AGONY AND INTEGRITY:
AN EROTIC PSYCHOLOGY FOR PREFIGURATIVE ETHICAL PRACTICE

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Philosophy

By

Christian Marchal Golden, B.A.

Washington, D.C.
April 25, 2012
The goal of this dissertation is to clarify and address a problem of integrity understood as the question of how one ought to live out one’s robustly motivating and normatively orienting attachments to persons, projects and ideals. My overall strategy is to consider two broad approaches to this question—one post-Kantian, one post-Nietzschean—arguing in favor of the latter as a superior account of the normative and psychological character of personal practical integrity.

To do so, I develop and defend an account of power as a productive, excessive and dynamic range of effects and relations that enter into the formation of normatively responsive human subjects and so condition our capacities for living out practical commitments. I call this view of subject formation and agency Agonistic Realism (AR) and contrast it with a post-Kantian strategy, which I call Sovereign Idealism (SI). AR is distinguished by its emphasis upon the role and value of contestation in the psychic and social lives of agents. I will argue that contestation over the meaning and value of our ends of action makes possible the experience of loss and transformation through which our committed activity remains vitally dynamic. SI elides these phenomena, and so underwrites an ideal of committed activity that denigrates them, whereas AR underwrites an ideal privileging them. Since AR does so by foregrounding psychic and social dynamics of contestation, the ideal of integrity it supports prescribes living out our deep commitments agonistically.
The conception of “agonistic integrity” for which I argue thus embraces three related hypotheses. First, post-Kantian ideals tend to presuppose the figure of “the horizon” as a pregiven or “background” source of normative authority, and so tend to construe personal integrity as steadfast adherence to demands already established, thereby eliding and undervaluing the role of loss and transformation in our committed activity. Second, recognizing the role and value of loss and transformation is a condition of adequacy for an ideal of how to live out our commitments. And third, an ideal of committed action grounded in an account of power’s integral role in subject formation and agency better satisfies this condition.
ACKNOWLEDGEMENTS

My deepest gratitude belongs to the members of my dissertation committee: Mark Lance, Gerald Mara, Mark Murphy and Terry Pinkard. They have always patiently and generously helped guide and inspire my efforts to emulate the intellectual virtues they model so effortlessly well. The support they have shown for my work and the contribution they have individually and collectively made to my intellectual development is immeasurable.

Special thanks are due to Wayne Davis and the members of the Graduate Committee for granting me the resources to see this project through.

I have also benefitted tremendously from the attentive care of a number of close colleagues in the Department of Philosophy at Georgetown, without whose unwavering support and encouragement I could not have completed this project. They are David Bachyrycz, Andrew Blitzer, Matt Burstein, Jeff Englehardt, Bryce Huebner, Maggie Little, Chauncey Maher, James Mattingly, James Olsen, Mark Pitlyk, Karim Sadek and Linda Wetzel. As friends and conversation partners, they have made Georgetown a place I have felt proud and privileged to call home for nearly a decade.

My gratitude also belongs to my friends and family outside of Georgetown, especially Michele and Gino Ballarin, Perry Davis, Andee Bingham, Bryan Buchanan, Andrew Butler, Carter Felder, Eileen Goode, David, Ginny and Richard Hansen, Brian Henriksen, Rowan Hildreth, Rachel Lipsey, John Marshall, Katherine Sydor, Mark and Anna Southworth, Joshua Stephens and Dustin Umberger. They have buoyed my spirits
and greatly enriched my life, and so have helped me to get the most out of the time I have spent at Georgetown.

Finally, I owe sincere thanks to the faculty of the departments of Philosophy and Religious Studies at the University of Virginia whose intellectual energy and integrity first drew me to philosophy. They especially include Talbot Brewer, John Marshall, Charles Matthewes, M. Jamie Ferreira and Peter Ochs.

The positive contribution to my life made by these extraordinary people is incalculable. This work is fondly dedicated to them and to the memory of Edward Robert Golden, Dee Dee Sharpe and Joseph and Lucille Lagler.
# TABLE OF CONTENTS

## Chapter I
Introduction: Taking Integrity Seriously .................................................................1
1: A Psychopathology of Agency .............................................................................1
2: Commitment and Integrity .................................................................................3
3: Erotic Ties, Erotic Ideals ......................................................................................7
4: Horizons of Erotic Agency ..................................................................................9
5: Taking Ourselves Seriously ..............................................................................12
6: Playing with Our Selves ...................................................................................29
7: Taking Things Seriously: Serious Play ...............................................................40

## Chapter II
Sovereignty of the Will: Christine Korsgaard’s Rational Psychology of Action ..........49
1: Introduction: Korsgaard’s Project in Outline .....................................................49
2: Normative Criticizability, Volitional Control, Psychic Wholeness ....................51
3: Responsibility & Authority ...............................................................................55
4: Korsgaard’s Psychology of Action ......................................................................62
5: Phenomenological Foundations ........................................................................77
6: Sovereignty & Horizonality ...............................................................................89
7: Conclusion: Three Theses ...............................................................................96

## Chapter III
Power: Genealogical Resources for a “Realist” Erotic Psychology .........................103
1: Introduction: Power as a Problem of Erotic Psychology ....................................103
2: Power’s Horizons ...............................................................................................107
3: The Conception .................................................................................................111
3.1: Productive Power: Genealogies of Subject Formation ..................................117
4: Conclusion: Against Shared Horizons ..............................................................162
| Chapter IV |
| Beyond Shared Horizons: Contested Practices, Practices of Contestation |
| 1: Introduction |
| 2: Social Horizons in Charles Taylor’s *Ethics of Authenticity* |
| 3: Heideggerian Horizons: “The Odd Stray” and Normative Breakdown |
| 4: Power’s Perspectives: Flexible, Dynamic Horizons |
| 5: Ambiguous Horizons of Action & Identity: Butler on Gay Self-Definition in the United States Military |
| 6: Love and Loss |
| 7: Conclusion: Paradigmatic Crisis |

| Chapter V |
| “Beyond Your Horizon”: Towards an Erotic Psychology of Agonic Integrity |
| 1: Introduction: Integrity’s Renewed Prospects |
| 2: Beyond One’s Horizons: Integrity as an Agonistic Ideal |
| 3: Against One Self: The Soul’s Plurality |
| 4: Against Oneself: Agonistic Self-Cultivation as an Erotic Ideal |
| 5: Worries |
| 6: Illustrations of Agonistic Integrity: Literature as Self-Overcoming |
| 7: Illustrations: Bodily Confrontation |
| 8: Is My View of Integrity “Horizonal”? |
| 9: Conclusion: Underground Overmen? |

| Bibliography |
**LIST OF ABBREVIATIONS**

All references to the writings of Friedrich Nietzsche have been cited in footnotes to the text according to the following abbreviations. In each citation, the part of the original text is identified with a roman numeral, followed by the original section number.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Uses”</td>
<td>“On the Uses and Disadvantages of History for Life”</td>
</tr>
<tr>
<td>BGE</td>
<td><em>Beyond Good and Evil</em></td>
</tr>
<tr>
<td>HATH</td>
<td><em>Human, All Too Human</em></td>
</tr>
<tr>
<td>GM</td>
<td><em>On the Genealogy of Morals</em></td>
</tr>
<tr>
<td>A</td>
<td><em>The Anti-Christ</em></td>
</tr>
<tr>
<td>EH</td>
<td><em>Ecce Homo</em></td>
</tr>
<tr>
<td>TI</td>
<td><em>Twilight of the Idols</em></td>
</tr>
<tr>
<td>GS</td>
<td><em>The Gay Science</em></td>
</tr>
<tr>
<td>Z</td>
<td><em>Thus Spoke Zarathustra</em></td>
</tr>
</tbody>
</table>

All references to works written by other authors contain the author’s last name and the cited work’s year of publication. Both of Judith Butler’s works cited in the text were published in 1997. *Excitable Speech* is designated with “(1);” *The Psychic Life of Power* is designated with “(2).”
It ought to make us feel ashamed when we talk like we know what we’re talking about when we talk about love.

Raymond Carver, 1981
CHAPTER ONE

INTRODUCTION: TAKING INTEGRITY SERIOUSLY

One should not be deceived: great spirits are skeptics. Zarathustra is a skeptic. Strength, freedom which is born of the strength and overstrength of the spirit, proves itself by skepticism. Men of conviction are not worthy of the least consideration in fundamental questions of value and disvalue. Convictions are prisons. Such men do not look far enough, they do not look beneath themselves: but to be permitted to join in the discussion of value and disvalue, one must see five hundred convictions beneath oneself—behind oneself. ¹

Life is far too important a thing ever to talk seriously about it.²

1: a psychopathology of agency

Dostoevsky’s Underground Man lacks integrity. He says so himself, voicing the accusations he expects from his imagined audience:

‘You thirst for life, yet you yourself resolve life’s questions with a logical tangle. And how importunate, how imprudent your escapades, yet at the same time how frightened you are! You talk nonsense, and are pleased with it; you say impudent things, yet you keep being afraid and asking forgiveness for them. You insist that you are not afraid of anything, and at the same time court our opinion. You insist that you are gnashing your teeth, and at the same time you exert your wit to make us laugh […]. You may indeed have happened to suffer, but you do not have the least respect for your suffering. There is truth in you, too, but no integrity; out of the pettiest vanity you take your truth and display it, disgrace it, in the marketplace…You do indeed want to say something, but you conceal your final word out of fear, because you lack the resolve to speak it out, you have only cowardly insolence. You boast about consciousness, yet all you do is vacillate, because, though your mind works, your heart is darkened by depravity, and without a pure heart there can be no full, right consciousness. And how importunate you are, how you foist yourself, how you mug! Lies, lies, lies!’³

¹ Nietzsche, A: 54
² From Oscar Wilde’s Lady Windermere’s Fan, Act I (1892)
³ Dostoevsky1993, p. 38
The failure of integrity so richly described in this passage is tied to the Underground Man’s caustic, ironic detachment from the series of shifting perspectives he adopts throughout the entire narrative. He either refuses to take them seriously, or lacks any critical detachment at all and spins out of control in fits of impassioned rapture and rage. Excessive absorption and excessive detachment are the poles between which his fragmentary consciousness oscillates willy-nilly as he gnashes his teeth at the absurdity of the role he plays in ramifying his own self-stultifying predicament. Like the collection of “notes” in which he ironically documents his condition, the Underground Man’s violently episodic psychic life is a disjointed mess, perpetually unstable but frozen in a state of irresolute paralysis.

The Underground Man represents a possibility of consciousness that is uniquely our own. His desperately embattled, disordered psyche has been seen by many commentators as symptomatic of an age in which traditional sources of normative authority have fallen deeply into question, post-capitalist social fragmentation proceeds apace, and increasingly atomized individuals are thrown back on ever-diminishing resources for coping with their alienated condition. To the extent that he represents such a psychopathological peril inherent in contemporary life, the Underground Man and his notes provide a dramatic analysandum for what follows.

So what is wrong with the Underground Man, exactly? Satisfactorily answering this question is more difficult than might be supposed. If the Underground Man lacks integrity, and this lack is supposed to be especially revealing of his peculiar deficiencies, then we face the question of what it is that he lacks. What does one lack when one lacks—and so what does one have when one has—integrity of the sort in question?
The case of the Underground Man dramatizes the problem of commitment that we will address at length in the following chapters. To better see the problem, recall that we began with the Underground Man’s own remark that he lacks integrity. We naturally ask: What kind of failing is this? Even better: What kind of condition is one in when one lacks integrity, and why is being in that condition a failing?

The first thing we should observe is that, strictly speaking, lacking integrity involves more than simply being uncommitted. For though having integrity surely requires commitment, it requires more than this, since even those who are committed can lack integrity. This means that if we want to know quite generally what the Underground Man’s lack of integrity means and why it is so bad, we must know more about the nature of commitment itself. Clarifying the positive condition will help us to understand the deficiency. After all, since being committed does not guarantee integrity, having integrity must somehow depend on the way in which one is committed. One must be properly committed, or committed to the right sorts of things, perhaps in the right way and to the right extent, or some such thing.

We can narrow things down with the help of some recent work on the nature and import of commitment. The problem we will address in the following chapters is to identify a worthwhile way of living out our deep, orienting commitments, namely the sort of passionate relationships to persons, projects or ideals which can invest our lives with richness and distinctive content that makes them our own. We will follow Robert Pippin’s recent work on Friedrich Nietzsche and call such psychologically robust,
normatively orienting commitments *erotic* ties or attachments. So we want to know what manner of erotic ties we should want to have and why, in order to understand why the Underground Man’s failure to have them explains his lack of integrity. But first we must try to understand what makes an attachment to a person, an idea, a practice, or whatever, an *erotic* one in the first place. What does it take to be committed to someone or something in this way?

Pippin claims to be following Friedrich Nietzsche and Martin Heidegger when he maintains that participants in temporally extended practices *must* have some sort of stake in them to count as committed to them at all. That this is surely true shows that Pippin’s account gets something right about the nature of commitment. Indeed, someone’s having a stake in a practice, its mattering to them in some way, is perhaps the simplest point of departure for understanding her continued involvement in it. But for that very reason, it is an uninformative one. By itself, seeing that doing something must *matter* to one to some degree in order for one’s doing it to make any sense does not refine or extend our pre-philosophical understanding of how human beings come under the spell of norms. It is just another way of saying that we take some sort of interest in what we do, that we find the practices we engage in to have some significance to us, and that we reliably exhibit this somehow. But everything of interest hangs on how the “some sort of” and “somehow” are to be filled out concretely.

Now Pippin does go farther, insisting that human practices are established and maintained “because human beings come to be committed to certain norms” by undertaking them “as if your life depended on it,” with a full or deep (or one might even

---

4 As we will see, what we want to know is not so much what objects to commit to, but how, in what spirit and with what style of activity, we should live out the commitments we have.

5 Pippin 2010, p. 11
say ‘existential’) commitment to the practice” in question. In other words, “in all distinctly human forms of life, we can…detect some basic, full-blooded or deep, ‘orienting’ commitments.” In Pippin’s view, then, enduring practices require erotically attached participants. He identifies such attachments partly in terms of a distinction between two senses of commitment:

The first we might call a thin or surface commitment of the sort involved when one agrees to play a game or participate in a social practice such as voting, and it consists in what obligations one is in fact undertaking from the point of view of any other player or participant. If you undertake to vote, you obligate yourself to vote in the proper precinct, not to vote twice, and so forth, whether you consciously acknowledge that or not...But there is another feature of your commitment that is rather a ‘depth’ commitment and, in this analogy, can be said to concern your commitment to the game itself, to its significance.6

In Pippin’s sense, “depth” commitments and “thin” commitments are two species of a class of phenomena surrounding human agency. They are different ways in which human beings may take up and live out their roles and responsibilities within practices that matter to them.

Pippin’s description requires that commitments entail obligations that those engaged in a practice understand one another to undertake as fellow participants. Let us call these default obligations in order to capture the sense in which their prescriptive content is said to correspond to a prior normative understanding shared by participants already engaged in the relevant practices. Pippin’s account of his distinction suggests that thin and depth commitments both entail default obligations. What distinguishes a depth commitment is the degree of first-personal investment it requires of the agent whose commitment it is. A ‘deeply’ committed agent in Pippin’s sense undertakes all the same obligations as one whose commitment is thin. The difference lies in the deeply

6 Pippin 2010, p. 27
committed person’s characteristically greater disposition to sacrifice what she otherwise values for the sake of honoring the obligations she thus undertakes.  

In this way, Pippin’s distinction helps mark the intensity of a given agent’s attachment to whatever is at stake within a practice. Recall that we want to know whether there is a way of taking up one’s commitments that manifests the seriousness so evidently lacking in the Underground Man’s style of activity. We know that taking a commitment seriously is possible since we recognize the failure so clearly in the case of the Underground Man. Agents function better in the relevant respect when they are not so disposed to mock and abandon their own perspectives willy-nilly, exhibiting indiscriminate irreverence, gnashing their teeth at everything, including their own feelings, judgments and concerns, just for the masochistic or exhibitionistic thrill of it. We want to say that people with integrity take their commitments seriously, and that whatever doing so entails, it does not look like that.  

We can use Pippin’s treatment of the intensity of erotic attachment to suggest an account of what taking commitments seriously entails. On this view, a person would take her commitments seriously in proportion to her disposition to sacrifice otherwise valued goods in order to promote what she is committed to. Weak or “thin” erotic attachments cannot sustain the willingness to prioritize the relevant object in relation to others that also matter to us and may compete for our motivational attention. But being weakly attached to something does not fully undermine one’s commitment, or sever one’s erotic tie, to it. It simply makes one’s commitment less serious. The seriousness of one’s commitment would thus correspond to its depth.

---

7 For our purposes, the terms ‘agent’ and ‘person’ are interchangeable.
3: erotic ties, erotic ideals

We should revisit the notion of ‘the erotic’ underlying Pippin’s account of depth commitments. Clarifying our understanding of the erotic dimension of human agency in light of his account will help us to further refine the problem of commitment to be addressed in the following chapters.

Concrete illustrations of what erotic attachments entail can be found in the intimate bonds existing between longtime friends, the devotion of an artist, musician or athlete to his training and performance in the creative or competitive activity he loves, the deeply felt patriotism, camaraderie or esprit de corps which binds a soldier to her fighting unit, military branch or nation, the attentive care a teacher brings to the lessons she prepares, so on. Among other things, an agent’s erotic ties to an enterprise are responsible for her capacity to feel the responsibilities associated with carrying it out to be motivationally salient for her. Those persons and projects to which we have erotic ties are those that matter to us, and such ties constitute our willingness to undertake personal sacrifices for their sake. Because this willingness can vary from person to person and from time to time for the same person, erotic ties in our sense embrace both thin and depth commitments in Pippin’s sense.⁸

In short, erotic ties psychologically motivate and normatively orient us by disclosing things that matter to us as mattering in some particular, experientially salient respect. And since the class of persons and things that can matter in this way to human beings is wide open, so is the range of possible erotic ties. For instance, the term ‘erotic’ is not meant to suggest that the attachments that can motivate and orient us in this way

---

⁸ By contrast, in *Nietzsche*, Pippin only characterizes depth commitments as erotic.
are uniformly or even characteristically ‘positive.’ Erotic ties are not always sources of felt contentment, satisfaction or assurance. Fear and anxiety often serve to powerfully motivate and orient us, perhaps by stemming from the felt prospect or perceived likelihood of neglecting or losing what matters to us. To take just one example, the fear of contributing to the failure of a project whose success matters a great deal to a person can serve to intensify her motivational perception of the demands to which she is subject.

Now we should distinguish erotic ties themselves from conceptions of how one is to live them out in practice. Let us call any such view of proper erotic practice, one that makes a claim on our normative allegiance, an erotic ideal. An erotic tie manifests the mattering of a person, project or ideal to a human being. It helps determine her normative and motivational relationship to what matters to her by disclosing as open, and by foreclosing, certain patterns of conduct affecting the erotic object. An erotic ideal, on the other hand, determines an agent’s relationship to her own erotic ties. Having an erotic ideal means it matters to one how things matter to one. Erotic ideals are a structurally distinct species of erotic tie, whereby one has an erotic tie to an ideal that concerns itself with one’s erotic activity. They have a second-order character, governing our relationships to our own commitments.

Whether any person needs erotic ties to function as an agent, that is, whether there is an erotic dimension to human action that is determinate but universal, is a question we must postpone for now. But we observe that, because having an erotic ideal requires

---

9 The sort of appeal exerted by an erotic ideal comes in as many varieties as erotic ties themselves: aesthetic, religious, moral, political, ethical, and so on.

10 It follows that erotic ideals may have a reflexive character, applying to themselves, regulating one’s adherence to them in accordance with their own requirements. We shall see an example of this in the Nietzschean erotic ideal of free-spirited ‘serious play’ to be set out and defended in Chapter V.

11 However, we have already seen that Pippin believes that there is—that any temporally extended social practice requires depth commitments of its participants.
having an erotic relationship of some kind to one’s way of living out one’s own commitments, erotic ideals are not necessary for erotic agency. Perhaps they are of late historical vintage, emerging alongside what Nietzsche calls “the historical consciousness” in the second of his *Untimely Meditations* and elsewhere. But we must leave this question aside as well.

