self views have.

Introduction

Deep self views remain among the most appealing contemporary theories of moral responsibility. Deep self theorists contend that an agent's deep self plays the grounding role in virtue of which an agent acts freely and is responsible for what she does (Frankfurt, 1971; Watson, 1975, 1987; Bratman, 1997, 2004, 2005; Shoemaker, 2015; Sripada, 2016). This is insightful. These views address the question that freedom understood in the traditional Hobbesian and Humean sense as in "lack of constraint" requires that a constraint arising from within one's own psychology be accounted for. They do so by providing resources to mark a division within an agent's self to capture her deep self—who she truly is. With the deep self

identified, when an agent acts freely her action is an expression of who she truly is. And when she acts freely, she is then a candidate for being accountability responsible, insofar as her action or omission is an expression of her deep self.[38]

But standard deep self views tend to involve a very strong condition for responsibility. Think about Susan Wolf's well-known Jojo (Wolf, 1987), who is severely indoctrinated by his evildoing father. Growing into an evil-doer himself, Jojo's morally objectionable behaviors issue from, and so express, his deep self. He thus counts as free and is responsible for his evildoing on these views. Nevertheless, given his turbulent and distorted personal history, one can reasonably question whether acting from a deep self really is sufficient for acting freely and responsibly.

Consider, again, someone who loves spicy food, but resolves to not eat spicy food for the sake of not irritating her stomach ulcer, who then freely and responsibly acts contrary to her deep self with regard to physical health, and puts too much red pepper powder in her meal out of weakness of will. One may question whether, after all, acting from a deep self is necessary for acting freely and responsibly. Despite renewed interest in deep self views (2016; Strabbing,

---

38 Note that not all deep self theorists claim to argue for accountability. While traditional deep self theorists such as Harry Frankfurt, Gary Watson and Michael Bratman argue for the purchase of an agent's deep self on grounding responsibility in the accountability sense, more recent development of the deep self view tends to focus only on responsibility in the attributability sense. According to these views, an agent is attributionally responsible for an action if and only if it stands in the right relation with her deep self, however the deep self or the right relation is understood. See Strabbing (2016b), Matheson, 2019, and Gorman (2019) for development in that direction. The purpose of this paper would be two-fold depending on which of these views my readers have in mind. The primary goal of my paper is to weaken the condition for accountability than traditional deep self views propose, so that it can explain cases in which an agent is blameworthy for a wrongdoing that is indeed *not* expressive of her deep self. But if my readers have in mind those recent deep self views that focus on attributability, and only attributability, then the secondary goal of my paper would be to offer a wider deep self view of moral responsibility to account for cases in which an agent can be blameworthy for her wrongdoing for which she is attributionally responsible. This is an advantage of my view because philosophers have argued that conditions for blameworthiness differ from conditions for attributability (Watson, 1996; Scanlon, 1998; Levy, 2005; Shoemaker, 2011; among others). Although different authors have different agendas on distinguishing blameworthiness from attributability, I take it as an underlying thought that judging an agent to be blameworthy concerns a broader set of facts about her than assessing her attributability does. These recent deep self views are thus ill-equipped to account for an agent's diminished blameworthiness in cases where she is indeed attributionally responsible.

2016b; Matheson, 2019; Gorman, 2019, 2022), these problems remain troubling, and deep self views have remained unpopular.

In this paper, I defend a novel version of the deep self view in which an agent must be responsible for her deep self to be responsible in the accountability sense for what she does, where the deep self on my view is understood as crucially expressed, albeit not exhausted by, her deep valuing and deep values.[39]

In what follows, I will lay the groundwork for my view in section 1. In it, I introduce what I take to be the deep self, and develop two novel claims regarding a long-neglected thesis of responsibility for the deep self. Drawing upon resources from my thesis on responsibility for the deep self, I present my deep self view of moral responsibility in section 2. I then illustrate my view with examples inspired by Susan Wolf's well-known case of Jojo in section 3. The examples I give will reveal the historical dimension unique to my view. In section 4, I further discuss the historical dimension, how it sets my view apart from standards views by absorbing Susan Wolf's criticism of those views, which will bring new life to the deep self view of moral responsibility.

1. Laying the Groundwork: Responsibility for the Deep Self

When we reflect on responsibility for what we do and its downstream consequences, one familiar line of thinking draws attention to how we are as persons who possess a particular set

---

39 Throughout the paper, I will discuss *value* both as noun and as active verb. This is because I believe that 1) an agent with a deep self is one who has things about her self and in the world that she deeply values, and 2) that which she values deeply are her deep values. The second claim concerns the relationship between *valuing* and *having what one values as one's values*. To further explore and defend this claim deserves a full project of its own, and I will have to leave that task to another paper. In the following discussions, I will sometimes drop either value as noun or value as active verb, depending on my emphasis on the discussion at hand.

of psychological traits. When our actions or omissions issue from, and so express, those traits, they are ours, and we are responsible for them.[40]

I develop a modified version of this approach: to be responsible in the accountability sense for what we do and all that we bring about, we must have had an unimpeded opportunity to develop and sustain a self-constituting ability[41] to fashion our selves, and thus be responsible for our selves. Furthermore, we must also retain the ability to deploy that self-constituting ability in order to modulate our behaviors.[42] This modification will help develop the intuitive idea that *to be responsible for what we do, we must first be responsible for who we are.* Despite lacking explicit development, this idea has long been shared by moral responsibility theorists from different camps (Wolf, 1990, 2015; Kane, 1999; Fischer and Ravizza, 1998; and Ishtiyaque Haji, 1998). As I will argue, developing a thesis of this long-neglected topic on *responsibility for the deep self* will reveal a far more refined relation between what we do responsibly and the deep self.

---

40 Views that take this approach most prominently consist of Harry Frankfurt's (1971) account on free will and hierarchical desires, Gary Watson's (1975, 1987) account on free agency and valuational system, and Michael Bratman's (1997, 2004, 2005) view on responsible agency and planning agency, among others. With renewed interest in this approach, philosophers such as David Shoemaker (2015) and Chandra Sripada (2016) argue for the condition for attributability in terms of an agent's cares and commitments, Jada Twedt Strabbing (2016b) offers a conjunctive sufficient and necessary condition for attributability in terms of an agent's judgments for normative reasons, Benjamin Matheson (2019) proposes an ideal narrator that connects an agent's moral identity in different person-stages that confers her attributability, and August Gorman (2019, 2022) puts forward a conjunctive sufficient and necessary condition for attributability in terms of an agent's partial and hypothetical approval for her behaviors were she to reflect on them.

41 Here, I leave it as an unsettled matter how this appeal to abilities trades in the dialectic between compatibilism and incompatibilism. My understanding of abilities is permissive between general abilities and specific abilities, and I take the self-constituting ability pertinent to responsibility for one's deep self to be an ability developed and retained by the agent over an extended period of time. When circumstances and opportunities obtain, the exercise of that ability would render an agent's relevant specific ability. But this is consistent with an agent's responsibility for her deep self when she does not exercise that ability. I thank Michael McKenna for pushing me to clarify this point.

42 See McKenna (2019: 10-12) for the discussion that inspires this point. And see Strabbing (2016b: 752-754) and Strabbing (2016a: 300-305) for a similar point, where she rightly points out the importance of possessing the responsibility relevant ability, rather than merely exercising it. Also see McKenna & Van Schoelandt (2015: 55-59) for their unprecedented effort in advancing a hybrid view of a mesh view (deep self view) of moral responsibility and a reasons-responsive view of moral responsibility, where they argue that the resources one can draw from one's psychological mesh (and the ability to do so) plays a critical role in understanding our moral identities.

1.1 The deep self

I take the deep self to be consisting of a set of psychological features that develop and persist *over time*, and they are expressed crucially through, albeit not exhausted by, an agent's *deep valuing* and *deep values*. So, a historical dimension of the deep self and a special kind of evaluative element are crucial to my understanding of the deep self. In addition, the focus on the historical dimension is relevant to this extra level of evaluation and reflection in understanding the deep self. Let me start with the evaluative element.

By deeply valuing something, I mean that an agent judges it to be good, and desires it primarily for those reasons for which she judges it to be good. In addition, she is susceptible to a range of emotions responsive to it. Here, these three aspects of deep valuing do not merely co-occur but relate to each other in a non-additive way. Specifically, they are connected by the reasons to which an agent is sensitive, and such a sensitivity need not be conscious, or in line with what is objectively or uncontroversially good. More importantly, that which she values in this way partly constitutes and crucially expresses her practical identity[43] in the realms of morality, society, aesthetics, and physical and mental well-being, albeit, possibly, in a disparate manner.[44]

---

43 Practical identity is a primitive of my view. I take it to be expressed by an agent's practical stances in the realm of morality, society, aesthetics, and other important realms of human life (albeit usually in a disparate way). The practical stances that a practical agent takes are explained by the multitude of her practical attitudes in the relevant spheres. Taking these practical stances make her the practical agents she is in the relevant realms of human life. But this is not to say that a practical stance that an agent takes is exhausted by her practical attitudes. Moreover, an agent's practical identity, though crucially expressed by her practical stances, is not exhausted by them, either. For instance, an agent must be able to put in practice her practical stances to *incorporate* her deep valuings into her practical identity *not* as means to incorporate her other valuings into her practical identity. So, a practical identity is not reduced to an agent's practical stances, practical attitudes, or deep valuings. I thank Carolina Sartorio's suggestion to clarify this point.

44 It is a vexing matter and so I mean to leave it as an unfinished philosophical project to fully state all the elements that bear on the constitution of one's practical identity. I thank Michael McKenna's suggestion to make this clarification.

Given how an agent may pursue and fulfill values in different ways in these different realms, an agent with a deep self does not have to have that deep self *as a whole*, and as a matter of fact, many of us do not. Instead, our deep valuings in different realms of human life express different parts of our practical identities and our deep selves. As a result, we may have a deep self in the realm of morality, but not one in the realm of aesthetics. Or, we may later develop a deep self in the realm of aesthetics but only with regard to, say, the aesthetics of food, but not regarding the aesthetics of fine art.

How deep is deep valuing? To compare, consider first Alfred Mele's characterization of valuing something in the *thin* sense, which involves a conjunction of a positive motivational element of desiring it and an evaluative element of judging it to be good (1995: 116).[45] For instance, consider Peta who desires to eat an ice cream sandwich because her friends bet that she would do it, or because it is the only thing left in the freezer and she craves snacks in the moment. Independently, she judges ice cream sandwiches to be good. However, she would rarely be motivated to go and get an ice cream sandwich without further enticement or under exceptional circumstances—like winning a bet or craving snacks with no other options besides an ice cream sandwich. She would count as thinly valuing ice cream sandwiches in Mele's sense because there is both positive motivational and evaluative components in her valuing ice cream

---

45 Mele further distinguishes between thinly valuing something that is of importance to an agent and her personal values as follows:

> "We can say that *S* at least *thinly values X* at a time if and only if at that time *S* both has a positive motivational attitude toward *X* and believes *X* to be good. Unfortunately, accepting this analysis does not settle what it is for something to be *among one's values*…Can we properly say that *X* is among a person's values if *X* is both valued by the person and of special importance to the person? No…[T]he range of personal values under consideration can be limited to things that are *valued by* valuers and are clear cases of the valuers' values" (Mele, 1995: 116).