For now we observe that the problem to be addressed in the following chapters is one of identifying an erotic ideal. This is because the problem of how to take our commitments seriously is one of identifying the right conception of how we should relate to and live them out. By “identifying” an erotic ideal, we mean clarifying and exhibiting its normative and psychological contours in order to show that pursuing it is both possible and desirable for human beings as we find them. We have already diagnosed the Underground Man’s lack of integrity as a failure to take his own erotic activity seriously. So one condition to be met by a worthwhile erotic ideal is that it requires the seriousness the Underground Man lacks, thereby mitigating the deficiency so clearly implicated in his lack of integrity.

4: horizons of erotic agency

We saw that Pippin’s account of commitment suggests a conception of integrity as taking one’s commitments seriously, where the relevant sense of seriousness is what appears to be necessary for maintaining depth commitments. So integrity requires depth commitment, the sort of erotic tie to something valued that compels one to undertake sacrifices for its sake, prioritizing it over competing goods. Such a view of integrity is an

---

12 Nietzsche, “Uses”
erotic ideal in our sense, and endorsements of it can be found in many guises in the philosophical literature on agency and integrity. It tends to rest on a broadly post-Kantian theory of what rational agency fundamentally requires.

Of course theories of agency are not equivalent to erotic ideals, where the latter specify constraints on how things should matter to one rather than norms governing action as such. But a view of what the active self is essentially like will have implications for the range of erotic ideals seen as possible and desirable for selves so construed. Pippin’s view of commitment entails one such strategy for making sense of being an agent capable of responding to normative demands and so of having the sort of integrity agents can have. We will call any such view of agency’s basic requirements “horizontal.”

Horizontal views regard agency as normatively intelligible only by reference to normative standards that are in each case prior to the activity in question. If the relevant activity is some discrete pattern of conduct, the norms may for instance arise from the agent’s own personal practical commitments—perhaps those embodied in the institutional contours of the community she inhabits as a participant, or more specifically those which she has first-personally endorsed upon reflection. As we shall see in Chapter IV, depending upon how it is worked out, the former captures the spirit of the neo-Hegelian and Heideggerian strategies respectively pursued by contemporary social theorists Charles Taylor, and John Haugeland and Jonathan Lear. And as we shall see in Chapter II, the latter is the neo-Kantian strategy adopted by Christine Korsgaard in her recent work on agency and integrity. In either case, according to the horizontal approach, a pattern of action can be represented as good or bad, efficient or inefficient, just or
unjust, tactful or crass and so on only by evaluating its salient features according to
normative standards seen as relatively stable and necessarily prior to the acts in question.
These are either seen as principally tied to the identity of the relevant agents, perhaps as
derived from the normative principles affirmed by them or endorsed by their community,
or as otherwise arising from or a function of the actions they perform.¹³

This explanatory strategy rests on a general doctrine about the basis of normative
authority that is unmistakably Kantian in spirit, namely that legitimate authority derives
from an ideally unanimous consensus about a normative demand. We will call this
assumption, that authority presupposes ideal unanimity,¹⁴ the consensus premise (CP) of
the horizontal view of norm-governed agency.¹⁵ Note that presuming the ideal unanimity
of a demand does not entail presuming its universality. This is why CP is Kantian only
broadly or in spirit. According to the horizontal approach, for a cultural template—say a
possible interpretation of a type of conduct or some other practically available resource
for evaluation or understanding—to be discursively available to real participants in a
social practice, it must be seen as the object of a normative consensus that is possible at
least under some ideal discursive conditions. This does not entail that it must be a
possible object of reflective endorsement by all rational beings or even each member of a
given, concrete community. So the sense of ‘ideal’ is not necessarily universalist. Rather
it makes reference to the possibility of articulating conditions for securing consensus

¹³ Jean-Paul Sartre sketches and defends a version of this view in his famous 1945 lecture, “The Humanism
of Existentialism.”
¹⁴ The idea that “ideal unanimity” principally conditions legitimate normative (at least political) authority
¹⁵ We will return to the significance of Christine Korsgaard’s reliance upon CP when we critique her theory
of agency by offering a rival account in Chapter III.
about a given norm that could be the object of the endorsement of some community, however parochial its boundaries.

Horizontal conceptions of self naturally support erotic ideals privileging behavioral consistency and uniformity. In the paradigm case, normatively intelligible agency consists in self-governance masterfully directed by a locus of first-personal responsibility. When ideally constituted agents are seen as centers of volitional control, their internal constitution is free of intractable conflict. They are unified and radically whole. Their integrity consists in the seamless integration of their psychic parts by a stable center of deliberative judgment and the effective execution of its self-given laws by an undivided will. It is a corollary of such a paradigm to treat polyvalence and ambiguity, on the side of norms, and anxiety and ambivalence, on the side of subjects, as deficient modes which obscure the normative import of the activity and compromise the effective agency of the subjects in question.

I want to argue that these implications are distortions of salient facts about erotic agency which make the ideal from which they emanate a misguided and unattractive one. But before we can see why, let us look more closely at a detailed account of erotic practice that equates integrity with seriousness in Pippin’s sense of depth commitment.

5: taking ourselves seriously

Let’s sum up. We are committed to people and things that matter to us. This ‘mattering’ motivates and orients our activity. Taking a commitment seriously means standing ready to sacrifice what we otherwise value for its sake. This is a formal characterization; what more can we say about taking things seriously? In particular, we
may wonder whether it is compatible with psychic conflict, or instead requires that we have no mixed or unresolved feelings about what we care about and what we should do in light of that concern. A sensible place to look for answers to these questions is Harry Frankfurt’s essay, “Taking Ourselves Seriously.”

In this essay and a companion piece entitled “Getting It Right,” Frankfurt’s essay extends his work on the relationship between agency and caring by offering a horizontal account of selfhood and a correlative erotic ideal of “self-integration” or “wholeheartedness.” Frankfurt’s guiding assumption is that we are creatures capable of caring about what sorts of creatures we are, which includes questioning the motives by which we wish to be moved. On his view, we are rational agents, creatures able to reflectively guide our own conduct, because our consciousness has a reflexive structure that makes “self-objectification” possible. What this means is that when one’s consciousness undergoes internal division and takes its own contents as reflective objects, one thereby becomes capable of guiding one’s conduct by second-order evaluations of first-order drives and perceptions. The critical distance involved requires a psyche whose structural complexity introduces “risk to varieties of inner fragmentation, dissonance, and disorder.”

For reasons we will come to shortly, Frankfurt thinks that maintaining this sort of fragmentation and disorder is a bad thing, a deficiency or paradigmatic failure of agency. But he believes we can avoid the deficiency by willingly identifying with or rejecting elements of the “psychic raw material” which we inherit from our forebears and circumstances. Having the ability to reflectively endorse or reject our own first-order

---

16 Frankfurt 2006, p. 18
17 Ibid., p. 6
motive is what makes us capable of having a will of our own. And though an individual agent is not responsible for the motivational dispositions which she inherits genetically or from the circumstances of her early psychic formation, she is responsible for determining whether her own dispositions have authority over her behavior. Only she can authoritatively answer the first-order practical question, What should I do?, and her answer determines what sort of a person she becomes.

Frankfurt’s core psychological doctrine is thus that human beings are capable of caring, of taking things seriously, and that this requires a distinctive order of psychic complexity. For one cares about something when one takes an interest in one’s taking an interest in it, when it matters to one that it matters to one. On such a view, the ability to form commitments is all about taking things seriously in this sense of caring: “When we do care about something, we go beyond wanting it. We want to go on wanting it, at least until the goal has been reached. Thus, we feel it as a lapse on our part if we neglect the desire, and we are disposed to take steps to refresh the desire if it should tend to fade. The caring entails, in other words, a commitment to the desire.”

This is very close in spirit to the language of erotic attachment that we encountered earlier when discussing Pippin’s work on the psychology of depth commitment. But Frankfurt draws some implications about the relationship between the sort of commitment involved in caring and the nature of the self that cares:

Willing freely means that the self is at that time harmoniously integrated. There is, within it, a synchronic coherence. Caring about something implies a diachronic coherence, which integrates the self across time. Like free will, then, caring has an important structural bearing upon the character of our lives. By our caring, we maintain various thematic continuities in our volitions. We engage

---

18 Frankfurt 2006, pp. 18-19
ourselves in guiding the course of our desires. If we cared about nothing, we would play no active role in designing the successive configurations of our will.\textsuperscript{19}

If Frankfurt is right that caring is “constitutive of our essential nature,”\textsuperscript{20} then the very possibility of agency, of playing an “active role” in our own volitional lives, depends on what caring entails, namely diachronic “thematic continuities in our volitions.” The paradigm of an agent is thus one who successfully purges her psychic life of internal conflict, unifying her psyche by observing stable practical requirements with which she has self-consciously identified herself:

…we become responsible persons…only when we disrupt ourselves from an uncritical immersion in our current primary experience, take a look at what is going on in it, and arrive at some resolution concerning what we think about it or how it makes us feel.\textsuperscript{21}

This is, by Frankfurt’s own admission, a robustly voluntaristic account of practical normativity, since he goes on to argue that we are only ever bound by those ends we consciously adopt and to whatever they instrumentally entail. Here we encounter the spirit of the Kantian enterprise: legitimate authority has its source in the self-determining activity of its subjects and derives from an ideally unanimous normative demand:

[The] willing acceptance of attitudes, thoughts, and feelings transforms their status…We have taken responsibility for them as authentic expressions of ourselves. We do not regard them as disconnected from us, or as alien intruders by which we are helplessly beset. The fact that we have adopted and sanctioned them makes them intentional and legitimate. Their force is now our force. When they move us, we are therefore not passive. We are active because we are being moved just by ourselves.\textsuperscript{22}

Frankfurt derives an ethical ideal from this sketch of the formal requirements of agency, one supposedly worthy of the allegiance of agents so construed. Frankfurt refers to the ideal as wholeheartedness or, following Spinoza, inner harmony or “acquiescence

\textsuperscript{19} Frankfurt 2006, p. 19
\textsuperscript{20} Ibid.
\textsuperscript{21} Ibid., p. 6
\textsuperscript{22} Ibid., p. 8
to oneself—that is, in acquiescence to being the person that one is, perhaps not enthusiastically but nonetheless with a willing acceptance of the motives and dispositions by which one is moved in what one does.” Where there is an absence of conflict between our first-order motives and our second-order evaluations of them, “we are thoroughly wholehearted both in what we are doing and in what we want…We are acting just as we want, and our motives are just what we want them to be.” In short, when we are who we will ourselves to be, we are who we are most willing to be.

Frankfurt claims not only that a condition of inner harmony provides “all the freedom for which finite creatures can reasonable hope,” but also that “being just the kind of person one wants to be” in this fashion is “unquestionably…a very good thing.” The reason is that “accepting ourselves reestablishes the wholeness that was undermined by our elementary constitutive maneuvers of division and distancing.” The idea is that psychic wholeness is a worthy aim because is constitutes the restoration of an intrinsically choiceworthy condition we once inhabited: “When we are acquiescent to ourselves, or willing freely, there is no conflict within the structure of our motivations and desires. We have successfully negotiated our distinctively human complexity. The unity of the self has been restored.”

So we can begin to see how a view of the self prescribes the sort of responsibility of which selves so construed are said to be ideally capable. Frankfurt’s conception of agency is essentially teleological since he derives an ideal of functional success from a prior conception of the psychic makeup agents characteristically have. Because his rational psychology emphasizes the ability to reflectively endorse the authority of first-

---

23 Frankfurt 2006, p. 17
24 Ibid., pp. 15-18
order motives by taking them as evaluative objects, it naturally evokes an ideal of selfhood in terms of the proper exercise of that ability. Moreover, the proper exercise of the ability consists in establishing stable patterns of identification with some of our motives over others. The account privileges the capacity for reflexive identification in this way because it is thought to be the route to a recaptured psychic wholeness.

Now, for our purposes, being wholehearted is more than a functional virtue of human agency requiring a kind of “unanimity” or “harmonious accord” among one’s first-order motives and one’s second-order, reflective lights. It entails an erotic ideal in our sense of a conception of how an agent should live out her own commitments. We saw that being wholehearted consists in “a willing acceptance of the motives and dispositions by which one is moved in what one does.”\(^{25}\) And we saw that reflectively identifying with one’s own motives commits one to the ends which they entail. Since a wholehearted person does this without internal conflict, she models an erotic ideal of being single-mindedly committed. She thus models an ideal of practical integrity because our earlier discussion tied integrity with taking one’s commitments seriously, and by remaining consistently motivated both to satisfy her commitment’s requirements and to remain so motivated, the wholehearted person appears to take what she cares about as seriously as possible. The assumption here, we should note, is that taking our commitments seriously is equivalent to being single-minded in our pursuit of what they require.

As we can see, framing Frankfurt’s notion of wholeheartedness as an erotic ideal requires some interpretation on our part, but it suits our purposes and we may do so for two reasons. First, commitments correspond to ends of action for which one is

\(^{25}\) Frankfurt 2006, p. 17
responsible, and in Frankfurt’s terms, taking responsibility for an end requires reflectively endorsing the first-order motive favoring it. And second, living up to a commitment means pursuing its fulfillment by consistently satisfying its requirements. And for Frankfurt, this in turn requires integrating oneself by resolving the conflict in one’s mental life, the conflict between competing motives, which necessitates practical deliberation. And so on Frankfurt’s view, through deliberation, agents aim to govern their own active lives by restoring unity to their psychic lives. We take on commitments to ends by endorsing the motives favoring action for their sake. And we take seriously our commitments to the ends we adopt by promoting their fulfillment through consistently performing the favored actions. Since a wholehearted person reliably endorses her own motives, she is a paragon of seriousness and, to the extent that we equate integrity with taking one’s commitments seriously, she is a paragon of integrity, too.

Two further points will conclude our discussion of Frankfurt’s views. First, we should clarify why his account of the self is a horizontal one. Second, we should draw out some implications of his notion of wholeheartedness that will help us to evaluate its merits, and so the merits of any account sufficiently like it, as a psychologically and ethically attractive erotic ideal. We will take these points in turn.

We have seen how Frankfurt’s account identifies personal integrity with unwavering adherence to one’s existing commitments. In the companion essay to “Taking Ourselves Seriously,” entitled “Getting It Right,” Frankfurt goes farther and identifies the capacity for practical rationality itself with a personal regard for the “unthinkable,” that is, with the “inability, or our inflexible refusal, to include among our
alternatives various actions that we are otherwise quite capable of performing.”

He calls the patterns of erotic identification that contingently constrain our wills in this way “volitional necessities.” They make certain courses of action such that one cannot bring oneself to pursue them.

The basic idea here is identical with the core doctrine of the horizontal strategy. It is that one can explain norm-governed agency—both its discrete evaluative character in particular cases and its possibility as a general class of phenomena—only in terms of the orientation supplied by a background set of standards or “horizon of intelligibility” discursively prior to the activity being explicaded. Frankfurt’s version of this claim is that one can only evaluate a given agent’s conduct by reference to patterns of identification she has already carried out, those which furnish her with the ‘volitional necessities’ distinguishing courses of action she can bring herself to carry out from those she cannot. Likewise, he thinks we can only evaluate the general capacity for normatively intelligible conduct, shared alike by all agents, by reference to the psychological capacity for identification already implicit in their shared identity as agents, one carved out by the reflexive structure of their consciousnesses.

Recall that Frankfurt equates taking oneself seriously, or integrity, with conformity to norms to which one is already committed. The stance is behaviorally prospective but normatively retrospective in that what counts as getting it right is treated

---

26 Frankfurt 2006, p. 32.
27 As we saw above, though we have defined the horizonal approach as an explanatory strategy concerned with the formal requirements of agency, it is deeply tied to a doctrine about legitimate normative authority, namely that legitimate authority derives from an ideally unanimous consensus about a normative demand. Both are post-Kantian ideas, and require us to conceive norm-governed agency as possible and possibly legitimate only against a background set of commitments seen as already taken on. The neo-Kantian version of the view specifies a determinate set of normative commitments said to be constitutive of agency because always already taken on by any agent as such. We will encounter this particularly robust version of the view in the next chapter when we examine Christine Korsgaard’s rational psychology of action and the ideal of integrity to which it leads.
by Frankfurt’s analysis as already settled for the agent by her existing concerns. We may compare Pippin’s view with Frankfurt’s to see that they amount to the same thing. On Pippin’s view, depth commitment entails a readiness to sacrifice for the sake of adherence to norms governing social practices, ‘default obligations’ assumed to be settled because they are at least implicitly understood by co-practitioners. Default obligations supply the ‘horizon,’ or background normative standards, supposedly governing social practices. We can see now that there is an analogue in Frankfurt’s erotic ideal of wholeheartedness: the volitional necessities governing a given agent’s conduct are analogous to the default obligations governing a given community’s social practices.

Finally, there are some implications of Frankfurt’s erotic ideal that we should consider. They all concern the way that an ideal of unwavering adherence to existing commitments compels us to regard the prospect of losing or changing with respect to our erotic attachments. After all, whatever else may be said about it, the human psyche is plastic and erotically dynamic. The patterns of identification whose firm maintenance Frankfurt believes is constitutive of success as an agent are subject to transformation. How an erotic ideal accommodates this class of phenomena, and the attitude towards them it prescribes, should be key factors in determining our allegiance to it.

Recall that wholeheartedness is supposed to constitute functional success for any agent to whom things can matter. Frankfurt asks us to regard an inner condition undisturbed by conflict as intrinsically successful and therefore choiceworthy, and to regard one marked by ambivalence as deficient and therefore unworthy of choice. But two questions arise. First, what exactly can agents control about their own erotic lives?
Second, if we are equally capable of ambivalence and wholeheartedness, why privilege one condition over another?

It turns out these two questions are connected in a way that reveals a serious defect in Frankfurt’s account of agency and integrity. He thinks agents are equally capable of ambivalence and of stable identification. The reflexive “fragmentation” of the psyche by our capacity to take our own drives and perceptions as evaluative objects guarantees this. So, to take our second question first, why attach the highest normative priority to steadfast adherence to commitments already made?

Frankfurt does not seem to have much of an argument for this feature of his view, apart from implicit appeals to the inherent desirability of restoring the “unity of the self” ostensibly compromised by the mature structural complexity of our psyches. But any condition preceding the psychic fragmentation occasioned by self-consciousness could not be anything like the achieved unity supposedly vouchsafed by wholeheartedness. Indeed, it is unclear that it even makes sense to refer to a psychic condition prior to psychic fragmentation as one that can be inhabited by persons, since by hypothesis it is not a possible object of any person’s experiential or reflective awareness. Recall that, for Frankfurt, the person as a self-responsible agent supposedly takes shape through her patterned evaluative responses to competing perceptions and drives within her psyche. So, having never experienced a condition of pure psychic unity, and being such that one cannot experience it, what grounds can there be for insisting on the importance of one’s striving to approximate it?

The problem is especially salient given Frankfurt’s horizontal method of analysis. The method requires norms to be dialectically prior to the practices they govern. Taking

---

28 Frankfurt 2006, pp. 15-18
our cues from Pippin, a norm that it is “prior” in the relevant sense is one which those
subject to it already have in hand, at least implicitly, as they hold one another to it. Now
the ideal condition Frankfurt identifies—the psychic unity initially compromised by self-
consciousness, which must be restored by wholeheartedness—may be seen as prior to
efforts to approximate it in at least one sense. Someone may in fact take himself to be
striving for the restoration of a long-lost, primal psychic wholeness.

However, this will not be sufficient to vindicate the coherence of the ideal, even
assuming a commitment to the horizontal approach. For the approach is designed to
explicate human action by showing what makes normative practices intelligible in
general and in particular cases. As we have seen, the general idea is a broadly Kantian
one: agents confer intelligibility and authority on their own practices as co-participants in
norm-governed activities in which each holds and is held by others to account. But how
can it be possible to rationalize someone’s behavior in terms of its subjection to an ideal
that does not make sense on its own terms?

To clarify. The unity prescribed by the ideal of wholeheartedness is not only a
condition never experienced by any agent, it is one that cannot possibly be inhabited by
them, at least if they are conceived as reflective evaluators with already fragmented
psyches. What would it mean to inhabit such a condition? By hypothesis, there is no
point of reference in our experience as mature agents from which to articulate an answer
to this question. If we cannot make sense of inhabiting the condition as a possibility for
agents as we supposedly find them, how can we hope to make sense out of their activity
by appealing to their subjection to it? If we follow Frankfurt, we wind up with an erotic
ideal that fails to meet the requirements laid down by the theory of agency motivating it.
Now Frankfurt has an obvious alternative to embracing this implication. He may abandon his insistence that the condition of psychic unity to be sought is identical with the primitive condition supposedly “restored” by wholeheartedness. But then it becomes completely mysterious why we should care about the ideal requiring it. Once we detach the ideal from a picture of what was once possessed and desirable but then lost, such that we find ourselves moved by the thought of recapturing it, an insistence on the importance of becoming whole seems *ad hoc*. What is so compelling about the kind of unperturbed self-satisfaction which Frankfurt claims is “unquestionably…a very good thing”? He has an answer, one that is interestingly different from the earlier, romantic insistence upon the inherent appeal of restoring a lost primal psychic wholeness:

…suppose that we are doing what we want to do, that our motivating first-order desire to perform the action is exactly the desire by which we want our action to be motivated, and that there is no conflict in us between this motive and any desire at any higher order. In other words, suppose we are thoroughly wholehearted both in what we are doing and in what we want. Then there is no respect in which we are being violated or defeated or coerced. Neither our desires nor the conduct to which they lead are imposed upon us without our consent or against our will. We are acting just as we want, and our motives are just what we want them to be. Then so far as I can see, we have on that occasion all the freedom for which finite creatures can reasonable hope.

There is a problem with Frankfurt’s inference from a condition of harmony between first and second-order desires to one’s immunity from “being violated or defeated or coerced.” It simply does not follow that someone who endorses her own first-order motives is therefore free from external control. One can easily imagine cases of psychological and behavioral conditioning so thorough that the subject happily identifies with the patterns of identification embodied in her own resulting mental and emotional states. To take a hackneyed example that makes the point well enough,

---

29 Frankfurt 2006, p. 15
consider a particularly severe case of Stockholm Syndrome in which an abductee identifies with his captors so deeply that he not only wishes to remain with them but intensely dislikes the thought of being talked out of his newfound fellowship. This is a familiar sort of worry about Frankfurt’s account of freedom as the capacity to form second-order volitions. But since he uses that doctrine here to ethically sanction his ideal of wholeheartedness, the usual objection applies. Since wholeheartedness is no guarantee that one is not being “violated or defeated or coerced,” Frankfurt is wrong by his own lights when he insists that it is “unquestionably…a very good thing.”31 So far, then, it looks like his erotic ideal is either incoherent or ethically problematic on its own terms.

But there is another problem with the ideal of wholeheartedness, one that arises whether or not the ideal is attached to the mythic, romantic quest for primal psychic wholeness. It raises both psychological and ethical worries about ‘seriousness’ as a sufficient basis for an account of integrity, at least when it is understood in Pippin’s sense as fulfilling default obligations tied to commitments already made, and especially in Frankfurt’s sense as doing so single-mindedly, “in thoughtful conformity to stable and appropriate norms.”32 And it brings us back to our first question: what exactly can agents control about their own erotic lives?

There is a curious asymmetry in Frankfurt’s account of the scope of an agent’s conscious influence over her own erotic activity. We have seen that, by his lights, we as agents can gain or abandon ends through our own conscious effort. We do so by endorsing or rejecting the authority of a drive or perception at work in our psyche. Moreover, we can gain objects of concern without or despite our effort, such as those we

---

31 Frankfurt 2006, p. 15
32 Ibid., p. 2
inherit from our formative childhood experiences, or those we gain through love, which outstrips our voluntary control in many ways:

…what we care about is not always up to us…We cannot have, simply for the asking, whatever will we want. There are some things we cannot help caring about. Our caring about them consists of desires and dispositions that are not under our immediate voluntary control. We are committed in ways that we cannot directly affect…Among the things that we cannot help caring about are the things that we love.33

Frankfurt’s treatment of love does qualify his general view that, fundamentally speaking, erotic agents can and should strive for psychic unity through wholehearted commitment. Nevertheless, there is a tension between this concession and the post-Kantian spirit of his general view of agency. Though he concedes that “we cannot bring ourselves to love, or to stop loving, by an act of will alone—that is, by merely choosing to do so,” he never retracts his earlier insistence that we cannot lose commitments against our will.34 We may not be able to willfully adopt or forfeit certain patterns of identification, such as those defining our loves, but if we decide to maintain one, we may.35

Putting aside the question of whether this tension in Frankfurt’s view can be maintained without undermining either of its poles,36 his overall view still seems to be that though we inevitably find ourselves caring about certain people and projects, once our reflective activity is up and running, we have an ideally, and largely actual, sovereign first-personal control over our own erotic lives. The figure of sovereignty is strongly implied by Frankfurt’s choice of a politicized vocabulary when discussing the way in which reflexive identification of our own mental contents supplies us with reasons for action:

33 Frankfurt 2006, p. 24
34 Ibid., p. 24-25
35 Note that the volitional capacity thus represented as invulnerable is essentially reactive.
36 For reasons that will emerge in the following chapters, I doubt that it can.
Some of the psychic raw material that we confront may be so objectionable to us that we cannot permit it to determine our attitudes or our behavior…we are resolved to keep it from producing any direct effect upon the design and conduct of our lives. These unacceptable intruders arouse within us, then, an anxious disposition to resist. By a kind of psychic immune response…we push them away, and we introduce barriers of repression and inhibition between them and ourselves…This means that we deny them any entitlement to supply us with motives or with reasons. They are outlawed and disenfranchised…The fact that we continue to be powerfully moved by them gives them no rational claim. Even if an externalized desire turns out to be irresistible, its dominion is merely that of a tyrant. It has, for us, no legitimate authority.\textsuperscript{37}

This vocabulary of freedom and tyranny entails something we encountered earlier, what can only be called a ‘legislative’ conception of selfhood: “The fact that we have adopted and sanctioned [attitudes, thoughts, and feelings] makes them intentional and legitimate. Their force is now our force.”\textsuperscript{38} The upshot is that it is ultimately up to us what matters to us, not because we have complete control over our psychic makeup, but because it is completely up to us what sense we make of our psychic makeup. If we find ourselves with a drive we cannot avow on reflection, we may resist its motivational pull. And if we find ourselves with one we endorse, it is equally within our power to hold the line and refuse to surrender that identification, whatever countervailing pressures may exist.\textsuperscript{39}

The worry is not simply that an account of sovereign agency like this one entails a psychologically dubious asymmetry of influence over our own erotic activity. More importantly, it psychologically legitimates an ethical ideal—wholeheartedness—that denigrates the prospect of loss and transformation in our erotic lives. In our discussion of wholeheartedness as an ethical ideal, we saw that it denigrates the experience of erotic

\textsuperscript{37} Frankfurt 2006, p. 10
\textsuperscript{38} Ibid., p. 8
\textsuperscript{39} Of course, particularly where love is concerned, we know this to be false. One can fall out of love, and lose the relevant commitments, just as easily as one can fall in love.
transformation by treating its prospect as an intrinsic failure. But Frankfurt communicates this implication in no uncertain terms.

His initial characterization of the psychic fragmentation introduced by self-consciousness carries strongly negative overtones, suggesting that it “impairs our capacity for untroubled spontaneity,” which “exposes us to psychological and spiritual disorders that are…not only painful; they can be seriously disabling. Facing ourselves, in the way that internal separation enables us to do, frequently leaves us chagrined and distressed by what we see, and well as bewildered and insecure concerning who we are.” In a word, “inhibiting uncertainty or ambivalence” can lead to “a nagging general dissatisfaction with ourselves.”

Frankfurt ruefully concedes that these forms of psychic discomfort are typically “too commonplace to be regarded as pathological,” in fact that they are “so integral to our fundamental experience of ourselves that they serve to define, at least in part, the inescapable human condition.”

But he hastens to contrast them with the byproducts of psychic differentiation which he so highly esteems: freedom, wholehearted commitment and love.

So what are we to make of wholeheartedness, the erotic ideal to which we have been led through Pippin and Frankfurt’s horizontal psychology of the self? We reviewed the reasons why it is either incoherent or ethically problematic on its own terms, but I think the most serious worry is that it denigrates erotic phenomena surrounding the experience of loss and transformation. For one thing, the capacity to form and adhere to commitments exists alongside, and depends upon, the ability to interrogate, adjust or abandon them. Frankfurt acknowledges this but insists that ambivalence is an inherent,

40 Frankfurt 2006, pp. 4-5
subjective failure of erotic agency. What he ignores is the way in which erotic loss and
transformation, and the conflicts over what matters to us that make loss and
transformation possible, are not only integral to the dynamic character of our erotic lives
but can be of immense ethical value.

Because horizontal views of agency naturally equate integrity with ‘seriousness’
understood as steadfast conformity with commitments already made, they tend to
downplay the psychologically constitutive and potentially valuable role played by
conflict and ambivalence. This is because conflict over the meaning and value of our
commitments critically spurs erotic loss and transformation. To that extent, contestation
is responsible for the dynamic lives of erotic practices and those engaged in them. My
task in the following chapters will be to show why and what this means for our problem
of identifying a worthwhile erotic ideal of integrity.

Specifically, I intend to demonstrate that the sort of erotic ideal suggested by
horizontal views of agency, one privileging psychic wholeness or unity and wholehearted
conformity to stable, prior normative demands stemming from an absence of inner
conflict, is in the end neither psychologically realistic nor ethically satisfactory. But, we
might ask, have we not been led to this view of integrity by responding properly to what
is so clearly lacking in the case of the Underground Man? If we are not to diagnose his
erotic pathology as a lack of seriousness, then how should we diagnose it? Must we start
from scratch?
6: playing with our selves

Fortunately, we do not have to start from scratch. There is something that the horizontal approach to integrity gets exactly right: the Underground Man’s lack of integrity stems from his failure to take anything seriously. But we have been supposing that Pippin’s talk of ‘depth’ adequately captures what is at stake in this failure. This is mistaken. It turns out that his account contains an ambiguity we overlooked.

Recall that Pippin’s horizontal view equates the seriousness of an agent’s commitment to something valued with the intensity, or “existential depth,” of her practical investment in promoting it. In turn, we took existential depth to correspond to one’s readiness to sacrifice other valued goods in discharging the responsibilities imposed by the values to which she is committed. And finally we took this to require conformity with the prior normative demands of a commitment already made. We vindicated this last interpretive choice by finding support for it in Frankfurt’s notion of wholeheartedness, whereby integrity requires nothing short of single-minded adherence to the demands imposed by commitments already made. Because both Pippin and Frankfurt defend horizontal conceptions of agency, it came as no surprise that they seemed to be offering kindred erotic ideals.

But something is awry here. Though existential depth is surely an important dimension of commitment, we have begun to suspect that it cannot ground an ethically and psychologically adequate erotic ideal, particularly when the ideal to which it leads is some version of wholeheartedness. I think this suspicion is right. The sense of seriousness required for integrity cannot be exhausted by the “depth” sense of commitment entailing wholeheartedness because this sense only grounds one of two
relevant senses of seriousness. Pippin contrasts depth with thinness, yet thinness in the sense we have attributed to him does not necessarily constitute a deficiency of seriousness. This is because an unserious person is one who *plays*, and while there is more than one sense in which one can be playful about one’s commitments, thinness in the sense we have attributed to Pippin only captures one of these.

If we look more carefully than before at the language Pippin uses to frame his distinction between depth and thin commitments, we will see that it equivocates between two different, relevant senses of play. Pippin tells us that someone whose commitments to a practice lack depth merely goes through the motions in carrying out the relevant responsibilities, performing them “in a bored and mechanical way just because everyone else is doing it, with little stake in the outcome (but observing the rules, your thin commitments).”41 We should notice at once that performing one’s responsibilities “in a bored and mechanical way” involves two distinct erotic failures corresponding to two different senses in which one can go through the motions in carrying out one’s role. Let’s consider them in turn.

First, Pippin tells us that a person lacking depth commitment to a practice performs her role with boredom. She observes her thin commitments or “default obligations” but has little personal stake in the role they prescribe. It is no wonder that someone just going through the motions in this sense participates in the relevant practice “just because everyone else is doing it.” We may call this erotic failure *subjective conformism*. The subjective conformist is indifferent to her role and the broader practice in which it fits. She is therefore unserious in the sense that she is “just playing around”

41 Pippin 2010, p. 27
rather than demonstrating the focused absorption we expect from someone who cares deeply about what is at stake in what she does. This form of play entails a low degree of existential investment in the relevant practice, an unwillingness to sacrifice for the sake of one’s contribution to its success (among other things).

Second, Pippin’s language identifies a “mechanical” way of performing a role. This failure constitutes going through the motions in a sense distinct from what is involved in subjective conformism. It entails an objective conformism. The objective conformist carries out the responsibilities of her role with little or no originality, perhaps by following instructions to the letter, failing to inflect them in anything like an idiosyncratic fashion. She reliably, perhaps even excellently, satisfies her default obligations but her performance lacks style.

Notice that these two kinds of erotic failure are distinct. The subjective conformist is indifferent whereas the objective conformist is unoriginal. An agent can fail in one respect without failing in the other. Indeed, one can fail in one respect and positively excel in the other.

For instance, the objective conformist lacks style but may exhibit a high degree of existential investment in her role. She may throw herself into the performance of her responsibilities, or prioritize them over other pursuits to whatever degree we like, however unoriginally mechanical her manner of execution may be. In other words, she may take her commitment to the relevant practice very seriously and so exhibit depth commitment in Pippin’s sense. Indeed, her commitment may be thoroughly wholehearted in Frankfurt’s sense, which suggests a further reason to doubt both his
insistence that wholeheartedness is an unquestionably good thing as well as our earlier supposition on his behalf that it is the key to a worthwhile account of integrity.

The crucial thing to notice is that the sense of play relevant to the failure of the objective conformist is not the same as the sense in which the subjective conformist does. For, unlike the subjective conformist’s failure, the objective conformist’s consists in an absence of serious play, not in the presence of unserious play. This is a critical distinction for our purposes, so let us try and clarify the point.

The sort of play engaged in by the subjective conformist is indeed an erotic failure. It is the unserious play involved in “mucking about,” the sense of idly, trivially or absent-mindedly going through the motions in fulfilling one’s responsibilities. In a sense, it is the failure for which the ideal of wholeheartedness is designed to supply a diagnosis and a cure. By contrast, for reasons we will come to shortly, it is at least sometimes an erotic failure to neglect the sort of play neglected by the objective conformist. What she neglects is serious play, a style of activity in which one expands upon given normative boundaries by improvising, experimenting with or otherwise challenging them. This is the sense in which one can play with one’s role in a practice in which something is at stake. And it is clear that someone who plays or experiments with the parameters of her role in a practice can thereby demonstrate that she takes what is at stake in it very seriously—so seriously that she would advance the state of the art of the practice by exploring its frontiers, exposing the deficiencies or exceeding the limitations of established ways of going on.

Importantly, Pippin departs from the horizontal framework when he touches upon the objective conformist’s characteristic erotic failing by referring to unoriginality in the
performance of a role. In his horizontal framework, seriousness and play are mutually exclusive yet they both entail the fulfillment of default obligations. What distinguishes depth commitment is one’s degree of existential investment in observing thin ones (one’s default obligations). We therefore need another erotic framework in order to properly diagnose the objective conformist’s erotic failing. So far, all we know is that it must be one in whose terms it can be an erotic failing to neglect a form of play that is equally a way of being serious about what matters to one where that is at stake in what one does.

To sum up, we have begun to identify a non-horizontal sense in which it might be possible to take commitments seriously, to live them out with an existential depth of investment. This means that we have identified the route to an alternate account of integrity: not an erotic ideal of seriousness in the fulfillment of one’s default obligations, but of serious play. But before we try to more precisely identify the erotic framework providing that route, we should note two more things about the ideal of integrity to which we want it to lead.

First, in such a framework, taking commitments seriously constitutes a form of play in part because it involves flouting default obligations. In addition to connoting experimentation or improvisation, “play” may be taken to suggest distance between the self that acts and the limiting effect of norms that define its actions in a particular way. This is the sense in which we may notice a degree of play, or looseness of fit, between

---

42 Someone who “plays seriously” flouts default obligations in that inter alia she challenges the meaning and import norms are already understood to have by those engaged in the relevant practice. I recognize that so far this is ambiguous. For one thing, if we abandon the horizontal framework, we must decide in what sense, if any, and to what extent, a normative understanding of what is at stake in a practice can really be shared by participants. And we will then have to explain what challenging norms consists in given our account of what norms are like. Doing all this is a massive undertaking, one that in some respects exceeds the scope of the present inquiry. Nevertheless, in the following chapters, particularly in our discussion of power and in our subsequent, extended defense of an erotic ideal of serious play, we will provide at least a detailed outline of what should be said about these matters. The next sentence gives some indication of the metaphorical direction in which the account will tend.
something bound and what binds it. Second, recall that all along we have been identifying our commitments with the responsibilities we assume when we take a stand on an issue we care about. By doing so, we live out an identity of a certain kind—we become the sort of person who cares about a given issue in a particular way and for certain reasons specified by the nature of the stand we take. This all seems right. But what does it mean for an account of integrity understood as serious play? It means that someone who lives out her commitments in a seriously playful manner thereby experiments with the boundaries of her role in a practice. And this means she is playing with the boundaries of her identity. She is playing with her self.

In his review of the contemporary discourse of authentic selfhood, *On Being Authentic*, Charles Guignon identifies a “postmodern” conception of agency that exalts a self that plays with its self. It is worth reviewing this conception in order to determine whether it can support an erotic ideal of serious play. Perhaps it is the non-horizontal framework we are after.

Guignon presents the “de-centered” self of postmodern thought as a reaction against the modern conception of a “nuclear self:” “To be human, according to the modern way of thinking, is to be a subject, a sphere of subjectivity containing its own experiences, opinions, feelings and desires, where this sphere of inner life is only contingently related to anything outside itself.”

Apart from being defined in terms of “inner space,” or a self-contained field of consciousness over against a world of publicly observable phenomena, the modern subject is defined in terms of its masterful control over its own conscious life. In this sense, its structure determines its proper function.

Being a self-subsistent interior space containing mental phenomena like beliefs and

---

43 Guignon 2004, p. 108
desires, its defining capacity is that of “objectifying and working over aspects of one’s situation, one’s body, even one’s own feelings and desires, on the basis of rational reflection about what one wants to attain.” The subject is thus conceived as “the source from which action springs. As an agent, the ultimate initiator of action, it is free in the sense that it can determine its own goals and decide its own course of action.” In short, “the self as conceived in modernity is a ‘bounded, masterful self,’ what Joseph Dunne calls the ‘sovereign self.’”

Guignon explains that notion of the subject, and to the extent that they are equivalent, the horizontal view of the self we have been considering, came under attack during the late twentieth century by “a cluster of movements loosely termed postmodern.” Postmodern thought “de-centered” the subject, “rethinking humans as polycentric, fluid, contextual subjectivities, selves with limited powers of autonomous choice and multiple centers with diverse perspectives.” The idea is there simply is no stable, unitary center of conscious experience and voluntary action underlying the multiple social roles played by an individual. There is no “one self playing many parts,” there is only the playing of the parts. There is nothing permanent, cohesive and ultimately authoritative to be found beneath the interplay of perspectives within and among the psyches of different individuals. Only the range of perspectives can properly

---

44 Guignon 2004, pp. 108-109. This description recalls our earlier encounter with Frankfurt’s view of the active self, which we saw wields sovereign power over its own erotic life. Admittedly, Frankfurt breaks with the modern view by denying that selves robustly determine their own goals, insisting that to a large degree we simply find ourselves with a given motivational character. Nevertheless, he believes it is robustly up to each individual whether to reflectively endorse or reject a given drive at work within her psyche, and that this voluntary, first-personal stance determines whether it has any proper authority over our conduct. To that extent, then, Frankfurt is committed to thinking that we are simply in charge of who we really are.