I accept this distinction between thinly valuing something that is of importance to an agent and having it among her values. My view on deep valuing is an extension of what it is to have something among one's values. But as I stated in the previous footnote, to further develop and defend this idea is beyond the scope of this paper.

sandwiches. However, ice cream sandwiches do not matter much to Peta and her practical identity in the realms of aesthetics of food or cuisine culture—it does not speak to who she is in those realms in life; she is no true gourmand and connoisseur of ice cream.[46] Deep valuing is deeper than Peta's valuing ice cream sandwiches, because the deeply valued item constitutes and expresses the valuer's practical identity in relevant realms of human life.

Or consider Peter who also desires ice cream sandwiches; but unlike Peta, he desires to eat them on a regular basis primarily for those reasons for which he judges them to be good. Peter would also count as thinly valuing ice cream sandwiches in Mele's sense, and in this case, ice cream sandwiches mean more to Peter than to Peta. But deep valuing is still deeper than that. For Peter, ice cream sandwiches do not constitute what he is in the realms of aesthetics of food or cuisine culture: like Peta, he is no true gourmand and connoisseur of ice cream, either. Indeed, to deeply value ice cream sandwiches, among other things, an agent could be a true gourmand and connoisseur of ice cream sandwiches, who judges ice cream sandwiches to be good, desires to eat and learn about them primarily for those reasons for which she judges them to be good, and is susceptible to a range of emotions in the prospect of not having access to sustain her pursuit, for example. Her deeply valuing ice cream sandwiches makes her who she is in the realms of aesthetics of food and cuisine culture.

Though deep valuing is important to the deep self, I suggest that it does not exhaust the deep self. Indeed, there might be other explicit or implicit attitudes that an agent holds that are constitutive of who she is, but are not parts of her deep valuings and deep values. So, characterizing the deep self as crucially expressed by an agent's deep valuings does not mean

---

46 This, of course, does not mean that she must not be a true gourmand and connoisseur of, say, spices. Rather, what is important to note here is that we would not know about this aspect of her practical identity with regard to the aesthetics of food and cuisine culture by her thinly valuing ice cream sandwiches.

that there is a privileged set of psychological features that *just is* one's deep self, like many deep self theorists would have us believe.[47] Nevertheless, I suggest that it is important to focus on an agent with a deep self as someone who has something in the world that she deeply values. This is because the way in which she deeply values something and acquires her values is critical to our assessment of her responsibility for her deep self: does she have a personal history in which she was afforded the unimpeded opportunities (i.e., a stable and healthy household growing up, access to education and affordable health care, and just society, and so on) to develop and exercise the ability necessary for responsibility? If she does not have such a personal history, as I will soon argue, then she is not responsible for who she is. This leads to the historical dimension of the deep self, a dimension unique to my view (more on this in section 4).

No one is born with a deep self. The true gourmand and connoisseur of ice cream sandwiches, for instance, does not deeply value ice cream sandwiches from age one. An agent obtains critical aspects of her deep self as she comes to acquire[48] certain values from which she acts, and generally matures into a functioning social, moral, aesthetic, etc., being who has things in her life that she deeply values and is able to fulfill them. And it is from this stage in life where an agent starts to have a deep self, and becomes a *candidate* of being responsible for her self and her behavior. And if she does something morally objectionable, she is then a candidate of being held responsible for it.

1.2 Two kinds of deep self

47 For an insightful identification of this problem for a lot of deep deep self views, see Gorman (2022).
48 Here, I take acquiring a value to be not only involving seeing something to be good or valuable, but also seeing it to be good or valuable *to the agent.* By seeing something as good to her, an agent may desire it, judge it to be good either consciously or subconsciously, feel certain emotions towards it, or all of these things together. In addition, to acquire a value is to be disposed to live up to it. Given the time and opportunities she is afforded, she acts on what she values and fulfills her values. Depending on what her deep valuings and values are, they will then be incorporated to who she is as a practical agent in the relevant realms of human life.

As I suggested, becoming responsible for one's self involves various value-engagements—we fashion what we are when we are in the pursuit of acquiring, reevaluating, retaining or rejecting, and fulfilling values. So, there is an aspect to how we shape our selves that concerns end-setting and value practice. When we successfully do so, we become a slightly, moderately, or drastically different version of ourselves. Different philosophers have characterized this aspect of shaping one's self as *aspirational.*[49] In line with this characterization, I offer a novel distinction between an *actual* deep self and an *aspired* deep self. They differ from each other in two following respects.

First, an agent's actual deep self is crucially expressed by values that she has already acquired; they constitute what she is *now.* An aspired deep self is expressed by values that she is able to acquire, judges or deems[50] worth acquiring, but has yet to possess; they would constitute what she wants to become *in the future*.

Second, an aspired deep self is obtained through aspiration. And as I will touch on momentarily, aspiration to a different self involves a conative state where one actively engages in taking courses of action to realize that self. An actual deep self, in comparison, can be obtained through the processes of either passive engagement or active engagement. For instance, one may have one's actual deep self obtained in an unreflective or superficial manner. Or one may also obtain it in an actively engaged manner.

1.3 Responsibility and aspiration

---

49 See Taylor (1976), Schmidtz (1994), Kane (1996). For a more recent development of this idea, see Callard (2018, Chapters 2, 6).