45 Ibid., p. 109

46 Ibid., p. 111
be said to exist, and they do not always or necessarily conform to the somatic boundaries of individual human beings.

The postmodern self is therefore not a singular locus of authority, a self-governing agent whose jurisdiction has determinate psychic and normative boundaries, as even Frankfurt suggests. Such boundaries are fluid and “always culturally and linguistically conditioned” since what kinds of selves exist is a historically contingent matter. For instance, in feudal Japanese society, “samurai” was an identity constituted and enforced by the complex but accidental constellation of rituals, institutions and practices peculiar to that time and place. In contemporary American society, where the relevant institutional and cultural support is absent, there simply cannot be individuals with samurai identities. But naturally there are others, and these too will collapse or dissipate in time and be overtaken or replaced by others as yet unknown. Accordingly, on the postmodern view, any given individual’s “own most basic commitments and defining ideals are ultimately up for grabs.”

The contingency of identity as conceived by postmoderns grounds a conception of the proper task of being a self. Citing Richard Rorty and Michel Foucault, Guignon explains that postmodern thought has tended to support an ethical ideal favoring “a stance of irony toward all self-definitions.” Given that there is no single, pregiven, underlying identity for one to realize and express through action, the postmodern version of authenticity enjoins one “to be that lack of self with playfulness and ironic amusement.” At least according to Rorty, Guignon tells us, this ideal is supposed to be liberating:

---

47 Guignon 2004, p. 119  
48 Ibid., p. 116  
49 Ibid.  
50 Ibid., p. 119
When I see myself as an ongoing event of self-creation that is answerable to no one, I can make up my own life story in any way I like, constrained only by the limits of what is on the table in my culture and my own imagination.\(^{51}\)

Guignon credits Friedrich Nietzsche with developing this ideal of ironic, playful self-creation, citing section 382 of *The Gay Science* in which he offers an image of “great health” in which one casts off traditional constraints and instead “plays naively—that is, not deliberately but from overflowing power and abundance—with all that was hitherto called holy, good, untouchable, divine.”\(^{52}\)

So the dispersed self of postmodern thought is ideally playful. But can it underwrite an erotic ideal of *serious* play? Guignon emphatically denies that it can. He insists that a view of the self as irreducibly polycentric and radically contingent cannot vindicate the discourse of commitment and integrity presupposed so far by our discussion. He seems to think that Nietzsche’s influence within postmodern thought guarantees this failure, reminding us that what Nietzsche and his followers find so exhilarating about the prospect of playful, ironic self-creation is that it presupposes “an image of the self standing outside the demands of responsibility and integrity.” Guignon counters that the ideal licenses moral and political irresponsibility with its rhetoric of liberation from all constraints as well as threatens “fragmentation and painful dissociation of the self as an agent in the world.” For “where the agent is nothing but shifting fragments, with no underlying unity, the ability to be an effective agent is undermined.”\(^{53}\)

So nothing less than the possibility of effective, responsible agency is threatened by the dispersed, postmodern self and the playful experimentation which is its privileged mode of action.

---

\(^{51}\) Guignon 2004, p. 116

\(^{52}\) Ibid., p. 119

\(^{53}\) Ibid., p. 123-124
Guignon ultimately argues that the best prospect for vindicating a worthwhile discourse of integrity in the wake of postmodern thought lies in a version of Frankfurt’s ideal of wholehearted commitment, one couched in terms of a dialogical conception of the self. Once we recognize that any human self is formed and sustained through its relationships with others with whom it necessarily shares a language and common concerns, we can see that its stable continuity, unity and capacity for self-responsibility are firmly grounded in its socially enforced need to answer to others for its avowals and conduct. Guignon therefore aligns himself with Pippin and Frankfurt by resorting to the normative priority and default authority of shared interpersonal horizons of meaning to account both for the possibility of agency and integrity. Those with integrity form and own up to the stands they take on issues of concern to the communities from whose shared resources and practices they draw their identity to begin with.

So we would seem to be back where we started. If Guignon is right to insist that talk of effective and responsible agency necessarily presupposes a horizontal conception of the self and an erotic ideal of something like Pippin’s depth commitment or Frankfurt’s wholeheartedness, then “serious play” is an oxymoron. In the chapters to follow, I will argue that Guignon dismisses Nietzschean talk of “play” too hastily. There is a perfectly intelligible and worthwhile sense of “serious play” that not only captures what has gone wrong with the Underground Man in particular but grounds a broad conception of integrity that is psychologically and ethically sounder than the horizontal ideal on which I tried to cast doubt as the best or only alternative. But as we have recognized already, an alternative erotic ideal requires the support of a view of the self capable of satisfying it. So we return to our earlier question, the one we hoped to answer
by turning to postmodern thought: where can we find a suitably non-horizontal conception of erotic agency?

The answer has already been suggested by our provisional discussion of why the horizonal strategy’s merits are questionable. We observed that since it naturally conceives integrity as requiring psychic unification through ideally wholehearted conformity with commitments already made, the view treats psychic fragmentation, conflict and ambivalence as deficient modes of subjectivity. We found this to be dubious in light of the role and value of loss and transformation in our erotic lives. Taking this as our hypothesis, we will seek resources for reframing erotic agency by exploring the domain of effects and relations which make these phenomena possible. Our name for that domain will be *power*.

This brings us to a subordinate hypothesis, that power makes contesting the meaning and authority of norms possible by constituting a dynamic manifold of differences which transgress ordered relations. The term “power” captures, first, its productive role in forming subjects’ erotic capacities; second, the way in which it is experienced by subjects as partly arising from ‘elsewhere;’ and third, the respect in which it imposes limitations on the control agents may exercise over the normative and performative shape of our own volitional lives. These features, which we will spell out more carefully in Chapter III, show how power infuses erotic activity with dynamics of precarity. Its effects and relations are persistently contested in psychic and social life and the dynamism thus generated gives psychic and social life its fluid and resilient character. Recognizing power therefore complicates the task of accounting for how we

---

54 Judith Butler’s recent work on precarity is closely aligned with the account of power and erotic agency to be defended below. See esp. her *Precarious Life* (2004) and *Giving an Account of Oneself* (2005).
are guided and governed by norms because it entails that we must exercise agency in a condition of ambivalence or ‘erotic anxiety.’ Doing so helps us to avoid the psychological mistake made by horizontal views of representing the psychic life of an agent and the normative constitution of a community as ideally static and discursively and performatively closed.

Now, since power makes conflict possible, the framework for understanding erotic agency based on it will be an agonistic one. Accordingly, we will call it Agonistic Realism to contrast it with Christine Korsgaard’s robustly horizontal framework that we will call Sovereign Idealism when we examine it in Chapter II. The erotic ideal of serious play that rests on Agonistic Realism will therefore constitute an analysis of agonistic integrity. What we hope to show is that Agonistic Realism vindicates the sense of mature erotic attachment Nietzsche explicitly described in terms of serious play:

A man’s maturity—consists in having found again the seriousness one had as a child, at play.\footnote{Nietzsche, BGE Part Four: 94}

7: taking things seriously: serious play

Here at last is the problem prompted by the case of the Underground Man, framed as a question: how should we live out our commitments in light of the possibility that we stand to lose, or lose our grip on, the erotic attachments that animate them? In this chapter, we found that horizontal approaches treat erotic loss and transformation as deficiencies to be overcome by establishing and enforcing stable boundaries of the self rather than as valuable resources for self-overcoming as something worthwhile for its own sake. They regard play as unserious by definition. Since we have found this to be

\footnote{Nietzsche, BGE Part Four: 94}
an apparent shortcoming, we have concluded that a satisfactory erotic ideal would be one that indicts the Underground Man but not the erotic plasticity of which his psyche is a pathological abomination. We want an erotic ideal that affirms the role and value of self-overcoming by resting on a conception of the self that accounts for the erotic anxiety that makes it possible. Because contestation fuels erotic anxiety, the ideal we want is an agonistic view of integrity as serious play.

Notwithstanding Guignon’s skepticism, our language of self-overcoming reveals that the ideal we are after is Nietzschean in spirit. Our full treatment of it will have to await our concluding chapter. But we should not conclude this one before clearing the way for that account by provisionally addressing Guignon’s concerns. For we will find that an ideal of serious play does not preclude the integral erotic role or value of a framework of issues that matter to the communities that agents inhabit. It merely refuses to assign social horizons of meaning the absolute analytical priority they enjoy within horizonal accounts.

What Agonistic Realism shares with the horizonal approach to integrity is the very first step in its analysis. This is the idea that if our commitments are constituted by the stands we take on the issues that matter to us, then we cannot take a serious stand, nor take our stands seriously, without taking seriously the issues themselves. Since we maintain with Frankfurt that integrity is all about taking our commitments seriously, then integrity requires taking seriously issues that matter, too. And it is an important feature of the account of erotic agency and integrity to be defended below that it sees taking important issues seriously as requiring fully owning up to their complexity.
This brings us to the import of the role played by power in forming subjects and conditioning the exercise of their erotic agency: the issues that orient our deep concerns elude our capacities for finally making sense out of them. This is because of how power shapes our capacities for framing and practically addressing the issues that matter to us. Taking the issues seriously therefore requires owning up to precisely this.

But what does it mean to say that the issues that matter to us elude our efforts to understand them? In Chapter II we will see that, according to Korsgaard’s robustly horizontal view of agency and integrity, it is through our role as participants in dialogue about matters of common concern to speech communities to which we belong that we work out our personal understandings of what we care about and why. But when we turn to Judith Butler’s account of agency in speech in Chapter III, we will see that the normative and performative status of our speech is always to some degree out of our control. This is what is meant by the claim that the meaning of anything of concern to us, including the normative and performative import of our own actions, persistently partially eludes us. Butler’s analysis will enable us to better understand what is going on when we take an issue seriously by trying to live out our convictions about it. If doing this is being committed, then taking that seriously, or having integrity, means acknowledging our limitations as agents. And, as I will argue in our final chapter, Nietzsche’s description of the playful erotic style exhibited by “the free-spirit” supplies a fruitful model of what this might look like.

But the point about taking our erotic objects seriously can be made in another way. Unless integrity is to be a form of narcissism, then in the first instance, the objects of our commitments should be the people, projects and issues that matter to us, not any
given conception of their significance which we happen to possess at a particular time.
Taking our commitments seriously in light of this requires not single-minded adherence
to a pregiven conception of what one already cares about but, first, owning up to the
complexity of the person or issue to which our convictions are subjective responses.
Caring about doing so gives us reason to doubt the adequacy of our own convictions as
committed responses to issues whose meaning and import no finite agent can exhaust in
thought, word or deed. In light of this, integrity takes on an agonistic character because it
has as much to do with challenging our own convictions as with enforcing their dominion
over our psychic life or challenging only or primarily those whose convictions differ
from our own.

So what I wish to defend in what follows is a view of agency that licenses an
erotic ideal privileging dynamics of internal and interpersonal conflict, transformation
and becoming, as opposed to the horizontal strategy favoring volitional continuity,
psychic wholeness, normative stability and an erotic fixation on first-personal control.
We will see that what taking our commitments seriously requires is living beyond our
horizons, not (or not only) living up to ones already established. What’s at issue in
questions of integrity is not, or not primarily, “taking ourselves seriously,” but taking
seriously the people, projects or values to which we are committed. Doing so means
recognizing their elusive complexity, and this must have priority if our erotic practice is
to avoid narcissism.

The same point can be made in terms of identity. Our identities are formed partly
through our erotic attachment to certain people, projects or ideals, and they are expressed
partly through our convictions about what these attachments require of us. Since our
convictions are subject to refutation and other forms of crisis and collapse, so are our identities. If integrity means taking our erotic objects seriously, and taking them seriously means welcoming potentially transformative challenges to our concrete investment in them, and, finally, who we are is largely a function of those investments, then integrity requires welcoming potentially transformative challenges to who we are. That is, it requires an erotic practice of engagement with difference oriented by a second-order commitment to, an erotic ideal of, self-overcoming.

If this is right, then Nietzsche is right to insist that what is needful is “having the courage for an attack on one’s convictions,” or, as he put it elsewhere, “to live dangerously,” to “live at war with your peers and yourselves!” But if this martial language is seen as shorthand for contesting formations of identity, including one’s own, then it is untrue that a playful self cannot be seen as taking responsibility for the stands that it takes on issues that matter to it. For this is the erotic possibility the horizontal strategy is wrong to implicitly deny: that taking a person, an issue or an ideal seriously may require not only risking but inviting and embracing the loss of oneself, of one’s ease and normative orientation, for the sake of that concern. Recall that “existential depth” of commitment in Pippin’s sense entails a readiness to sacrifice for the sake of one’s erotic object. On an agonistic reading of integrity, one must be prepared to sacrifice

---

56 Nietzsche, GS Book Four: 283, “Preparatory Human Beings” (my emphasis)
57 In the closing pages of On Being Authentic, Guignon gestures in this direction, citing Hans Georg Gadamer’s work as a point of departure for thinking about what this involves. To this extent, Guignon breaks with the horizontal strategy at the last. However, I doubt that these remarks can be reconciled with his earlier insistence that effective and responsible agency requires an essentially stable, unified and continuous, albeit “dialogical,” self. Ultimately, Guignon’s thought remains in thrall to the post-Kantian figure of transcendentally necessary shared horizons of meaning. The argument we will provide below is designed to show why this is a serious mistake.
oneself, and even the accustomed form of one’s devotion, for the sake of the object to which one is devoted.\textsuperscript{58}

By inviting confrontations with differences excluded by her commitment, an agent with agonistic integrity opens herself to the prospect of erotic transformation, of losing her grip on what she cares about by losing her understanding of what that care concretely requires of her. She does this because she recognizes that the issues that matter to her, to which her convictions are serious subjective responses, are ones she cannot master. She lives out an acknowledgment of the provisional status of those convictions, knowing that her understanding of the issues at stake may change, and that taking the issues seriously in light of this fact may therefore require that she modify or abandon them. Again, Nietzsche’s psychology is a critical resource for articulating what serious play in this sense can be like. In a characteristically provocative passage, he identifies the failure to remain open to potentially transformative encounters with difference as an erotic pathology afflicting much philosophical practice:

\textit{Being philosophically minded.} – We usually endeavor to acquire a \textit{single} deportment of feeling, a \textit{single} attitude of mind towards all the events and situations of life – that above all is what is called being philosophically minded. But for the enrichment of knowledge it may be of more value not to reduce oneself to uniformity in this way, but to listen instead to the gentle voice of each of life’s different situations; these will suggest the attitude of mind appropriate to them. Through thus ceasing to treat oneself as a \textit{single} rigid and unchanging individuum one takes an intelligent interest in the life and being of many others.\textsuperscript{59}

\textsuperscript{58} Note the resonances with the paradoxical existential movement Kierkegaard ascribes to the Knight of Faith in \textit{Fear and Trembling}. The Knight of Faith fully relinquishes the object of his love (in Abraham’s case, his son Isaac), simultaneously believing “by virtue of the absurd” that it will be fully restored at one and the same time. Moreover, unlike alternative figures of commitment with which Kierkegaard contrasts him, the Knight of Faith cannot perform these movements under his own power; he is no sovereign. Unfortunately we cannot presently explore the extent to which Kierkegaard’s account of the Knight of Faith’s style of commitment mirrors serious play as we propose to understand it here. But both are erotic ideals in our sense, and I think there is a considerable affinity between the two.

\textsuperscript{59} Nietzsche, HATH Part 9, “Man Alone With Himself:” 618
If ‘seriousness’ is understood along these lines, then I want to suggest that there is such a thing as taking one’s commitments, and by extension oneself, too seriously, and that this commonplace expression captures something of great general importance about the style of commitment appropriate for finite creatures like ourselves. Recognizing this puts us in a position to enrich our understanding of how properly to take one’s commitments, and by extension oneself, seriously—one which is not subject to the same objection because it is the sort of seriousness which Nietzsche believes is compatible with a healthy measure of playfulness. Recall the passage from *Beyond Good and Evil* with which we concluded the previous section. Provocatively, Nietzsche’s choice of words suggests that the horizontal strategy gets things almost exactly the wrong way around. From Nietzsche’s point of view, the condition to be “restored” through the pursuit of integrity is not the stable being of primal psychic oneness, as Frankfurt insists, but a “mature,” or passionately committed, version of the inventive becoming exhibited by the serious child who energetically grows into and out of her own perspectives. As a figure of continuous life amidst transformation, the child and her serious form of play is an apt erotic symbol for our purposes.

Finally, here is a brief overview of the argument to follow. In the next chapter we will turn to Christine Korsgaard’s rational psychology of action because it supplies an especially clear, robust and thorough statement of the core commitments of the horizontal approach we introduced in this chapter. As we have already indicated, we will refer to her theory as Sovereign Idealism. We will address its deficiencies primarily in Chapter III where we sketch an account of power as something her view elides, explicating it in terms of a class of phenomena of agency that flouts the three core commitments of
Sovereign Idealism. We will call the view of erotic agency resting on our account of power Agonistic Realism. In Chapter IV we will apply that account to some contemporary social-theoretic statements of the horizontal strategy, *mutatis mutandis*, supplementing the criticism of Sovereign Idealism and showing in outline the social and political relevance and potential value of an erotic ideal instead grounded in Agonistic Realism. Finally, with a view of agency conditioned by power in hand, we will complete our criticism of the horizontal strategy by drawing upon Nietzsche’s figure of the “free spirit” to articulate an erotic ideal of serious play that casts the prospects for practical integrity in a suitably agonistic light.
CHAPTER TWO

SOVEREIGNTY OF THE WILL:
CHRISTINE KORSGAARD’S RATIONAL PSYCHOLOGY OF ACTION

1: introduction: Korsgaard’s project in outline

In this chapter we will take a close look at a rich account of the distinctive metaphysical grounds, normative character and psychological contours of human agency. The theory of agency we will examine is Christine Korsgaard’s, but though we will focus on the version she most recently developed, the full view she defends brings together years of her work on issues of normativity and agency. Korsgaard’s position is a complex one, so our treatment of it must be selective. Our task in this chapter is to lay out certain salient features of her conception of agency so that we may later fairly evaluate its plausibility and attractiveness as a general view of the psychological constitution and normative powers of human agents with large normative implications for ethics and politics. Our present business is therefore preparatory for subsequent chapters where we will bring Korsgaard’s theory into conversation with an alternative political psychology and ethical ideal to be defended later.

Looking ahead, the three elements of her view most salient for our purposes are its robust assertions of (1) first-personal volitional control and (2) psychic wholeness as paradigmatic for human agents, and its insistence upon the analytical necessity of (3)
stable, general horizons of intelligibility for properly thematizing (1) and (2) as constitutive conditions of agency. We should remember that the views within philosophical psychology, ethics and social philosophy that interest us operate at two levels. A general view of the psychological and normative dimensions of human agency is articulated which provides the basis for normative reflection about what we called “erotic practice” in Chapter I, namely the way human subjects live out their deep commitments. Keeping this in mind, our examination of Korsgaard’s position will proceed in three main stages which roughly correspond to those of her broadly transcendental mode of argumentation.  

First, we will review the main tenets of her theory of agency by stating the intuitive case she makes for its guiding convictions about the psychology of human actors and the ethical orientation such convictions help establish. There we will encounter the three salient features enumerated above in the informal guise of paradigmatic assumptions ostensibly underlying our everyday discourse about what’s at stake in being an agent. Second, having thus set out the basic conceptual vocabulary which Korsgaard believes best captures our experience as agents, we will more closely examine the formal psychology of action developed in her recent work and review her phenomenological defense of the technical conceptions on which it relies. Third, we will shed light on the intended appeal of Korsgaard’s theory of action by comparing it with a model of agency addressed by Judith Butler in her work on linguistic performativity and hate speech regulation. Doing so will enable us to complete our discussion of Korsgaard’s work and

---

60 By “transcendental,” I mean a discursive strategy aimed at explicating the conditions for the possibility of some experienced phenomena as such.
connect it with the concerns of the previous chapter by illustrating the sense in which her theory requires stable analytical horizons for thematizing the preconditions of agency.