50 Here, by deeming something as worth acquiring, I mean that an agent has not yet formed a judgment (consciously or not) either that it is good, or that it is good for her. But she may notice that she has a desire to learn about it and to take it as her own, or she may have a vague impression that it is good from other people's testimony (cf. Callard, 2018), among other things.

With these two kinds of deep self in mind, I suggest that when an agent aspires to a different deep self—namely, an aspired deep self, from her actual deep self, it involves an active engagement on the agent's part. This active engagement manifests her own contribution to the making of her deep self in two following respects.

First, an agent's aspiration to a different self involves the employing of her internal sensitivity and an external awareness to values she currently has and values different from hers. Employing this sensitivity and awareness will lead an agent to revise, retain or reject her current values, and acquire new values. It thus marks an agent's own doing in the forming of her actual deep self.

Second, an important component of actively forming one's deep self is to realize one's own conception of one's self. Aspiration to a different deep self is prominent in this process. It helps an agent to understand what she is and what she wants to become. Moreover, it helps her realize her own conception of her self by leading her to actively take actions with various value-engagements. This also marks an agent's own contribution to the shaping of her actual deep self.

Connecting aspiration and responsibility complements the intuitive idea that one is responsible for something if one contributes to the making of it with active engagement.[51] Indeed, aspiration to a different self specifies an important way in which a responsible agent contributes to the making of her self with active engagement, and thus makes her responsible for her self. But is aspiration the only way to shape an agent's self? what if she never aspires? Indeed, some people shape who they are in a passive way; their self-shaping involves not an active effort, but rather, only the possession of an ability to do so. As I will soon lay out, the possession of an ability to aspire is partly necessary for responsibility for who we are.

---

51 See Kane (1996) for discussions of this idea.

1.4 Responsibility for the deep self

Let me put forward two novel claims about responsibility for one's deep self where both trade in

the two kinds of deep self discussed above.

Sufficient An agent is responsible for her actual deep self if it is obtained through her

aspiration to transform her previous deep self to an aspired deep self, in which case the

previously aspired deep self is now her actual deep self.[52]

Necessary An agent is responsible for her actual deep self only if she possesses the ability

to aspire to transform her actual deep self to an aspired deep self, given the time and

opportunity to do so. In addition, her not exercising this ability is by her own making, not

something beyond her control.

Offering a sufficient condition and a necessary condition as two separate principles has

important implications. First, an agent may not have aspired to a deep self, and yet she may still

be responsible for her actual deep self on the condition that she is able to so aspire, given the

time and opportunity afforded, but does not exercise that ability through her own making. For

---

52 Here, I acknowledge that there is a challenge from manipulation and brain engineering of the sort such that an
    agent might come to aspire as a result of such manipulation and brain engineering. In that case, it is argued that
    such manipulative causes can be responsibility-defeating. See Mele (1995) for relevant discussions. These
    challenges will not be addressed in this paper, and for simplicity, I will leave the condition of non-manipulation
    implicit throughout.

    I also acknowledge that there is another challenge from the regress problem, in that responsibility for an actual
    deep self must be consequences of previous actions for which an agent is responsible. But if this is the case,
    who is responsible for the self that issued those actions for which the agent is responsible? This could go on and
    on. For relevant proposals in response to this problem, see Kane (1999: Chapter 5) and Callard (2008: Chapter
    5, 6). Although significant, I will not explore this topic in this paper.

example, an agent who obtains her first deep self in a passive and unreflective manner, and stays idle with it despite the time and opportunity given to her to examine and shape her self for better or for worse, would still be responsible for her deep self in the way identified.

Second, an agent may have the necessary ability to aspire, and is afforded the time and opportunity to do so, yet she may not satisfy any complete set of sufficient conditions. For example, suppose she is severely indoctrinated the extent to which there is no control left in her, in that case, her not exercising her ability to aspire is brought about by things beyond her control. She is then not responsible for her deep self.[53]

These two claims about responsibility for one's deep self and the conceptual space left by them will show their significance in advancing my deep self view of moral responsibility. They will provide indispensable explanatory power for an agent's responsibility, and in particular, her diminished blameworthiness for her wrongdoing in cases I will focus on.

2. A Novel Deep Self View of Moral Responsibility

In my modified approach to understand moral responsibility in terms of the deep self, I suggested that to be accountability responsible for what we do and all that we bring about, we must have had an unimpeded opportunity to develop and sustain a self-constituting ability to fashion our selves, and thus be responsible for our selves. To complete this approach with details from my two claims about responsibility for the deep self—namely, the aforementioned self-constituting ability pertinent to responsibility for one's deep self is the ability to aspire to a different deep self —I now offer a new formulation for the deep self view of moral responsibility as following.

---

53 A complete theory of responsibility for one's deep self would fill the gap in the latter case; however, here I only mean to argue for a more modest theory, one that advances one sufficient condition for attributional responsibility for one's deep self, and one that advances a distinct necessary condition. I thank Michael McKenna for pushing me to clarify this point.

NewDS An agent acts freely and is morally responsible for what she does that renders her blameworthy for her wrongdoing only if she possesses a deep self for which she is responsible. Responsibility for her deep self in turn requires that an agent was afforded an opportunity to develop and exercise an ability to aspire to a different deep self. As she acts freely and responsibly, she retains the ability to deploy such an ability to draw upon resources from her deep self to regulate her behaviors.

My formulation turns on a far more refined relationship between the deep self and responsible agency than standard views do, while sustaining the explanatory power of the deep self. To explain, consider four important implications that my formulation has on understanding moral responsibility, and in particular, blameworthiness.

First, when an agent actively shapes her deep self by performing courses of action as she exercises the ability to aspire, she acts freely and is morally responsible for those actions *in virtue of* actively shaping and taking responsibility for her deep self. If those actions involve wrongdoing, then she is blameworthy for them.[54] This, in turn, reveals a more refined sufficient condition for responsibility for what one does drawing upon resources from how one shapes one's self—to the extent that one is actively taking actions to shape one's self by exercising one's ability to aspire, one is responsible for them. Specifically, if theses actions involve wrongdoing, she would be blameworthy for them.

---

54 It has been argued that conditions for blameworthiness can be met without the occurrence of any wrongdoing. For some of the relevant discussions, see Khoury (2011) and Capes (2012). It is beyond the scope of this paper to discuss instances of blameworthiness without wrongdoing. Instead, I will focus only on cases of blameworthiness for wrongdoings.

Second, an agent may act freely and responsibly in performing some act *A* and she might also be responsible for her deep self by exercising the ability to aspire. Nonetheless, the performing of *A* does not involve such an exercising of that ability. She thus may be responsible for *A* without it being a product or expression of her actively fashioning her self. Specifically, if *A* involves a wrongdoing, she is blameworthy for it. Indeed, such a free act, *A*, might be incongruous with the manner in which she actively shapes her deep self. A weak-willed action is an example of such a free act *A*. If the weak-willed action involves a wrongdoing, she is blameworthy for it.

Third, an agent may act freely and responsibly, but she has not previously actively engaged in fashioning her deep self by exercising the ability to aspire. For example, she acquires characteristics of her first and current deep self in a passive and unreflective way, and remains idle despite the time and opportunity given to her to actively shape her self, for better or for worse. Nevertheless, she is responsible for her deep self by possessing the ability to aspire, and retaining the ability to deploy that ability to draw upon resources from her deep self to regulate her behaviors, given that she has access to do so. She thus may be, as with the last example, responsible for what she does in the sense that renders her blameworthy if it involves any wrongdoing, despite the fact that it is not a product or expression of her actively shaping her deep self.

Fourth, an agent may possess the ability to aspire but does not exercise it, and her not exercising it results from a responsibility-defeating condition arising through no fault of her own. In that case, she is not responsible for her deep self. As I will illustrate in the next section, if an agent commits any wrongdoing that is nonetheless expressive of her deep self, her not being responsible for her deep self diminishes her blameworthiness for it.

3. Jojo, Dodo, Momo

Consider Susan Wolf's Jojo (1987). Jojo grew up with his dictator father who is an evildoer. He was raised to idolize his father and grew into a person just like him—he desires evildoing, endorses these desires wholeheartedly, and he acts in accordance with values he has acquired learning from his father. According to Wolf, Jojo has a deep self expressed by a value system he has adopted. His evildoing expresses who he is.

Now, allow me to fill in more details into this example. Suppose that the original Jojo is someone like the following.

> *The OG Jojo* Despite going through powerful indoctrination, Jojo still develops an ability to employ his internal sensitivity and external awareness to values including those different from his own. However, he is blocked from having access to learn those different values, or to see tensions among his own values. This is because the indoctrination has rendered his way of living and being so fixed that he is rarely and only superficially presented with values different from his own. In addition, the indoctrination deprives him of all the relevant knowledge with which he can recognize values different from his own *as values.* He feels content with himself and his life, and never exercises his ability to employ his sensitivity and awareness.

Given the conceptual space left by *Necessary* and *Sufficient*, we can acknowledge that despite having characteristics of a deep self—namely, he has things in life that he deeply values and is able to live up to them, Jojo is not responsible for his deep self. Despite possessing an ability to

aspire to a different self, he lacks access to exercise that ability or gain relevant knowledge to do so, thus cannot exercise that ability due to lack of conditions. Given *NewDS*, Jojo's blameworthiness for his evildoing is diminished on the condition that he is not responsible for the deep self he has, even though his evildoing is indeed expressive of who he is.

Now suppose Jojo has a triplet sister, Dodo, another evildoer who has gone through the same indoctrination process, but has since been assigned to deal with affairs that involve frequent interaction and cooperation with people from different backgrounds, and thus has been directly exposed to values different from her own.

The People Person Dodo Like Jojo, Dodo too develops an ability to employ her internal sensitivity and external awareness to values including those different from her own, to tensions among her values or between her values and those different from her own. But unlike Jojo, she has been consistently presented with different values from her own. Besides being presented with them, she is involved in activities that are interactive and communicative to the effect that she cannot avoid recognizing that there exist values different from her own, and that there are tensions between those values and her own. Indeed, she recognizes these things. Nevertheless, at no point did she ever contemplate the differences between the two set of values, and no change of heart ever took place.

Given *Necessary*, we can say that Dodo *is* responsible for her deep self on the condition that she possesses an ability to employ her internal sensitivity and external awareness. More importantly, despite never exercising this ability, she is afforded access to do so, and is thus able to aspire to a different self, for better or for worse. And her not exercising that ability is through

her own making. Given *NewDS*, her responsibility for her deep self renders her blameworthy for her evildoing no less than any other.[55]

Suppose Jojo and Dodo has a triplet brother, Momo, yet another evil-doer in the family, who has gone through the same indoctrination process, and has been assigned, alongside Dodo, to deal with affairs that involve frequent interaction and cooperation with people from different backgrounds. He thus has been directly exposed to values different from his own just like Dodo.