2: normative criticizability, volitional control, psychic wholeness

In her recent book *Self-constitution*, Christine Korsgaard develops sophisticated answers to questions about the nature of action, what constitutes a person’s identity, what’s involved in having personal commitments, and what are the character and prospects of personal integrity. The book’s subtitle, *Agency, Identity, and Integrity*, already suggests that its target of explanation includes a set of normative conceptions. As it turns out, the sense of agency Korsgaard wants to explicate is not that of causally generated behavior as conceived by the physical sciences, nor that of the ostensibly value-neutral phenomena of identity construed as a purely metaphysical relation, but of human conduct insofar as it is subject to appraisal according to norms for which participants in discursive practices may hold one another responsible. Korsgaard’s theory of agency is thus primarily designed to explain everyday interpersonal practices of responsibility ascription. To see why this is her central target of explanation, let’s consider her dialectical starting point.

Korsgaard opens the preface of her book with the observation that human actions can be right or wrong, whereas the otherwise intelligent behavior of non-human animals cannot.61 She then states the overarching goal of her inquiry: to identify and explain the distinctive feature of human choice and action which accounts for this difference. For now, let’s put aside the question of whether specifically moral appraisal, Korsgaard’s

---

61 Korsgaard 2009, Preface xi
“right or wrong,” is a legitimate mode of practical evaluation. She thinks so but knows she must defend this claim and doing so is one of the primary aims of her book. Her initial claim must therefore be weaker than any about the deep character or validity of specifically moral requirements.

The initial claim is instead that human behavior is, at least paradigmatically, normatively evaluable. By itself, this is not yet enough to distinguish it from the behavior of non-human animals, for “when an agent acts, even a non-human agent, the agent may succeed or fail.” But what it does show is that “the very idea of action is a normative one: it is subject to a standard of efficacy.” Accordingly, Korsgaard argues that specifically human conduct is distinguished by being subject to normative standards in a special way, one that is tied to rationality as a distinctively human capacity for self-consciously evaluating the adequacy of normative claims. So, human conduct is distinctively norm-governed in virtue of being evaluable by reference to discursively available norms.

“Discursively available” proprietary norms are those suitable for articulation and application by participants in discursive practices. An individual’s conduct may be judged effective or ineffective; excessive or deficient; more or less tasteful, timely, picturesque, provocative, fitting, felicitous, damnable or decorous across indefinitely many possibly relevant dimensions of goodness and badness. When referring to “normativity” in the following discussion of Korsgaard’s view, I will use the term as she does to denote the dimension of any human activity along which it is subject to

---

62 Korsgaard 2009, p. 109
63 This is my term, not Korsgaard’s, though I think it sheds light on how she motivates her kind of explanatory and normative project.
proprietary evaluation by agents themselves in these and other ways. Each of these and other possible dimensions of proprietary evaluation are encompassed by Korsgaard’s use of the term “efficacy,” since we just saw her use it to embrace all the varied forms of success or failure to which human action is susceptible.

Human action is thus norm-governed in a way that non-human action is not: we play an active role in this normative governance by evaluating it for ourselves. Starting from this simple idea, Korsgaard develops a rich view of the deep character of the volitional activity possible for human beings. Of the further claims she makes about the essential character of human agency, perhaps the most salient for our purposes are two having to do with the first-personal dimension of choice and action.

First, Korsgaard insists that each human agent stands in a special relationship to herself first-personally, one for which a radical sort of psychological and normative control is paradigmatic. The relevant sense of “psychological control” is that of its ultimately being up to each agent what she does; she determines which patterns of conduct she shall perform. The relevant sense of “normative control” entails each agent’s ultimate discretion regarding a crucial aspect of the normative status of her relation to her

---

64 The implication that there is a dimension of human activity—and so, to that extent, a way in which human activity itself—is in the end unitary and therefore finally subject to a single all-embracing analysis is one Korsgaard’s theory entails. As I explain below, this makes her view of human agency “horizontal” in the sense described in the previous chapter.

65 Throughout our discussion of Korsgaard’s recent work on agency and integrity, I will attempt to characterize the aspects of her view which are of greatest interest here and in sufficient detail to permit a charitable interrogation in subsequent chapters when we turn a critical eye to the psychology of action she is offering. We should note here at the outset that the features of her position which are of greatest interest to us—its requirement of first-personal volitional control, psychical self-unification and stable (in her case, formal/proceduralist) horizons of agency—are to be understood as paradigmatic for concrete human agents as we find them in historically situated social practices; that is, as robustly predicated only of the ideally constituted, fully realized agent. In other words, Self-Constiution offers an ‘ideal theory’ of the nature and possibility of rational agency. Whether or not this constitutes an objection to Korsgaard’s view depends in part on how we evaluate ‘horizontal’ views as such, at least insofar as horizontal representations of agency require questionable forms of idealization to get underway. We will turn to this question in the following chapter.
own conduct—namely the respect in which she is free to determine upon reflection the *reasons for which* she acts as she does. The overall claim, then, is that human beings are distinguished as agents by their capacity to determine how and why they behave the way they do. Why is control so important to an account of distinctively human agency? To return to the idea with which Korsgaard begins, namely responsibility, we might say on her behalf that one cannot be held responsible for, or called upon to justify, some behavior unless that behavior was at one’s discretion. We’ll return to this idea shortly.

There is another critical element of Korsgaard’s account we must identify before proceeding. When we say that I, as an agent, determine the action I perform, this is by Korsgaard’s lights another way of saying that I determine *myself* to act in the relevant way. And to make sense of this possibility, the thought goes, I must regard myself as causing whatever is brought about through my action. And to see myself as *the* cause of what is produced through my conduct, I must take myself to be something whole and unified over against my always multiple and potentially conflicting motives. For though my desires and inclinations may conflict with one another or tempt me to contrary action, being an agent is ideally all about guiding oneself volitionally through reflective choice—not standing still amidst conflict, paralyzed into inaction by a divided mind. The second claim we want to focus on, then, is that *volitional wholeness*, specifically as constituted by some robust form of *psychic or unity*, is paradigmatic for distinctively human agency.

So, in a nutshell, agency on such a view is norm-governed behavior guided and determined by a unified self. Being capable of carrying out self-regulated patterns of conduct under one’s reflective control is to be capable of performing an *action* as opposed to being at the mercy of one’s inner conflicts or helplessly driven to fits of
random behavior. For Korsgaard, an agent’s mantra is: “Get a hold of yourself! Pull yourself together, make a decision and act!”

This is only a first pass at two dominant themes within Korsgaard’s theory of agency: first-personal volitional control and psychic wholeness or unification. These will be the themes of greatest interest to us here and in later chapters, so our discussion of the theory will be organized around unpacking their intended meaning and rationale. However, we will endeavor to understand Korsgaard’s guiding convictions regarding the psychology of action and the proper method of inquiry about it. For though her book is richly argumentative, it is equally a sustained effort to bring readers around to a certain way of engaging with the questions about agency to which her arguments provide answers. And ultimately we will be interested in evaluating not only the content of her view of agency but also this comprehensive way of addressing the pertinent questions. So let’s look more carefully at the conception of agency by which her approach is informed and in which her arguments terminate.

3: responsibility & authority

In her preface, Korsgaard identifies the unique feature of human action, that which makes it distinctively normative, as “rationality:’’

As I understand it, reason is a power we have in virtue of a certain type of self-consciousness—consciousness of the grounds of our own beliefs and actions. This form of self-consciousness gives us a capacity to control and direct our beliefs and actions that the other animals lack, and makes us active in a way that they are not. But it also gives us a problem…the problem of deciding what to count as a reason for belief or action. To put the point another way, this form of self-consciousness makes it necessary to take control of our beliefs and actions, but we must then work out how to do that: we must find normative principles, laws, to govern what we believe and do. The distinctive feature of human beings, reason, is therefore the capacity for normative self-government.66

66 Korsgaard 2009, Preface xi.
So rational agency is paradigmatic for human beings: human beings, *qua* agents, are rational ones. And it is our basic task as rational agents to form beliefs and perform actions from some conception of why it worthwhile to do so in each case. The idea that what we are basically called to do as agents is govern our own volitional lives through critical reflection is of a piece with Korsgaard’s initial representation of the task of explaining human action as a problem about what makes norm-governed human conduct possible.

Now once we are taken in by the idea that agency as such is all about self-government, our thinking readily moves from attention to psychological capacities, such as first-personal volitional control, to a normative concern with responsibility. For those living within modern democratic societies are expected to understand that self-government is no idle pastime. Self-rule is everywhere represented as especially serious business, an arduous task imposing a heavy burden of responsibility. Moreover, much of modern political thought in the West, especially within the liberal tradition, represents proper democratic governance as fundamentally subject to formal constraints with relevantly universal scope, such as legal protections for individual citizens, and as structured by a concern with the moral rightness of deliberative and legislative procedures. Communitarian thought, on the other hand, tends to emphasize the

---

67 Since we are concerned only with her views regarding the psychology of *human* action, in the following discussion of Korsgaard’s position we will use the terms “agent” and “rational agent” and their cognates interchangeably. We should note that she begins with the assumption that there is something distinctive about human action, namely its normative criticizability by reference to standards of “right and wrong,” and derives from this a highly contentious theory of rational agency as critically reflective, normative self-government. To put the point in terms of Rawls’s distinction between ‘concept’ and ‘conception,’ “rational agency” so construed is Korsgaard’s worked-out conception of the broad concept of human agency with which she starts. We may note in passing that she does elaborate a sense of ‘agency’ of which she thinks some non-human animals are capable, but it is a limited sense that amounts to a riff on the paradigm of rational agency paradigmatic for (presumably, developmentally normal) human beings.
importance to a free society of social harmony and normative consensus, at least regarding basic rights and institutions. As our discussion of Korsgaard’s theory proceeds, we shall see that liberalism’s general preoccupation with formal deliberative constraints on lawmaking and the communitarian concern with political unity and social harmony are repurposed within her moral-psychological imaginary, functioning as core elements of her general view of agency. This will help illustrate the strongly politicized dimension of her metaphysics and psychology of action.

But, to return to Korsgaard’s prefatory remarks about the distinctive character of rationality, once her talk of self-government prioritizes issues of first-personal responsibility, our thinking about agency is channeled in certain directions by the politicized vocabulary of self-rule. Certain further questions seem unavoidable once we believe ourselves to be robustly empowered to settle for ourselves first-personally “what is worth doing for the sake of what.”68 For instance: There must be some central locus of responsibility for carrying out the relevant task, mustn’t there? And (so the thought continues) why shouldn’t this question arise in the course of our reflections about the deep character of fully realized human agency? Individual adult human beings are designated as such in our society precisely by an enumeration of responsibilities for self-regulated conduct, are they not? Indeed, individuals may consult with one another in their efforts to deliberate, and they may jointly deliberate in order to “unite their wills” and act as one, but the deliberative labor associated with any collaborative effort is ultimately indexed to the individual agents whose joint undertaking it is, and collective deliberation is possible only through the union of otherwise singular, self-enclosed centers of agency. We can join our efforts and act together, in other words, but in the end

68 Korsgaard 2009, p. 173
coordinated action is an achievement principally attributable to each contributing party, and such disaggregated attributions of responsibility to autonomous individuals should enjoy ultimate priority in our thinking about collective undertakings. Or, again, so we may naturally suppose once pointed in this direction.

The general idea here is that to the extent that we are rational agents, we determine for ourselves what we do and the reasons for which we do it, for which choices we each consequently bear ultimate first-personal responsibility as an independent person. So to the key notion of control already introduced we must now add that of the responsibility borne by whoever is charged with the task of self-regulated conduct to carry out that task in appropriate ways. All and only agents have the capacity for carrying out the task of agency as norm-governed self-control; each agent has the responsibility to do so.

These notions of responsibility and control are so tightly interwoven in Korsgaard’s conceptual vocabulary that they stand or fall as one depending upon whether her basic, guiding convictions about them are vouchsafed. Indeed, Korsgaard’s view of the significance of these concepts for understanding the nature and possibility of rational agency in human affairs is a basic point of departure for her inquiry. Her representation of first-personal responsibility and volitional control is the cornerstone of her theory of the psychological and normative character of human action. But since disentangling these two notions will help shed even more light on the psychological doctrines they support, let’s first consider the salient conception of responsibility.

---

69 As we shall see, Korsgaard intends the phenomenology of everyday choice and action to supply the ultimate vindication of these conceptions and the basic picture of agency they help compose. In this sense, the theory of agency she offers, and all it specifically entails about identity, responsibility, integrity, and so on, rests on the phenomenological credibility of the conceptual vocabulary in whose terms she articulates it.
What is it to be first-personally responsible, on Korsgaard’s view? To be clear, what’s at stake in this question are general standards for personal accountability. The question here is what it takes for an agent to be in a position to answer for what she has done, not what it takes for her behavior to be ‘responsible’ in the sense of good or appropriate in whatever respect happens to be relevant. Answering the latter question presumably requires, among other things, close attention to salient features of the decision-making context in which the agent in question is faced with a choice. Whereas Korsgaard offers a conception of agency whose basic normative and psychological character is quite universal in scope. Recall her stated explanandum; she wants to identify the feature of human action which distinguishes it in toto from the intelligent behavior of non-human animals and other natural phenomena.

So what must be true of an agent for her conduct to be such that she may take responsibility for its status as right or wrong, truthful or erroneous, effective or ineffective, and so on? We have already begun to see that Korsgaard believes personal accountability for some pattern of conduct requires that it be ultimately mine. That conduct is mine whose normative character—its success or failure, moral rightness or wrongness, and so on—I may answer for. And this in turn requires that I am in a position to freely determine the character of the conduct in question. To the extent that I have greater or lesser control over a particular movement of my body, for instance, and so may be in a better or worse position to answer for it, the more or less it is mine in the required sense.70

---

70 On Korsgaard’s account, the relevant first-personal subject can be a plurality, a collective, self-determining “we.” Indeed, it is part of the intended appeal of her view that its normative upshot for politics favors collective will-formation through free, interpersonal reason-exchange—in short, some version of deliberative democracy. Only a group whose members reason and act together can be properly regarded as
This is already suggested by Korsgaard’s initial observation that a person’s conduct may be \textit{judged} relevantly effective or ineffective, excessive or deficient, and so on. Her thought is that if one’s conduct may be evaluated as better or worse in some way, then there must be some sense in which she is especially responsible for doing better rather than worse in the relevant respects. If the individual does not have some special responsibility for the character of her own conduct, then why introduce the idea that the conduct \textit{belongs} to her, that it is in some special sense \textit{her conduct}, at all? And if she does have some such special responsibility, then that must be because she has some special \textit{authority}, some especially effective influence, over what her conduct ends up being like. So the “pre-theoretical” commitment to the notions of proper authority and responsible action Korsgaard takes her readers to share requires us, she thinks, to regard ourselves first-personally as having an authoritative say in how we behave and why, sovereign jurisdiction over our movements and their normative significance.\footnote{We will return to the import of the figure of sovereign authority for Korsgaard’s theory.}

In light of all this, it seems clear that Korsgaard’s insistence upon the paradigmatic “mineness” of those actions for which one is responsible addresses itself to a question about the nature and ultimate source of proper normative authority. As such it has far-reaching implications for matters of interest throughout practical philosophy from ethics to politics. The basic idea is that there is some normatively significant conduct—some better or worse behavior—for which each agent is in the end the most appropriate free and therefore as a legitimate source of normative authority—a responsible agent. But in the political case, as analogously in the individual one, the relevant sense of freedom is that of the “self-determination” of a democratic public functioning as a single agent through the harmonious integration of its deliberative and executive powers. And as we’ll see shortly when we turn to her conception of willing as the signal power of agency, self-determination in Korsgaard’s sense entails a radical measure of control over the normative and performative character of one’s conduct. In a word, what we do must be in some ultimate sense up to us or else we cannot be held responsible for it because we cannot be said to have done it ourselves. So in Korsgaard’s moral and political philosophy alike, agents’ authority over and responsibility for their conduct isn’t possible without the first-personal exercise of tight volitional control.
ultimate object of attribution. An agent is paradigmatically responsible for her own conduct in that she is properly subject to criticism for its salient propriety or impropriety.\textsuperscript{72} And her responsibility depends upon her ability to first-personally regulate and control, and therefore be properly subject to criticism for, such conduct. Such conduct is \textit{ipso facto her own}. The idea is then that she is especially in control of, and hence responsible for, the conduct properly regarded as her own. Here is Korsgaard:

\begin{quote}
To be an agent, according to Kant, is to be a self-conscious causality: to think to yourself that you will bring about a certain end, and somehow, through that thought, to bring it about. When you think of yourself as an agent, you think that your effects are your own, and had you not taken thought to realize them, they would not, other things being equal, have happened at all. So you are their cause, and these effects are yours, and the world is different, because you’ve made it so. There’s a touch of the divine in being an agent, and as Aristotle says, we love the effects of our actions, as poets do their poems, and parents their children, because we see them as our handiwork, and we see ourselves realized and completed in them.\textsuperscript{73}
\end{quote}

In fact, having ultimate control over and responsibility for some conduct is precisely what makes it one’s own in the first place. On Korsgaard’s view, this is what makes it the case that one is one’s own person at all, the bearer of a unique identity rather than a haphazard assortment of spatiotemporal parts bound together by contingent causal relations and without an enduring, normatively consistent pattern or shape. Having responsibilities that are one’s own—conduct whose \textit{appropriate} character one may regard as one’s own business—is just what it is to have such a shape by Korsgaard’s lights.\textsuperscript{74} And so being a uniquely identifiable self and having the ability and responsibility to act for oneself are regarded as co-constitutive conditions for the possibility of one’s rational agency. In all these ways, Korsgaard’s theory of action rests

\textsuperscript{72} That is, its goodness or badness in whatever respects are contextually relevant from case to case.  
\textsuperscript{73} Korsgaard 2009, pp. 84-85  
\textsuperscript{74} In this way Korsgaard’s analytical method prioritizes findings arrived at from the standpoint of reflection about the conditions of possibility for our experience as agents over those disclosed by other sorts of metaphysical inquiry. Her moral psychology is to this extent methodologically pragmatic, grounded in a phenomenology of choice and action.
heavily on the establishment of a tight conceptual relationship between an agent’s first-personal control over, responsibility for and authority regarding her own volitional life. The first-personal standpoint of practical deliberation, of self-conscious reflection upon “the grounds of our own beliefs and actions,” is thus assigned high priority in her account.  

We have already seen that Korsgaard’s theory of action is couched in a normative and psychological vocabulary in which the concepts of responsibility, unity and control are densely interwoven. Let’s look more closely at her technical representations of these ideas.

4: Korsgaard’s psychology of action

The will

Korsgaard’s full view is complex but grows out of a commitment to the central Kantian doctrine regarding the nature and possibility of rational volition: the will. On the Kantian view, having a will is constitutive of rational agency: all and only rational agents have wills. Simply put, the will so construed is an agent’s ability to put herself in motion through her ability to deliberate about what it is best to do. It is her ability to act from her conception of how she should act under the relevant circumstances.

So, on this view, the exercise of the will is an expression of the practical rationality of which human agents are distinctively capable. It is a power of choice of normative principles for thought and action that constitute one’s evaluative outlook and regulate one’s conduct. One establishes and expresses one’s identity—one makes oneself

75 Korsgaard 2009, Preface xi. We’ll see below that Korsgaard’s conception of the standpoint of practical deliberation constitutive of agency is reflective and procedural in a robust sense.
a particular kind of person with a certain qualitative character—through one’s choice of the practical principles governing one’s conduct. In this sense, because our actions express the values to which we are committed, we as agents choose who we are—we constitute our own identities—through the exercise of choice:

…when you deliberate, when you determine your own causality, it is as if there is something over and above all of your incentives, something which is you, and which chooses which incentive to act on. This means that when you determine yourself to be the cause of the movements which constitute your action, you must identify yourself with the principle of choice on which you act.  