The Inhibited Momo Despite being involved in dealing with affairs that consistently put him in exposure to values and reasons for actions different from his own, Momo is so fixed in his way of living and being that he can barely recognize values different from his own, or tensions among his own values or between his and others'. This is because the indoctrination has inhibited him from developing an ability to employ an internal sensitivity and an external awareness to values in the first place. He goes through the motions as he interacts and does business with those who are different from him. Never at any point did he question his siblings' evildoing or his own.

In Momo's case, his lack of ability to aspire to a different deep self is more straightforward in that he did not get to develop it in the first place. Given *Necessary*, he is not responsible for his deep self on the condition that he is simply not able to aspire. Given *NewDS*, Momo's not being responsible for his deep self renders him less blameworthy, perhaps even blameless, for his evildoing.

---

55 Again, this diagnosis of Dodo would coincide with Strabbing's diagnosis of instances of attributional responsibility. According to her Judgment Responsiveness View (JRV) (2016b: 744), an agent's being attributionally responsible for her action needs not be responsive to reasons that are *correct*. But also notice that Strabbing's JRV cannot differentiate Jojo from Dodo in terms of their difference in blameworthiness.

The skepticism raised by Wolf (1987, 1990) facilitated by the case of Jojo as it is originally displayed, is meant to show that merely having a deep self and acting in accordance with it is not sufficient for responsibility. Thus, the deep self view is problematic. Nevertheless, I argue that the original Jojo case would only work in favor of Wolf's criticism if we understand Jojo's way of living and being as fixed as it is for the triplet brothers Jojo and Momo, but not Dodo. Having Dodo's case specified helps to reveal important features in an agent's *personal history* relevant to her free and responsible agency overlooked by standard deep self theorists. Namely, having a history of powerful and thorough indoctrination is one thing, having a history of powerful and thorough indoctrination *that left an agent with no opportunity to develop or exercise her ability to shape her deep self* is another. The case with Dodo shows that the former does not entail the latter. And it is the latter that matters to our assessment of blameworthiness, and more broadly, accountability. I turn to further explore this point now.

4. History Matters

History is crucial to my view. It sets an important difference between *NewDS* and standard deep self views in the literature.

To explain, the historical dimension on my view is two-fold. First, an agent develops and sustains her deep self *over time* as she acquires her values and realizes her deep valuing. Suppose in becoming responsible for her deep self, an agent obtains her actual deep self at t_1, aims at realizing an aspired deep self (or does not exercise her ability to do so through her own making) at t_2, and successfully realizes her aspired deep self (or does not do so) at t_3. Then active engagement through aspiration (or passive engagement through her own making) connects her deep self from t_1 to t_3.

Now suppose at t_2, while the agent is able to aspire to a different deep self and is going to aspire, she undergoes covert manipulation. As a result, a new set of desires, values, cares and commitments, judgments about normative reasons, and self-governing personal policies is implanted in her. More so, it effectively dominates her mental life, which involves but are not exhausted by a change regarding how much weight she gives to the same matters. For example, she may now give zero weight to matter *B* to which she used to give above zero weight.

On my view, she then no longer aims to aspire to a pre-manipulation deep self. Instead, she now either aspires to a post-manipulation deep self that she never would have aspired to were she not manipulated, or ceases to aspire altogether.[56] Because her active engagement with shaping her deep self before manipulation is disrupted through external manipulative influences, and these influences bypass her ability to evaluate, retain, revise or reject aspects of her self drawing upon resources from her deep self at the time, then when she aspires, her active engagement can no longer connect her deep self from t_1 to t_3. In that case, she ceases to be responsible for her deep self in virtue of being so covertly manipulated (at least for now). Given that she is not responsible for her deep self after such manipulation (at least in the current stage), she is not responsible for actions issued from that self. If she does something morally objectionable, her blameworthiness is diminished (at least for now).

This aspect of the historical dimension on my view concerns an agent's diachronic moral and practical identity, which differentiates my view from a number of familiar deep self views that do not require diachronic identity (Frankfurt, 1971; Watson, 1975; Sripada, 2016; Gorman, 2019). On those views, what matters for moral responsibility is whether an agent identifies with certain elements in her psychological make-up, whether that results in a mesh between second

---

56 For an example like this, see Mele's Ann and Beth (Mele, 1995: 145).

and first order desires or between an agent's motivation and her values, or whether the identification is understood as less wholehearted than with full wholeheartedness. According to those views, were an agent to undergo covert manipulation described at t_2, as long as she identifies either wholeheartedly or partially with the newly implanted and dominating set of psychological elements and acts in accordance with them, she acts freely and is morally responsible for her actions.

Although those deep self views that do account for an agent's diachronic identity share this first aspect of the historical dimension of my view, such as Bratman's cross-temporal self-governing policies that connect an agent's practical identity over time (Bratman, 1997, 2004, 2005), or Matheson's ideal narrator providing narrative explanations that confer psychological connectedness between different stages of a person (Matheson, 2019),[57] these views do not always share the second aspect of the historical dimension of my view, to which I turn now.

The second aspect of the historical dimension of my view further requires that an agent act freely and responsibly only if she was afforded an unimpeded opportunity in her personal history to develop and exercise the ability to form her deep self.[58] On my view, Dodo was afforded such an opportunity, but Jojo and Momo were not. And this is why the assessment of their blameworthiness differs.

Without an explicit endorsement of the historical dimension, it is not apparent on many standard deep self views that there be a difference in assessment of blameworthiness of Dodo from Jojo and Momo—as along as the triplets act from the deep self with which they identify,

---

57 Although, note that Matheson (2019: 469) argues that when emotions are involved in the ideal narrator's explanation of an agent's actions, the explanation is essentially *dispositionally* diachronic, rather than temporally diachronic, thus is not inherently historical.

58 For a relevant rationale on a positive historical thesis, see McKenna (2012: 170).

fully or partially, actually or hypothetically, either via a hierarchy of desires, what they value, care, or their self-governing policies, they are free and morally responsible for their evildoing.

The lack of internal resources in many standard views to account for the diminished blameworthiness of impaired agents like Jojo and Momo may mislead us to fittingly withhold emotions towards them. This brings us back to Wolf's criticism of deep self views. As she points out, acting in accordance with one's deep self might not be sufficient for responsibility, because whether one is responsible for one's deep self factors into our judgment of their responsibility for what they do that is expressive of that self, as quoted below.

[W]e sometimes do question the responsibility of a fully developed agent even when she acts in a way that is clearly attributable to her real self. For we sometimes have reason to question an agent's responsibility for her real self. That is, we may think it is not the agent's fault that she is the person she is—in other words, we may think it is not her fault that she has, not just the desires, but also the values she does (Wolf, 1990: 37).

Some features in an agent's deep self may indicate her non-responsibility. For instance, Wolf argues elsewhere that responsibility for what one does issuing from one's deep self requires *sanity*. As she defines it, sanity is understood "as the minimally sufficient ability cognitively and normatively to recognize and appreciate the world for what it is" (Wolf, 1987: 56). Further, sanity enables an agent to know the difference between right and wrong, and to correct her behaviors and improve herself accordingly (60). An agent is not morally responsible for what she does if she is unable to do so—namely, if she is insane. So, merely acting in accordance with her

deep self is not sufficient for her responsibility for it.[59] It is further required that the deep self from which she acts be sane.

It is not explicit in Wolf's view that sanity is required by responsibility for one's actions or omissions *because* it is further required by one's responsibility for one's deep self. Nonetheless, I argue that it is so: sanity indicates that one has developed the necessary ability to be responsible for one's deep self. On my view, the necessary ability is the ability to aspire to a different deep self. I now suggest that my view can accommodate the sanity requirement on the condition that it be understood in a particular reading.

The sanity requirement says that an agent with a deep self is sane only if she is able to know the difference between objective right and wrong in the world. Knowing the difference between right and wrong can be understood in the following two readings. In the stronger reading, knowing the difference amounts to understanding and appreciating the right *as* right, and the wrong *as* wrong. Such an understanding and appreciation could potentially lead an agent to act accordingly. In the weaker reading, knowing the difference merely amounts to realizing that there is a difference between right and wrong, without involving the further understanding and appreciation of the right *as* right, the wrong *as* wrong.[60]

It is unclear that this distinction is implied in Wolf's sanity requirement. For, to the extent that Wolf's original Jojo is concerned, the sanity requirement can simply show that he is not responsible for his evildoing because he is unable to know right and wrong in the stronger reading—he does not know right as right, wrong as wrong. However, to make sense of the difference in Jojo's and Momo's blameworthiness from Dodo's, this distinction is important. Their difference does not lie in their ability to know in the stronger sense—none of them know

---

59   See both Watson (1996: 240) and Scanlon (1998: 192, 279) for a similar idea.
60   I thank Yili Zhou for her discussions with me on this point.

right as right, wrong as wrong. Rather, the difference lies in their ability to know in the weaker sense—only Dodo knows in the weaker sense. Recall that she is able to employ her internal sensitivity and external awareness to her values and values different from hers. In addition, she is afforded access to values different from her own to an extent that the successful proceedings of her social and work life render her a recognition that there is such a difference between what she deems right and what others deem right. And yet, she does not truly understand or appreciate the difference as an indication that her evildoing is wrong.

If we want to capture the difference in blameworthiness between the triplets by the sanity requirement, then it cannot merely mean the ability to know right and wrong in the stronger reading, as someone like Dodo is able to know in the weaker reading without knowing in the stronger reading. I thus suggest that responsibility in the accountability sense requires the ability to know right and wrong *in the weaker reading*.

The ability to aspire critical to my view involves sanity in the weaker reading, but not in the stronger reading. This is because one does not need to know the objective right as right and objective wrong as wrong to aspire to a different self, but one does need to possess at least some moral knowledge in the weaker sense to aspire. The aspired deep self is not constrained exclusively by objective standards of right and wrong. Indeed, one can be blameworthy for degrading oneself as long as one possesses the ability to aspire to, say, a more satanic self, and is afforded access to exercise this ability to deepen one's evildoing. As I have shown, Dodo is someone like this. *NewDS* absorbs Wolf's criticism to familiar deep self views and accommodates Wolf's sanity requirement to account for what is necessary for accountability. All of this is achieved by focusing on features in an agent's personal history where the necessary ability for accountability is developed and exercised given access to relevant opportunities.

To conclude, I have developed a novel deep self view of moral responsibility in which an agent's responsibility for her deep self is necessary for her responsibility in the accountability sense for what she does, and more specifically in the cases I have focused on, her blameworthiness for a wrongdoing that is either expressive or not expressive of that self. My account retains the familiar core of standard deep self views in which an agent's deep self is essential to our understanding of her moral responsibility. Beyond that, my view has the internal resources to account for cases in which we deem it justified to blame an impaired agent less. This has been achieved by paying closer attention to the historical dimension of the deep self than standard deep self views have, which I hope has brought new life to the deep self view.[61]

REFERENCES

Bratman, Michael (1987). *Intention, Plans, and Practical Reason*. Cambridge: Cambridge, MA: Harvard University Press.