The Kantian doctrine of the will invoked here rests on ascribing radical first-personal volitional control to every agent. For the will is the capacity for self-conscious reflection upon the grounds of belief and action, and the exercise of choice as an elective volitional power sufficient to motivate appropriately corresponding behavior. As such, the doctrine is also designed as more than a description of one dimension of our experience as deliberators and actors. It is supposed to capture what essentially distinguishes agents from non-agents by representing the relevant deliberative and volitional capacities as paradigmatic for every instance of the former. On the Kantian view, it is the ability to exercise this capacity that distinguishes an agent from a “mere” causally governed behavioral system. Here is the key passage from Kant’s *Groundwork* in which he sets out this opposition between rational beings and purely causal phenomena, representing it as basic for any serious attempt to understand agency:

> Everything in nature works in accordance with laws. Only a rational being has the capacity to act in accordance with the representation of laws, that is, in accordance with principles, or has a will. Since reason is required for the derivation of actions from laws, the will is nothing other than practical reason.  

---

66 Korsgaard 2009, p. 75  
77 Kant 1997, 4:412
Because Kosgaard’s explicates agency in terms of this Kantian doctrine,\textsuperscript{78} her psychology of action is shot through with the voluntaristic conviction that we previously encountered in an informal guise. Namely that rationality endows human agents with the ability and responsibility to first-personally determine the way in which and the reasons why we act as we do. Or, as Kant might say, all natural phenomena are governed by norms (or “laws”), for instance those of causality, but only an agent, as a being endowed with the power to derive actions from rational principles through the exercise of her deliberative faculties, can govern \textit{herself} through the mediation of the capacity for reflective choice. And so all and only agents are capable of the signal task of self-government: giving laws to oneself. So all and only agents are free.\textsuperscript{79}

\textit{Principles and incentives}

So, on the Kantian view defended by Korsgaard, agents are distinguished by being free in virtue of their capacity to determine through rational reflection the character and grounds of their behavior. This means they are capable of determining \textit{themselves} to carry out actions they regard as licensed or required by practical considerations whose validity they affirm on reflection. When an agent wills a particular action, she therefore puts herself in motion in the particular way and for the particular reasons she regards as authoritative for her under the circumstances:

For instance, maybe I am climbing a mountain, and if I don’t go back down now I won’t make it home before sunset. So I say to myself: ‘I guess I’d better go back down now, in order to make it home before sunset.’ That thought—I will go back down now, in order to make it home before sunset—is my maxim and it is also my conception of a law. And it is because my maxim determines what I do that my movements count as willed—that is

\begin{itemize}
  \item \textsuperscript{78} Korsgaard 2009, p. 90
  \item \textsuperscript{79} “Autonomy” and “freedom” are technical terms within Kant’s philosophy, and neither he nor Korsgaard treats them as equivalent, but the significance of this detail is beyond the scope of our present interest in the broadly Kantian psychological doctrines underlying Korsgaard’s position.
\end{itemize}
Here Korsgaard suggests that agency—‘willing’ as rational self-determination—paradigmatically entails acting from one’s conception of a law one takes to be valid for oneself. Following Kant, Korsgaard refers to such subjectively adopted principles of volition as ‘maxims.’ Reflectively guided, self-governed conduct thus requires acting from one’s ‘conception’ of a law, or action-guiding endorsement of the validity of a practical principle. So, as we have seen, agents make their conduct their own—they become accountable, self-governing normative subjects as distinct from random, causally governed behavioral systems—through willing. And the above example shows that willing in the required sense entails adopting maxims and endeavoring to comply with their specific volitional requirements.

The last point about behavioral conformity is critical. For one’s choice of practical principles is not motivationally inert. Rather, on Korsgaard’s reading of the Kantian doctrine of the will, one’s adoption of a practical principle is paradigmatically expressed through conduct determined appropriate by it. So being characteristically moved to act by certain sorts of considerations—for instance, by the fact that one is in a position to render assistance to another in need—just is for one’s conduct to be governed by a practical principle; in this case, by some version of the “principle of charity.” Perceiving and responding to the legitimate practical authority of the principle of charity is in large part to be so moved.

What we have said so far about practical principles addresses one important dimension of the functioning of ‘the will’ through which rational activity is constituted.

---

80 Korsgaard 2009, pp. 68-69
Principles play an essential but partial role in Korsgaard’s psychic economy. The functional role played by *incentives* is equally important. After all, something has got to initially prompt an agent to act, and ‘incentive’ is Korsgaard’s name for any such “potentially motivating factor” for action, including occurrent drives, impulses, and desires.

But the role of incentives isn’t limited to prompting action. Our efforts to abide by the requirements imposed by our maxims—such as the principle of charity—may well take place over against the contrary pull of counter-inclinations, misaligned urges, desires for conflicting ends and other psychological or physiological phenomena which threaten to compromise our motivational allegiance to the ends we set. Such sub-personal obstacles to self-determination are, by Korsgaard’s lights, categorically distinct from the obstacles to action furnished by events in the world understood as a causal order, where all sorts of phenomena beyond an agent’s control may frustrate her attempts to bring about the end of action she wills. When an agent is faced with incentives to action precluded by her maxim, she encounters an *internal* sort of motivational resistance—the countervailing pressures arise from her own psychic depths but she does not identify with them. Rather, they appear to her under the aspect of alien forces to be overcome through her steadfast commitment to the end of action she has determined *herself* to realize through an exercise of her will:

...suppose you experience a conflict of desire: you have a desire to do both A and B, and they are incompatible. You have some principle that favors A over B, so you exercise this principle, and you choose to do A. In this kind of case, you do not regard yourself as a mere passive spectator to the battle between A and B. You regard the choice as yours, as the product of your own activity, because you regard the principle of choice as expressive, or representative, of yourself—of your own causality.  

---

81 Korsgaard 2009, p. 75
According to Korsgaard, you must identify yourself with your practical principles, for she thinks the only alternative is to regard some such principle (or any other relevant consideration) “as some third thing in you, another force on a par with the incentive[s]” on either side of the question. If you consider yourself a “mere passive spectator” in this fashion, she concludes, “then you cannot regard yourself as the cause of the movements which constitute your action. Self-determination, then, requires identification with the principle of choice on which you act.”

Wholeness as unity

Though we are not especially concerned here with the “metaphysics of normativity” which occupies an early chapter of Korsgaard’s book, briefly reviewing her views on this score will help shed light on one of the main tenets of her psychology of action which does concern us—namely that rational agents are subject to a requirement of volitional wholeness in order to be able to function properly as agents. But first let’s consider the programmatic support the theme of wholeness derives from the formal doctrine of the will as a self-determining causality constitutive of rational agency. Doing so will enable us to understand Korsgaard’s reasons for interpreting the requirement of wholeness in light of a metaphysical conception of unification.

As we have seen, Korsgaard argues that the exercise of rational agency consists in self-determination. This is the capacity to critically evaluate the candidate reasons for action supplied by one’s impulses and acting consistently—embracing or resisting the motivational pull of whatever incentive—in light of one’s principles. When someone

---

82 Korsgaard 2009, p. 75
acts from principle she is thus expressing her rational identity which she shares with all other agents. When she acts from specific maxims, tailored to suit her idiosyncratic circumstances, she constitutes and expresses the personal practical identities associated with them. If an agent consistently acts from the principle of charity, for instance, which requires (let us suppose) that one express one’s concern for others’ welfare by rendering assistance to those in need whenever possible, then she is *ipso facto* a charitable person, and not unless.

All of this, Korsgaard thinks, requires agents to see themselves as capable of standing aloof from each particular incentive to action—indeed, as independent of the *sum total* of the potentially motivating elements within their psychic makeup. The notion of seeing *oneself* in the required way in turn presupposes the singularity, or integrated wholeness, of the agent *as a self*. For each agent’s *self* is the person whose identity is fashioned from her deliberative and volitional activity: as we said, I am a charitable person if I consistently abide by the requirements of the principle of charity, and not unless. Fashioning a “practical identity”—becoming who one is—is therefore something that happens in and through each agent’s efforts to carve normatively consistent patterns of behavior out of the multitude of motives afforded by their psyche. The self that acts is thus a self-integrating whole constituted from its own psychic parts through the deliberative activity thereby attributable to it as a whole. The crucial idea here is that the radical elective discretion constituted by the Kantian will requires a strict and unequivocal distancing of the self, seen as necessarily unitary and ideally self-consistent, from its motives, seen as multiple and conflicting:

---

83 The latter is of course a much stronger claim than the former. Nevertheless, Korsgaard’s Kantian doctrine of the will entails it.
84 A nod to Nietzsche whose ironies subsequent chapters will help illuminate.
However it goes, reasoned or arbitrary, chosen or merely the product of circumstance, the sorts of identities I am talking about remain contingent in this sense: whether you treat them as a source of reasons and obligations is up to you. If you continue to endorse the reasons the identity presents to you, and observe the obligations it imposes on you, then it’s you…you can walk out even on a factually grounded identity like being a certain person’s child or a certain nation’s citizen, dismissing the reasons and obligations that it gives rise to, because you just don’t identify yourself with that role. Then it’s not a form of practical identity anymore: not a description under which you value yourself.  

Denying the necessary unity of the active self would be to deny that practical deliberation requires a singular independent subject empowered to assent to or refuse the motivational pull of its incentives, that is, “something over and above all of your incentives, something which is you, and which chooses which incentive to act on.” For Korsgaard, such a denial goes hand in hand with a doctrine of the will she distinguishes from her own by labeling it “particularistic.” Willing particularistically “would be a matter of willing a maxim for exactly this occasion without taking it to have any other implications of any kind for any other occasion.” Thinking of willing in this way entails denying the necessity of a unitary subject for deliberation because, on this view, one’s active self, the one doing the deciding, must be radically identified with whatever incentive one acts on: 

…this means that particularistic willing eradicates the distinction between a person and the incentives on which he acts. And then there is nothing left here that is the person, the agent, that is his self-determined will as distinct from the play of incentives within him. If you have a particularistic will, you are not one person, but a series, a mere heap, of unrelated impulses…Particularistic willing lacks a subject, a person who is the cause of his actions. So particularistic willing isn’t willing at all.

---

85 Korsgaard 2009, p. 23
86 Ibid., p. 72
87 Ibid., p. 75
88 Ibid., p. 76. The details of Korsgaard’s full “argument against particularistic willing” are beyond our scope here, but its intended upshot is that, as practical deliberators, we must take our practical principles to have at least “provisionally universal” scope. The bit we have rehearsed is meant to show that one cannot otherwise preserve the sense in which there is really a self empowered to decide what to do across like cases.
So as an enduring, responsible person, an agent is paradigmatically whole. This is to say that her volitional activity must be seen as subject to a requirement of self-unification. Being whole over time, across deliberative episodes, makes her an enduring subject; being a self-unifier is what makes her first-personally responsible. Generally speaking, the parts to be integrated through the exercise of one’s will are one’s incentives to action; the active self is the whole constituted through action consistent with one’s principles. In short, “Since actions must be assignable to the person as a whole, the work of practical deliberation, the work that leads to action, is also a kind of reunification.”

Returning to Korsgaard’s intuitive starting points, we may sum things up as follows. Since being a rational agent is all about normative self-government, and since this requires constructing a single life out of many possibilities of action, being an agent means becoming a self-constituting whole: “Pull yourself together and act!”

**Teleology & normative authority**

We have just revisited the theme of first-personal responsibility earlier addressed informally, reviewing it in the light of Korsgaard’s formal conception of the will as a self-determining causality. We saw before that a rational agent is supposed to be capable of determining himself to act in one way rather than another as a whole. We may now add that the active self is a self-constitutor; it paradigmatically aims at self-constitution. But we might still wonder whether the technical conception of the will as a power to engage in a self-constituting activity, a controversial, formal thesis within the psychology of action, is not underdetermined by the everyday idea that, in order to see an action as

---

89 Korsgaard 2009, p. 133
mine, I must be capable of seeing myself as having undertaken it *as a whole*. Korsgaard has an answer ready. It is that the metaphysics of normativity suggests an ontology of action according to which action *is*, by its very nature, self-constitution.

Now if we hold that agency itself (the activity of acting) is structured by an essentially guiding aim (self-constitution), our metaphysics of action will naturally be a teleological one. Indeed, Korsgaard defends such a view. She attributes the metaphysical conception to both Plato and Aristotle, whom she says thought the authority of norms governing activities is best accounted for by reference to the intrinsic point and purpose of whatever activity happens to be in question. On this view, an activity’s point and purpose is captured by the standards defining the activity itself. They are its *constitutive standards*, or those that distinguish doing *that* activity from doing another. They both constitute and govern the activity in question, for by specifying the conditions which must be satisfied for an activity to be of the relevant kind, they constitute the non-negotiable ground rules applicable to anyone engaging in it. That is to say that, to the extent that one fails to strive to satisfy them, one fails to so much as engage in the relevant pursuit. As long as one remains so engaged, then, one’s efforts must be seen as subject to evaluation in light of at least those standards.

It is this idea which guides Korsgaard’s analysis of what an action *is* such that particular actions can be subject to normative requirements at all. For we have seen that she maintains with Kant a sharp distinction between entities capable of action and those which are not: the latter behave in accordance with laws, the former in accordance with their conception of laws. Since agents occupy a distinct category defined by the sort of activity peculiar to it, Korsgaard concludes that there are constitutive standards governing
the activity of action itself. Coupled with her view of the human condition as a predicament in which our rationality and self-consciousness “condemns” us to find reasons for thought and action, her conception of agency as a determinate activity in which we are always already engaged entails that we are subject to the standards constitutive of it.

On Korsgaard’s view, the normativity of practical reason—the binding obligatory force of the practical “oughts” supplied by our maxims—therefore derives from the universal reach of certain authoritative standards which are prior to any subjectively held practical principles. The standards in question are those that Korsgaard believes to be pragmatically inescapable for any agent; it is from their priority so construed that they derive their unconditionally binding authority. Such norms are constitutive for any action at all and as such impose unconditional requirements on any subject that acts, for one’s movements only constitute an action insofar as they are governed by the standards that distinguish action from non-action in the first place.

So the source of Korsgaard’s insistence upon the necessary unity of the active self is really twofold, though the two ideas dovetail in her analysis. The intuitive source of the insistence, supposedly drawn from our everyday understanding of the normative contours of our own agency, is that “actions must be assignable to the person as a whole” in order for her to be able to regard herself, rather than forces at work within or without her will, as their (sole) author. The metaphysical source of the insistence, supposedly derived from an area of overlap between Plato’s and Aristotle’s metaphysics, is that the authority of norms governing activities can best be accounted for by reference to the standards constitutive of the activity in question.
To sum up. According to Korsgaard, *action is self-constitution*. Actions make those who carry them out the cause of whatever is to be realized through some appropriate pattern of behavior: “To act is to constitute myself as the cause of some end.” Agency is exercised through one’s making *oneself* the self-governing cause of something brought about through one’s behavior. Recall that, in a way, this whole picture is all about what is involved in being responsible. The idea, then, is that one can only take responsibility for something brought about through a pattern of behavior if one brings it about oneself *as a whole*, as its single author. Otherwise, Korsgaard insists, whatever is brought about is not an end realized through action performed by an agent, but a mere event produced by causal, whether psychological or physiological, forces at work within one. The power of determining oneself to be the cause of an end of action—“self-determination”—is the essence of agency, the basis for the difference between an action and an *event* as a pattern of causally generated behavior.

*Integrity*

We are now in a position to understand the role played within Korsgaard’s theory by her representation of integrity. Integrity for her is not an adventitious virtue of character which fully formed agents can either have or fail to have, but is constitutive of agency itself. An agent’s integrity consists in the psychic unity that we have seen she thinks makes being a self-determining causality possible. Integrity so construed can be achieved, Korsgaard believes, only through conformity with the constitutive standards for action, since only these lay down the guidelines for agents’ inescapable efforts to unify

---

90 Korsgaard 2009, p. 79
their own psychic and volitional lives. If we must always act, and action is
d paradigmatically a self-constituting, self-unifying activity, then we must always take
ourselves to be subject to the conditions requisite for a pattern of behavior to constitute
an action. Since she argues that these are, in turn, the hypothetical and categorical
imperatives of Kantian moral philosophy, integrity-as-unity—a structural requirement for
the proper constitution of an agent’s psyche—and integrity-as-moral-rectitude—a
normative ideal for conduct—are really two dimensions of one and the same
phenomenon.

So for Korsgaard the prospects for personal practical integrity are contingent upon
one’s conformity with the constitutive principles of volition and action as such. Good
actions are those characteristically performed by a good agent in the sense of someone
who is good at being an agent. Such a person self-consciously regulates her conduct in
light of the volitional conditions for the very possibility of being an agent. Only such a
person has integrity because only she conforms to the practical law which makes action,
and hence being an agent, possible in the first place: “The necessity of conforming to the
principles of practical reason comes down to the necessity of being a unified agent.”91

**Efficacy and autonomy**

We’ve seen that, by Korsgaard’s lights, regarding oneself as the self-determining
cause of an end realized through one’s own movements requires seeing oneself as an
enduring, unified subject. Agency presupposes psychic wholeness, and it is in the
unification of our volitional life made possible by the Kantian imperatives of practical

---

91 Korsgaard 2009, p. 25
reason that integrity as a crucial personal normative achievement consists. For the hypothetical and categorical imperatives respectively command agents to adopt effective means for realizing the ends they set and to set only those ends which have a suitably universal formal character. The former specifies the constraint of efficacy on action; the latter imposes a requirement of autonomy. Efficacy and autonomy so construed are standards which supposedly constitute and govern the practice of agency itself.

On this view, the difference between good and bad action only shows up against a prior constitutive standard, one which specifies the fundamental conditions for something’s being an action in the first place. One is not acting poorly when one’s efforts fail to be subject to the requirement of making oneself, *as a whole*, the effective cause of one’s aims, for instance. In such a case one simply *fails to act*, and this is because in such a case there is by hypothesis *no one agent*, literally *no one around*, to serve as the locus of effort and responsibility. Hence autonomy, as the basic requirement that each agent perform as a single, unified locus of responsible choice, vindicates its status as a valid, inescapably binding norm of agency. Similar reasoning applies to the constraint of efficacy supposedly imposed by the Kantian hypothetical imperative.

Now remember why psychic wholeness and volitional unity are supposed preconditions of agency. It is supposed to be impossible for us to regard ourselves as responsible for the realization of an end through the exercise of our will as a self-determining causality unless we see ourselves as doing so *as a whole*. But wholeness is not the whole picture. We also have to be able to see ourselves, as a whole, as empowered to stand aloof from each and every motivational incentive operating upon our will [“potentially motivating factors” such as occurrent drives, impulses, desires, etc.].
in order to authoritatively order and govern their influence over our movements. Korsgaard explains the connection between having and exercising a will in her sense and the idea of making a personal commitment in terms of the distinction between deciding to do something and predicting that one will do it: “…deciding is different from predicting. Deciding is committing yourself to do the thing. That is another way of saying acting is determining yourself to be a cause.”

According to Korsgaard, agency is thus a matter of making yourself the cause of a particular pattern of behavior which thereby becomes an action responsibly performed by you as its author—your action. The idea of the action being yours is the same, and seems meant to have the same sort of appeal, as that of private property: something belonging to you alone, being your exclusive dominion or jurisdiction, exclusively subject to your unfettered discretion and so something for which you are alone ultimately responsible. The contrasting case where action is concerned would seem to be a lack of ultimate first-personal authority over your mental or bodily movements deriving from your inability to carve out a path from those suggested by your occurrent motivations as the one you have chosen. Where one lacks discretion or control over something’s disposition, the possibility of ownership or of radically asymmetrical entitlement, is attenuated accordingly.

Korsgaard poses the question she takes to be basic for action theory, and states her response in terms of the notion of mineness which we already encountered, thus:

What makes an action mine, in the special way that an action is mine, rather than something that just happens in me? That it issues from my constitution, rather than from

---

92 Korsgaard 2009, p. 77
some force at work within me; that it is expressive of a law I give to myself, rather than a 
law imposed upon me from without.\footnote{Korsgaard 2009, p. 160}

5: phenomenological foundations

So the Kantian imperatives express constitutive standards for action, standards 
which govern any attempt to undertake a performance assignable to one as its author. 
Since action requires normative attribution to a person as a whole, that is, as a 
volitionally integrated locus of responsibility, every action is \textit{inter alia} aimed at 
constituting the self that acts, and the constitutive standards of action tell agents how to 
do this. That is, the Kantian imperatives tell us how to unify and integrate our lives 
through our conduct so that we can fully realize our agency. Fully realized agency in this 
sense is a regulative ideal, a universal paradigm for agents, and it is a purely formal one 
since the Kantian imperatives express procedural constraints on the exercise of our end-
setting capacity. Our performances count as actions and we count as self-guiding 
subjects to the extent that we effectively strive to satisfy the requirements of efficacy and 
autonomy imposed by the standards constitutive of the activity in which we are 
necessarily engaged as agents.