Bratman, Michael E. (1996). Identification, Decision, and Treating as a Reason. Philosophical Topics 24 (2):1-18.

Bratman, Michael E. (1997). Responsibility and planning. *The Journal of Ethics 1 (1)*:27-43.

Bratman, Michael E. (2000). Reflection, planning, and temporally extended agency. *Philosophical Review* 109 (1):35-61.

Bratman, Michael E. (2003). Autonomy and hierarchy. *Social Philosophy and Policy* 20 (2):156-176.

Bratman, Michael E. (2004). Three Theories of Self-Governance. *Philosophical Topics* 32 *(1/2)*:21-46.

Bratman, M. (2005). Planning agency, autonomous agency. *Personal autonomy*, 33-57.

Bratman, Michael E. (2007). *Structures of Agency: Essays*. Oxford University Press.

Callard, Agnes (2018). Aspiration: The Agency of Becoming. Oup Usa.

Capes, Justin A. (2012). Blameworthiness without wrongdoing. *Pacific Philosophical Quarterly 93 (3)*:417-437.

Dover, Daniela (2022). The Conversational Self. *Mind* 131 (521):193-230.

Dworkin, Gerald (1981). The Concept of Autonomy. *Grazer Philosophische Studien* 12 (1):203-213.

Frankfurt, Harry G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy 68 (1)*:5-20.

Frankfurt, Harry G. (1977). Identification and externality. In Amelie Rorty (ed.), *The Identities of Persons*. University of California Press.

Frankfurt, Harry (1987). Identification and Wholeheartedness. In Ferdinand David Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge University Press.

Fischer, John Martin & Ravizza, Mark (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge University Press.

Gorman, August (2019). The Minimal Approval View of Attributability. In David Shoemaker (ed.), *Oxford Studies in Agency and Responsibility 6*. Oxford University Press.

Gorman, August (2022). Demystifying the deep self View. *Journal of Moral Philosophy 19 (4)*:390-414.

Haji, I. (1998). Moral Appraisability: Puzzles, proposals, and perplexities. Oxford University Press.

Harman, G. (1993). Desired desires. *Value, welfare, and morality*, 138-157.

Kane, Robert (1996). The Significance of Free Will. Oxford University Press USA.

Khoury, Andrew C. (2011). Blameworthiness and Wrongness. Journal of Value Inquiry 45 (2):135-146.

Levy, Neil (2005). The Good, the Bad, and the Blameworthy. *Journal of Ethics and Social Philosophy* 1 (2):1-16.

Matheson, B. (2019). Towards a structural ownership condition on moral responsibility. *Canadian Journal of Philosophy, 49(4)*, 458-480. doi:10.1080/00455091.2018.1480853

Mele, Alfred R. (1995). Autonomous Agents: From Self Control to Autonomy. Oxford University Press.

McKenna, Michael (2012). Moral Responsibility, Manipulation Arguments, and History: Assessing the Resilience of Nonhistorical Compatibilism. *The Journal of Ethics 16 (2)*:145-174.

McKenna, M., & Van Schoelandt, C. (2015). Crossing a Mesh Theory with a Reasons-Responsive Theory: Unholy Spawn of an Impending Apocalypse or Love Child of a New Dawn?. In *Agency, freedom, and moral responsibility* (pp. 44-64). Palgrave Macmillan, London.

McKenna, M., Coates, D. J., & Tognazzini, N. (2019). Watsonian compatibilism. *Oxford studies in agency and responsibility, 5*, 5-37.

Kay Nelkin, D. (2020). Responsibility, Reflection, and Rational Ability. *The Monist, 103(3)*, 294-311.

Scanlon, Thomas (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press.

Schmidtz, David (1994). Choosing ends. *Ethics* 104 (2):226-251.

Shoemaker, David (2011). Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility. *Ethics 121 (3)*:602-632.

Shoemaker, D. (2015). Ecumenical attributability. *The nature of moral responsibility: New essays*, 115-40.

Smith, A. M. (2005). Responsibility for attitudes: Activity and passivity in mental life. *Ethics*, *115*(2), 236-271.

Sripada, Chandra Sekhar (2010). The deep self Model and asymmetries in folk judgments about intentional action. *Philosophical Studies 151 (2)*:159-176.

Sripada, Chandra (2016). Self-expression: a deep self theory of moral responsibility. *Philosophical Studies 173 (5)*:1203-1232.

Strabbing, J. T. (2016a). Attributability, weakness of will, and the importance of just having the capacity. *Philosophical Studies, 173(2),* 289-307.

Strabbing, J. T. (2016b). Responsibility and judgment. *Philosophy and Phenomenological Research, 92(3)*, 736-760.

Strawson, Galen J. (1994). The impossibility of moral responsibility. *Philosophical Studies* 75 (1-2):5-24.

Taylor, C. (1976). Responsibility for self. *The identities of persons*, 281-99.

Watson, Gary (1975). Free agency. *Journal of Philosophy 72 (April)*:205-20.

Watson, Gary (1996). Two Faces of Responsibility. *Philosophical Topics* 24 (2):227-248.

Watson, Gary (1987). Free action and free will. *Mind 96 (April)*:154-72.

Wolf, Susan (1987). Sanity and the Metaphysics of Responsibility. In Ferdinand David Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge University Press. pp. 46-62.

Wolf, Susan (1990). *Freedom Within Reason*. Oup Usa.

Wolf, Susan (2015). Character and Responsibility. *Journal of Philosophy 112 (7)*:356-372.