But even if we grant that the Kantian imperatives are constitutive for the activity 
of self-governed self-constitution, we may still wonder why we should believe that our 
actions are subject to them. Here’s why. So far we’ve reviewed Korsgaard’s intuitive 
case for thinking that, as rational agents, we have the power of willing. And we’ve 
reviewed the psychology of action which she thinks best accounts for this power. But we 
should accept Korsgaard’s claim that we \textit{must} take ourselves to have this power only if 
we really cannot do without the conception of ourselves as self-governing subjects
described by it. Korsgaard presumes we all implicitly share this self-conception, but even at this point in the argument, we may wonder why we should buy into this picture of what we are all really like. Can’t all the universalistic talk of normative self-government, self-determination, self-constitution, and so on, be so much mythological or rhetorical vapor?

As before, Korsgaard has an answer ready, and the strategy it embodies infuses her analysis. Like Kant, she ultimately recommends her account of agency and normativity on the strength of its phenomenological support. Its central picture of what human action is like and why it is special offers us a philosophical vindication of our supposed self-conception as free deliberators, first-personal sovereigns governing our volitional lives by our own reflective lights. It is in such a phenomenological register that, at the outset, Korsgaard observes in a loosely Sartrean vein that “the human plight” consists in being “condemned to act”: “We must act, and we need reasons in order to act.”

This is the lynchpin of Korsgaard’s phenomenology because, as we know by now, she is ultimately out to defend a series of related claims about the nature of action and of the authority of the Kantian principles of practical rationality, the hypothetical and categorical imperatives. She wants to show that the function of an action is to constitute the identity of a subject as its author: “An action chosen in a way that more successfully unifies and integrates its agent is more authentically, more fully, an action, than one that does not.” And she wants to show that the Kantian imperatives have unconditionally binding authority on all rational agents because they make action so construed possible in the first place. As she explains in an early chapter, “...I am going to argue that the

94 Korsgaard 2009, p. 1, 23
principles of practical reason are principles by means of which we constitute ourselves as unified agents. And I will argue that that explains their normativity. The principles of practical reason bind us because, having to act, we must constitute ourselves as unified agents.\textsuperscript{95}

Framing the aspirations of Korsgaard’s project in this way helps us to see that everything hangs on whether we assent to what she means when she claims that human beings must act and need reasons in order to do so. What this means for her is embodied in her Kantian doctrine of the will as a self-determining causality and the correlative claim that normative self-government is the essence of rational agency. The point of these abstract formulations is to capture the philosophical import of the form of self-consciousness Korsgaard ascribes to human beings, namely “consciousness of the grounds of our own beliefs and actions,” which “makes it necessary to take control of our beliefs and actions.”\textsuperscript{96} Her account echoes Frankfurt’s by holding that self-consciousness is what fragments the human psyche: amidst the wide range of alternative possibilities for belief and action seemingly available to us, we find that the question of what to believe and do is not settled for us by our instincts. Rather, through the use of our critical powers, we must determine for ourselves in each case how to proceed and why. The principles of practical reason lay down binding requirements for how to act in each case because they are constitutive of action itself. They reunify and integrate our psychic and volitional lives such that we are whole and effective, which we must be if we are to be agents at all.

\textsuperscript{95} Korsgaard 2009, p. 25
\textsuperscript{96} Ibid., Preface xi
We may now rephrase our skeptical challenge as follows. Suppose we grant that we are conscious, or reflectively aware, of the potential grounds of our beliefs and actions. That is, suppose we are aware of the conflicting incentives to belief and action presented by our consciousness of the world and of our own experience. Why suppose that we may therefore take control of what we think and do in light of this awareness? Korsgaard answers that the proof that we may take control of our own conduct can be found in the content of our lived experience of choice and action. She adduces a number of illustrative cases. We will consider two of those featuring most conspicuously in her narrative. The first addresses the first-personal volitional control which the Kantian doctrine of the will ascribes to rational agents. The second is meant to bear out the way in which deliberation serves to unify and so constitute the self that acts.

**Volitional Control**

First, Korsgaard appeals to the experience surrounding one’s decision to work productively, say on writing a book, throughout the day. She uses this case to illustrate the way in which the hypothetical imperative unifies and constitutes the will by requiring one to take effective means for realizing one’s ends. Here is how she describes the experiential contours and motivational pitfalls of choice in an everyday case like this one:

At this moment, now, I decide, I will, to work; at the next moment, at any moment (importantly, maybe even at this moment), I will certainly want to stop. If I am to work I must will it—and that means I must determine myself to stay on its track. Timidity, idleness, and depression will exert their claims in turn, will attempt to control or overrule my will, to divert me from my work. Am I to let these forces determine my movements? At each moment I must say to them: ‘I am not you; my will is this work.’ Desire and temptation will also take their turns. ‘I am not a shameful thing like timidity,’ desire will say, ‘follow me and your life will be sweet.’ But if I give in to each claim as it appears I

---

97 This is a crucial step of Korsgaard’s argument, though she makes no real effort to vindicate the claim. She seems to think that it is patent once we introspectively examine any situation in which we are faced with a choice, which for her is any situation in which we are conscious.
will do nothing and I will not have a life. For to will an end is not just to cause it, not even if the cause is one of my own desires and impulses, but to consciously pick up the reins, and make myself the cause of the end.\textsuperscript{98}

I have included so much of this passage because it conveys a number of salient features of Korsgaard’s phenomenology of choice and action. Three are especially relevant to us.

First, there is a strict partitioning, or bifurcation, of the psyche sufficient to underwrite the power of volitional self-monitoring which the doctrine of the will ascribes to all rational agents. The claim is that one’s consciousness, the reflectively mediated awareness of one’s situation, always stands ready to obtrude itself into one’s activity, distancing one’s focal awareness from what one is doing to allow one “at any moment,” or even “at each moment,” to exert one’s will in the face of contrary motives. Call this Korsgaard’s \textit{Reflective Distance Thesis} (RDT). Here is Korsgaard on the reflective distance opened up by our conscious awareness of our own incentives to belief and action:

\begin{quote}
The first result of the development of this form of self-consciousness is liberation from the control of instinct. Instincts still operate within us, in the sense that they are the sources of many of our incentives...But instincts no longer determine how we respond to those incentives, what we do in the face of them. They propose responses, but we may or may not act in the way they propose. Self-consciousness opens up a space between the incentive and the response, a space of what I call reflective distance.\textsuperscript{99}
\end{quote}

Consider this idea of reflective distance in light of Korsgaard’s description of the everyday case of trying to remain motivated to pursue our ends in the face of contrary inclinations. We notice that the deliberative payoff of RDT is that our \textit{active self}, the responsible center of agency, the self-determining locus of deliberative authority, is always sufficiently distanced from our otherwise engaged activity that it (that is, that \textit{we}

\textsuperscript{98} Korsgaard 2009, p. 69
\textsuperscript{99} \textit{Ibid.}, p. 116
construed as, as identified with, our will) can carry out its (our) constitutive business of motivational surveillance and volitional enforcement. The will must be seen as always monitoring since it must be prepared to enforce at any time.

Second, there is the claim that, owing to the partitioning of the psyche by RDT, exerting one’s will in the face of contrary motives is necessary for sustaining one’s committed activity. To stay with Korsgaard’s example, in the presence of some incentive contrary to my end of remaining productive, the reflective distance opened up by my consciousness of the incentive as a possible ground of action requires that I myself, that is, my will as a causality, intervene to resolve the matter: “If I am to work I must will it.” Call this Korsgaard’s Necessity of the Will Thesis (NWT). Here she is not making the clearly false assertion that sustaining an activity necessarily requires an exercise of the will at every moment of action: perhaps we’ll get lucky and our commitment to our end and our inclinations will align splendidly. But should contrary motives appear, she is committed to the idea that an exercise of the will is necessary if I am to “determine myself to stay on its track.”

Finally, there is the claim that exercising one’s will in the face of contrary motives is sufficient for determining the character of whatever one wills. Call this Korsgaard’s Sufficiency of the Will Thesis (SWT). This formulation of SWT is more abstract than we might like because Korsgaard never precisely specifies the range of

---

100 This is, or appears to be, the Kantian will as the panopticon of Foucault’s Discipline and Punish. Since the one “practical identity” that is necessary and universal for human beings is our identity as rational agents (as self-determining wills), RDT entails that the deepest, constitutive structures of human identity are panoptical and disciplinary, for they are designed to provide an ongoing, double function of surveillance and enforcement. Of course Foucault and Korsgaard part ways over the question of the contingency of this structuring of agency. For Foucault, it is a historically contingent and ultimately variable formation of social power. For Korsgaard, RDT expresses a fixed and universal structure: the very possibility of rational agency, of willing, is defined in its terms. We will address this question directly in the following chapter.
possible effects which the will is specially empowered to bring about. However, two
domains of direct volitional control through the will’s exercise seem vouchsafed by
SWT.

First, it seems clear that Korsgaard thinks an exercise of one’s will is sufficient to
establish any end of action (some particular act or acts performed for some particular
reason or reasons) as one’s willed end (SWT-a). That an agent is guaranteed control over
this much is required by the very definition of the will as a self-determining causality.
For, by adopting principles for the governance its own conduct, a self-determining
causality sets its own ends rather than having them established for it, say by its instincts.

Second, the doctrine of the will entails that agents are by their very nature
empowered to strive to make an effort of a certain kind—to make an internal volitional
effort to make an external bodily effort—even though the external effort may founder on
all sorts of natural and interpersonal obstacles (SWT-b). For a self-determining causality
is a power of making oneself the cause of an effect in the world. Though talk of
“internal” efforts may sound odd, Korsgaard is clear that “not every action is physical:”

…it is not accidental that the only thing I can do without a method or a means is effect a
change in myself. It does not have to be in my body, but it does have to be in myself.
Even when we imagine cases in which someone produces an effect magically, we
imagine the agent disposing himself somehow: chanting, or concentrating, or staring. For
that matter, even when we imagine God creating the world, we imagine God most
immediately effecting a change in himself: thinking a thought, or perhaps even uttering
one, like ‘let there be light.’ So although the movement I am talking about [moving one’s
limbs] is the effecting of a change in the world, it is essential to the idea of action that the
agent produces the change in the world by producing a change in himself. The reason all
this matters is that it is another way of saying that action essentially involves self-
determination.\footnote{Korsgaard 2009, pp. 95-96}

In other words, SWT should come as no surprise when we reflect on the nature of
the Kantian will. For willing an end just is the power of committing oneself to bringing it
about through one’s efforts. And insofar as making the relevant efforts is up to one in the

\footnote{Korsgaard 2009, pp. 95-96}
robust sense which the will as a self-determining causality alone can supposedly
underwrite, then making the efforts itself must also require, but only require, the will’s
eexercise. Distinguishing these two sub-theses (a & b) of SWT thus helps us to clarify an
ambiguity lurking in Korsgaard’s usual way of characterizing the self-determination of
which she thinks rational agents are uniquely capable:

A human being acts when she self-consciously determines herself to be the cause of a
change in the world. Unlike the other animals, we are conscious of our causality, and it is
therefore up to us how we exercise it. We choose the laws or principles of our own
causality.\textsuperscript{102}

“Choosing the laws of our own causality,” understood as setting our ends by freely
adopting our volitional principles, is what SWT-a asserts. But in order for our end-
setting capacity to have any sufficiently reliable, concrete payoff in our volitional lives,
we as agents must also be equally assured of at least being able to \textit{endeavor to make} overt
efforts to promote our ends. Notice that \textit{making} the overt efforts is not something that
even a self-determiner can expect to count upon in a causal and social order full of
potential hindrances to successful action. Someone or something might always get in our
way, hold us fast or otherwise gum up the works in the very moment of overt action. But
without the guaranteed ability to at least \textit{try} to make overt efforts to promote our ends
(SWT-b), we would simply fail to count as self-determiners in Korsgaard’s robust
Kantian sense.

For instance, if I am chained to a boiler in a villain’s lair, his well-fastened fetters
may prevent me from so much as lifting a finger to liberate myself. But my will
empowers me to make the effort, however futile it may be. To be clear, the claim here is
not that there is some reliable pattern of physiological or psychological activity with

\textsuperscript{102} Korsgaard 2009, p. 127
which our effort to make a volitional effort can be identified in concrete practical contexts. There may well be some such reliable correspondence between ‘internal’ and ‘external’ effects of our will, but whether there is or is not is supposed to be of no concern to Korsgaard, for she joins Kant in sharply distinguishing discourse about purely causal mechanisms from talk about rational agency as self-determination. As we saw when we reviewed the Kantian distinction between rational beings and everything else in nature, exclusively causal discourse supposedly cannot suffice to capture or validate our self-experience as agents. This is why she believes phenomenological support is both necessary and sufficient to authenticate her claims about what human agency is like.

So when I say on Korsgaard’s behalf that having a will empowers me to resist even when escape is hopeless, I am not talking about straining my muscles or gritting my teeth, though these may coincide more or less reliably with the supposed internal volitional act. There is nothing relevantly necessary about such coincidences. Rational necessity, not causal necessity, is what’s decisive within Korsgaard’s Kantian psychology of action, and it’s in the sense of rational necessity that I am assured of my ability to resist a causally irresistible force like the villain’s chains. Our distinction between ‘internal’ and ‘external’ volitional efforts helps illuminate the cardinal distinction between rational and causal necessity which is so critical to Korsgaard’s project: “The ideal of agency is the ideal of inserting yourself into the causal order, in such a way as to make a genuine difference in the world.”¹⁰³ One can only sensibly talk of inserting oneself into the causal order if there is some good sense in which one exists and acts independently of it. The requisite independent realm of existence and action just is the “internal” domain of the will we have identified. It is the sphere in which each

¹⁰³ Korsgaard 2009, p. 89
individual, even as a concrete, non-ideal rational agent, nevertheless enjoys a limited measure of *guaranteed* volitional control over her own conduct. SWT-(a-b) maps out two salient regions within this internal jurisdiction.

In short, *determining ourselves to be the cause of changes in the world* can mean either framing an end or initiating the attempt to realize it. Korsgaard’s Sufficiency of the Will Thesis empowers self-determining, that is, rational, agents to perform either species of internal act with unconditional discretion.

There is one last implication of Korsgaard’s phenomenology that we should highlight. In an important sense entailed by our distinction between internal and external volitional efforts, willing is a private affair. When Korsgaard says that, quite generally, one “produces a change in the world by producing a change in himself,” it seems to follow from our characterization of the internal domain of volition that setting an end to be realized through my action is something which I do *by myself*. Now we should be careful here. As one agent among many in a community of discourse and action, I may (even ideally) require all sorts of help in accomplishing the ends I set. After all, our external or bodily mediated efforts to realize our ends may misfire, backfire or otherwise fail due to the interference of the world or other agents. Korsgaard thinks this much is true. However, she insists that even any collaborative undertaking must ultimately be traced to exercises of individual agents’ wills.\(^{104}\) So on the view offered, I cannot fail to make or maintain a commitment without consciously refusing to perform, or

---

\(^{104}\) Korsgaard defends an account of personal interaction in terms of the idea of “uniting wills” through cooperative reasoning and a sharing of ends (2009, Ch. 9). But within her account, the self-determining individual enjoys supreme analytical priority. Indeed, as we have already seen, that is precisely how rational individuals as such are metaphysically individuated. Idealized individual willing (the primary analysandum of *Self-Constitution*) furnishes the paradigm case, or basic analytical horizon, against which joint deliberation and collective action is to be understood.
subsequently determining myself to abandon, the internal volitional act whereby I “consciously pick up the reins, and make myself the cause of the end.”

For Korsgaard, then, my conscious, ‘internal’ volitional effort is necessary and sufficient for both setting and ‘privately’ endeavoring to overtly promote an end. This is required by the doctrine of the will because having a will just is to have and rely upon this capacity for self-motivated conduct. The world or other people may well frustrate my efforts to accomplish the realization of the end to which I am committed, but as a rational agent, the questions of whether I am committed and whether I strive to keep faith with my commitment are, by Kant’s and Korsgaard’s lights, simply and entirely up to me. That agency, in the robust sense of self-determination, requires a volitional prime mover is clearly enough suggested by Korsgaard’s decision in the above passage to invoke divine volition as a paradigm for understanding the human equivalent.

To sum up, we may say that though the external, or worldly and interpersonal, success or failure of our efforts to promote the ends we set is never guaranteed merely by the exercise of our wills, Korsgaard’s view requires that the ends set and the (internal) volitional efforts to realize them (SWT-a & SWT-b) are two elements of an internal domain over all of whose possible normative and performative dimensions we as agents have direct and absolute first-personal control.

---

105 Korsgaard 2009, p. 69
106 After all, metaphysically speaking, action just is such self-motivated conduct on Korsgaard’s view.
107 It may help to refer the discussion back to the terminology we introduced in the first section of this chapter where we stated the “intuitive case” Korsgaard makes for her conception of agency. Note that since, in her view, we as agents establish our ends by selecting practical principles, or maxims, and since our maxims express our reasons for action, it is through the absolute discretion to set our own ends asserted by SWT-b that we are free to determine upon reflection the reasons for which we act. This is what we earlier characterized (in Section 1.2) as an ascription on her part of “normative control” to all agents. The form of volitional control we called “psychological control,” that is, first-personal discretion regarding one’s own conduct, or the freedom to determine what one does, we disambiguated and more carefully articulated on Korsgaard’s behalf in our discussion of SWT-a.
Deliberative self-unification

Korsgaard gives an account of “the phenomenology of deliberation” designed to motivate the idea that reflection upon our grounds for belief and action, and the exercise of choice to which it is prerequisite, serve to unify and so constitute as whole the active self. She begins with a concise restatement of the three theses just introduced, reflective distance and the necessity and sufficiency of the will, explaining that when we reflect generally upon the experience of exercising our deliberative powers, we will find that “it is as if there were something over and above your incentives, something which is you, and which chooses which incentive to act on. That much just follows from the fact of self-consciousness. You must decide whether to act on the incentive, and in effect that means you must decide whether to identify your will with it or not.”

Korsgaard immediately suggests that none of our own incentives appear to us anonymously, as though they could belong to anyone. Rather, owing to the fact that our dispositions to belief and action have their source in our guiding principles, which in turn have their basis in “our conceptions of our practical identity, from the various roles and relationships that we think of as giving our lives and actions value,” an incentive arising in our own reflective consciousness characteristically appears to one “as something that wants to be you.” Korsgaard thinks it follows that in cases where our inclinations conflict because only one competing alternative may be pursued, we feel “that we ourselves are being pulled in both directions.” She explains that when we feel torn in this way, “in such a case that is literally true—it is you, your will, your agency, that is in

108 Korsgaard 2009, p. 126
danger of tearing.” This is why it makes perfect sense, in her view, for us to encourage
ourselves or others to decisive action in cases of conflict by saying, “Make up your mind,
or even better, Pull yourself together.” She concludes: “The work of practical
deliberation is reunification, reconstitution: and the function of the principles that govern
deliberation—the principles of practical reason—is the unification of the self.”

6: sovereignty & horizonality

Korsgaard’s brand of idealism underwrites a psychology of action and an account
of agents’ normative powers that is robustly horizontal. To see why, it will be helpful to
consider the sort of normative authority in whose terms Korsgaard’s psychology
articulates agents’ constitutive powers.

Recall our earlier observation that Korsgaard’s doctrine of the will, as a power of
self-determination aimed at psychic unification, represents agents as first-personal
sovereigns exercising constitutive powers of self-governance. Indeed, the figure of the
sovereign nicely represents the conceptual tie between Korsgaard’s psychology of action
and her theory of normative authority. Only an unequivocally empowered, singular locus
of authority, a self-determining center of responsible agency, would be capable of
ordering its own psychic life and univocally directing its conduct with the autonomy
required of a self-governing entity answerable only to laws it freely gives itself.110 This
is, in a nutshell, the transcendental argument Korsgaard offers for the claim that, as
agents, we must take ourselves to be subject to the Kantian categorical imperative. For

109 Korsgaard 2009, p. 126
110 As we’ve seen, Korsgaard’s transcendental analysis does treat both autonomy and efficacy as essential
characteristics of rational agency, but she emphasizes that, in the end, these two standards are really two
sides of one and the same command, namely the categorical imperative which requires
she argues that this command imposes the requirement of autonomy that helps distinguish *willing* as self-determination from random behavior as determined by “alien” causes without one’s volitional self. And she in turn motivates a phenomenological narrative of choice and action designed to help us see that willing is what we must take ourselves to be doing whenever we take ourselves to be acting responsibly—that is, governing our volitional lives by using our powers of critical reflection and ‘the will’ to respectively evaluate the grounds of our beliefs and actions and regulate our conduct accordingly.

Korsgaard does not overtly deploy the figure of the sovereign subject in her articulation of the normative and psychological contours she believes to be constitutive of agency. However, there is an influential notion of sovereignty from which the conceptions she is defending draw much of their inspiration and intended appeal. Judith Butler explains and interrogates this notion of sovereignty at length in her work on hate speech. There she develops an account of linguistic performativity and the constitution of human agents through social matrices of power inhering in and expressed through language. Drawing on the work of J.L. Austin, Louis Althusser, Michel Foucault and Jacques Derrida, Butler argues that human identity is both inaugurated through and potentially threatened by socially instituted and enforced linguistic practices of definition, naming and calling.

In her view, the dual power of language to constitute and assault subjects is far-reaching but nowhere automatic or immediate in its effects. Rather, since linguistic practices are social conventions with contingent, unfolding histories, the agency possible through speech is heavily ritualized and “citational.”¹¹¹ This means that, in and through our activities as linguistic agents, we always draw upon a tradition of past practice and

¹¹¹ Butler1997(1), p. 27
contested semantic and performative resources. Everything we do as speaking subjects is consequently subject to unforeseen dissonances and disruptions that have their source both within and without our bodies. Accordingly, a dominant theme of Butler’s work on hate speech is that attributing an immediately efficacious performativity through speech to language users—promoting “the sovereign conceit”\textsuperscript{112} in hate speech discourse by ignoring the “open temporality of the speech act”\textsuperscript{113}—is a mistake with grave political implications.

Whether we should be persuaded by Butler’s attack on the figure of sovereign agency in speech is a question to which we will return. For now we want to focus on her representation of sovereignty as a radical possibility of agency that Korsgaard’s theory affirms in its own way. As Butler’s counter-assertion of the citationality of discourse suggests, sovereignty in her sense invests subjects with the ability to constitute the content and produce the intended effects of their speech by an unmediated exercise of the will, a conscious volitional faculty of irresistibly naming or calling into existence an effect as the object of one’s intentions. Since Butler is concerned with linguistic practices in her work on hate speech, she casts her central examples of sovereign agency in those terms…

The sovereign conceit emerges in hate speech discourse in several ways. The one who speaks hate speech is imagined to wield sovereign power, to do what he or she says when it is said. Similarly, the ‘speech’ of the state often takes a sovereign form, whereby the speaking of declarations are, often literally, ‘acts’ of law.\textsuperscript{114}

…and in terms of Austin’s conception of illocutionary speech as a species of performative utterance:

\textsuperscript{112} Butler 1997 (1), p. 16
\textsuperscript{113} Ibid., p. 15
\textsuperscript{114} Ibid., p. 16
The illocutionary act is one in which in saying something, one is at the same time doing something; the judge who says, ‘I sentence you’ does not state an intention to do something or describe what he is doing: his saying is itself a kind of doing. Illocutionary speech acts produce effects. They are supported, Austin tells us, by linguistic and social conventions.\footnote{Butler 1997 (1), p. 17}

Now, as we saw, Butler maintains that the dependency of all speech upon histories of conventional usage supported by socially instituted authority renders linguistic agency vulnerable to disruption and failure. Consistent with this, she goes on to deny that any speech act, including ostensibly “illocutionary” performatives, is ever insulated from misfire and counter-appropriation. Part of the political upshot is that hate speech can fail to produce the injurious effects for whose production it is designed:

If a performative provisionally succeeds…then it is not because an intention successfully governs the action of speech, but only because that action echoes prior actions, and \textit{accumulates the force of authority through the repetition or citation of a prior and authoritative set of practices}…When the injurious term injures…it works its injury precisely through the accumulation and dissimulation of its force. The speaker who utters the racial slur is thus citing that slur, making linguistic community with a history of speakers.\footnote{Ibid., p. 51}

Sovereign speech is thus speech which is not paradigmatically subject to deficiency, excess, disruption or failure. An agent that is sovereign in this sense is thus one empowered to engage in activity not paradigmatically subject to deficiency, excess, disruption or failure.

 Accordingly, sovereign agents have mastery, or unequivocal control, over some dimension of their activity.\footnote{Ibid., p. 15 et passim.} At least some of their efforts are immediately first-personally authoritative; nothing without their volitional self can suffice to frustrate or undermine the efficacy of those undertakings. It is in this sense, one adapted from Butler’s narrower engagement with issues of linguistic agency, that Korsgaard affirms the

\begin{flushright}
\small
\textit{et passim.}
\end{flushright}
sovereignty of the agency of which she thinks humans are distinctively capable. For though she acknowledges that my volitional efforts as an agent may indeed fail—the world or other agents may frustrate my attempt to realize some outcome or perform some action which I intend—we have seen that her doctrine of ‘the will’ as a power of self-determination native to agency attributes a kind of normative and psychological invulnerability to the self-consciously critical and end-setting activity represented as paradigmatic for agents.

To see why, consider first the two sub-theses of the Sufficiency of the Will Thesis which we reviewed in the previous section. As a self-determining causality, I simply cannot fail *inter alia* to make a certain end my own if I so choose (SWT-a). Additionally, I cannot fail to commit myself to the realization of an end through my action if I so will it (SWT-b). Moreover, recall that Korsgaard claims that efficacy, as commanded by the Kantian hypothetical imperative, is a constitutive standard of volition. This is another, more programmatic, way of asserting the sovereignty of rational agency. For though concrete human beings lack fully realized sovereign power since others and the world may interfere with the effective exercise of their will, the Kantian conception of agency as rational self-determination invests the paradigm agent with precisely this capacity to immediately bring about those effects it wills. This is just what it means to hold with Korsgaard that agents are paradigmatically subject to a standard of efficacy, or that this standard is *constitutive* for any action: insofar as I act at all, I must constitute myself as a

---

118 Since the Kantian laws of practical rationality have an imperatival form, there is an interesting sense in which Korsgaard’s normative constructivism is founded on a conception of sovereign speech. Therefore, though her concerns are more comprehensive than Butler’s, Korsgaard’s view of the authority of agency falls under the scope of Butler’s question whether sovereign speech as mastery is a viable model for linguistic agency.
unity, as one who acts, by striving to realize my willed ends through my volitional efforts.

To sum up. Korsgaard’s assertion of sovereignty as paradigmatic for human beings as embodied rational agents derives from the fundamental Kantian psychological thesis to which we have devoted so much attention, the doctrine of the will. This is the idea that to effectively bring about one’s end as its self-determining cause is the constitutive aim, the telos, of being an agent. Indeed, this is Korsgaard’s deep analytical rationale for endowing partially realized, embodied\textsuperscript{119} human agents with the one measure of ineluctable volitional mastery available to us—the ability to set our ends and make the relevant efforts, where the ends set and the efforts made are guaranteed to have precisely the normative and performative character we elect to invest them with.

This is what is entailed by the assertion of volitional control on which Korsgaard’s theory relies: to be a rational agent is for it to have to be up to you, full stop, to determine how and why you are to behave. The idea is certainly not that there are no standards beyond your personal whims which have authority over your conduct. The idea is rather that the standards for action to which every agent is subject, the Kantian imperatives, have their authority precisely because they shape and govern the structure of the will that endows us with the capacity for unfettered choice of our ends and of our reasons for action.\textsuperscript{120} Consider the alternative. Suppose that the what and wherefore of our conduct were ineliminably subject to normative or performative deficiency, excess, disruption or failure, and that these constraints were phenomenologically borne out by

\textsuperscript{119} Cf. Butler’s characterization of “the body” as always partly out-of-control, a fund of potentially subversive and transgressive differences: \textit{ibid.}, pp. 10-12 \textit{et passim}.

\textsuperscript{120} As expressed by the maxims or practical principles with which our volitional freedom requires us to identify ourselves as agents whenever we act, regardless of which maxims we adopt.
our experience as agents. If human agency were thus not at least paradigmatically sovereign, what reason could there be to assert that, as agents, we must take ourselves to have first-personal control over our own volitional activity in the ways we have seen Korsgaard ascribes to agents?

If ideally rational agents are fully realized sovereigns in Butler’s sense, then real human agents are necessarily aspiring ones. On such a Korsgaardian view, failures of sovereignty are not reflections of the “citationality of discourse” or any other paradigmatic dimension of the concrete social contexts for agency inhabited by human beings. Instead such failures must always already be seen as deficiencies owing to the insurgent embodiment of our agential capacities within causal and social orders endurably characterized by dynamic cycles of normative flux, eruptive difference and discursive and performative contestation. Relative to the paradigmatically efficacious exercise of the will as a sovereign’s proper power, embodied difference and enduring contestation in human life take on the aspect of threats aimed at the very heart of our highest capacities and so at the possibilities for self-cultivation which they underwrite. Korsgaard’s view therefore capitalizes on a representation of difference and contestation as essentially menacing and lamentable rather than integral to and potentially fruitful for the social and psychic life of human agents. Whether this should raise concerns about the ethical and political tenor of Korsgaard’s theory of agency is a question to which we will return in subsequent chapters.

Being a sovereign subject is to be robustly empowered, endowed with the godlike capacity to call into existence precisely that which the divine agent envisions and wills through an exercise of that selfsame faculty:
According to the biblical rendition of the performative, ‘Let there be light!’ it appears that by virtue of the power of a subject or its will a phenomenon is named into being.\textsuperscript{121}

Against Butler, whose “presumption is that speech,” and so the agency exercised through speech, “is always in some ways out of our control,”\textsuperscript{122} Korsgaard commends her Kantian psychology of

action to us with the thought that “there’s a touch of the divine in being an agent.”\textsuperscript{123}

7: conclusion: three theses

1. Psychical \textbf{wholeness} is constitutive of responsible agency. That is, an agent’s psychic makeup is paradigmatically whole (Psychic Wholeness Thesis), since…
   a. Volitional self-unification is the distinguishing metaphysical function of an action (Action Unity Thesis), and “we must act” (RDT & NWT). It follows that…
   b. The inescapable task facing each agent is self-unification (Unification Task Thesis) through conformity with the standards constitutive of action, which fact makes those standards supremely and universally authoritative (Moral Sovereignty Thesis).

2. \textbf{Sovereign} power: radical, first-personal volitional control is constitutive of responsible agency\textsuperscript{124} (Deliberative Control Thesis). DCT embraces the following volitional doctrines, just reviewed:

\textsuperscript{121} Butler 1997(1), p. 50. Cf. Genesis 1:1-3: “In the beginning, when God created the heavens and the earth, the earth was a formless wasteland, and darkness covered the abyss, while a mighty wind swept over the waters. Then God said, ‘Let there be light,’ and there was light…”; & Genesis 2:19: “Out of the ground the LORD God formed every beast of the field and every bird of the sky, and brought them to the man to see what he would call them; and whatever the man called a living creature, that was its name” (from the New American Standard Bible).
\textsuperscript{122} Butler 1997(1), p. 15
\textsuperscript{123} Korsgaard 2009, pp. 84-85. Cf. Genesis 2:19 in footnote 28 (above). In this biblical passage, humanity’s own sovereign linguistic agency is represented as modeled on God’s power of creation through authoritative naming, and so as deriving from the “touch of the divine” in humanity’s form and likeness to its sovereign progenitor.
a. Reflective Distance Thesis (RDT)
b. Necessity of the Will Thesis (NWT)
c. Sufficiency of the Will Thesis (SWT-(a-b))

3. A ‘final’, that is, an unequivocal, analytically transparent, stable and universal, horizon of intelligibility is required for properly disclosing the constitutive standards of volition and so the conditions of agency as such (Final Horizon Thesis). As we have seen, this requirement is cashed out in terms of a teleological metaphysics of action (AUT) and is entailed by DCT inter alia.

Each claim represented above by an acronym is a distinct thesis about rational psychology or the character and sources of practical normativity, but Korsgaard’s view is situated at their intersection. Sovereignty, wholeness and the figure of a final horizon of analysis—are the three most salient elements of Korsgaard’s theory for our purposes. Jointly they constitute a comprehensive view of the normative, metaphysical and psychological character of agency of which Korsgaard’s project in *Self-Constitution* is a systematic exposition and defense. For ease of reference, we will call the comprehensive view of agency delineated by PWT, DCT and FHT Sovereign Idealism (SI). A few more words of clarification are in order regarding each of the three theses we have isolated.

We’ll take them in turn.

First, we have seen that the Kantian doctrine of the will imposes a strong requirement of first-personal control—paradigmatically speaking, reflective, self-conscious deliberative control—to be exercised by an agent over his own volitional life. We’ve called this paradigmatic requirement the Deliberative Control Thesis. We have

---

124 For our purposes, “responsible agency” entails the capacity for practically responsive uptake of the normatively salient features of contexts for belief or action. As we’ve seen, Kosgaard’s rhetorically charged term for this broad characterization of what essentially distinguishes agency is “rational agency.”
seen that, by Korsgaard’s lights, the ability to exercise such control is established by the reflective structure of human consciousness, and that this capacity entails a heavy burden of first-personal responsibility. In Korsgaard’s terms, the reflective fragmentation of the psyche makes fashioning a coherent identity out of one’s ongoing practical activity an inescapable task for one; the task is one’s own, and the resulting identity is one’s own life, precisely because this self-constituting activity is, after all, what makes one a single self rather than a Humean contingent bundle of desires and perceptions. Here are these two salient aspects of DCT in Korsgaard’s own words:

There is no question that we human beings are self-conscious in a very particular way. We are aware, not only that we desire or fear certain things, but also that we are inclined to act in certain ways on the basis of these desires or fears. We are conscious of the potential grounds of our actions, the principles on which our actions are based, as potential grounds. And this…sets us a problem that the other animals do not have. For once we are aware that we are inclined to act in a certain way on the ground of a certain incentive, we find ourselves faced with a decision, whether we should do that…Our awareness of the workings of the grounds of our beliefs and actions gives us control over the influence of those grounds themselves.  

The identity of a person, of an agent, is not the same as the identity of the human animal on whom the person normally supervenes…We are self-conscious in a particular way: we are conscious of the grounds on which we act, and therefore are in control of them.  

Korsgaard acknowledges the centrality of this assertion within her larger project of developing an account of the normative sources and psychological character of “the psychic unity that makes agency possible.” The idea, familiar from Korsgaard’s earlier work, is a Kantian one. The claim is that because of the self-reflective character of human consciousness, normal individual human beings, qua rational agents, can freely determine and so claim ultimate responsibility for their own conduct.

Next, there is the claim that taking responsibility for what we do requires governing our own conduct through the adoption of principles sufficient to unify our

125 Korsgaard 2009, pp. 115-116  
126 Ibid., p. 19  
127 Ibid., p. 7
psyche. For otherwise we cannot be whole. If what we do cannot be intelligibly attributable to us as a whole, then we cannot be properly said to have done it at all.

Insofar as we act, then, we must be seen as whole. This is the Psychic Wholeness Thesis (PWT). DCT and PWT operate in tandem within Sovereign Idealism:

These conditions—the need to work at being unified and the need for practical deliberation—are brought about together. And this means that the function of deliberation is not merely to determine how you will act, but also to unify you...those are not two different things, for your movement will not be an action unless it is attributable to you—to you as a whole or a unified being—rather than merely to something in you. And the task of deliberation is to determine what you—you as a whole or a unified being—are going to do.128

So DCT and PWT entail that the constitutive task facing agents is that of working at being unified (UTT above) and that first-personal practical deliberation just is the work required.

Finally, one further observation about the absolute priority assigned within Korgaard’s theory to norms said to be constitutive of action tout court. We have seen that her inquiry into the nature of agency is ultimately aimed at disclosing the constitutive standards for action as such, that is, as distinct from but embracing all other discrete forms of norm-governed human activity (like fly fishing, jazz drumming, HVAC maintenance, etc.). But we might wonder why we should invest in the search for such standards. Why think that there really are any norms which govern the activity of acting itself? The suggestion that, in order to properly understand the relationship between human psychology and our normative activities, we have to answer such an abstract question as that may strike us as odd. If you want to develop “a self-conscious appreciation of what you are and of how you work,”129 why not simply conduct an open-ended series of narrower, concrete inquiries into culturally local human practices—say,

---

128 Korsgaard 2009, pp. 125-126
129 Ibid., p. 67
for instance, the ones in which your personal or your community’s idiosyncratic, lived concerns happen to be situated and at stake?

Well, it seems we have reason to invest in the search for constitutive norms for action at least as long as we care about the prior question, What is an action, tout court, as distinct from non-action? But why care about this question? Korsgaard’s answer is that answering this question is a prerequisite to answering narrower questions about specific domains of human activity. Her reason for thinking this derives from her dialectical point of departure, where we saw her establish a cardinal distinction between human and non-human action in terms of the former’s susceptibility of evaluation in terms of rightness and wrongness. And as we saw, morally evaluable conduct is something for which each individual agent must be capable of taking responsibility, something which we must therefore each determine in our own case after the manner of a self-governing sovereign. Understanding what makes the sovereign a sovereign, then, precedes and frames any sensible effort to understand how any particular sovereign functions.

In short, there is such a thing as human activity tout court because human beings are distinguished from non-humans by their uniquely active role in directing their own activities. As agents we must determine our own choice and pursuit of ends, guided by reflection upon our reasons for belief and action. Such is, after all, “the human plight:” “We must act, and we need reasons in order to act.”

Therefore, for Korsgaard, her Kantian question about the constitutive norms governing action has by its very nature absolute priority over any attempt to understand discrete, contextually situated or foreground, agency. The norms in question therefore have supreme authority for every

---

130 Korsgaard 2009, p. 23
agent by virtue of the absolute priority of their derivation: “the laws of practical reason
govern our actions because if we don’t follow them we just aren’t acting, and acting is
something that we must do. A constitutive principle for an inescapable activity is
unconditionally binding.”\footnote{Korsgaard 2009, p. 32}

As we saw in the previous chapter, these methodological commitments are
characteristic of horizontal models of agency. So though Korsgaard’s tacit ascription to
agents of paradigmatically sovereign power overdetermines the ‘final horizontality’ of her
position, her theoretical starting points implicitly require it.
CHAPTER THREE

POWER:

GENEALOGICAL RESOURCES

FOR A “REALIST” EROTIC PSYCHOLOGY

1: introduction: power as a problem of erotic psychology

Let’s take stock of where we are. In our first chapter, we used the figure of Doestoevsky’s Underground Man to dramatize an apparent pathology of psychic life that many believe is symptomatic of late modern social conditions. Following the Underground Man’s own lead by diagnosing his deficiency as a lack of integrity, we identified it as a problem of erotic psychology, since erotic ties are the orienting and motivating commitments which shape our lives and the Underground Man’s failure to take anything seriously undercuts his ability to form and maintain any such attachments. We then introduced the idea of a ‘horizontal’ view of the normative and psychological contours of psychic and social life. Because horizontal views share a broad approach to all sorts of philosophical issues pertaining to agency, they tend to address them in a distinctive way.

---

132 I mean “realism” roughly in the sense defended by Raymond Geuss in Philosophy and Real Politics (Princeton: Princeton University Press, 2008). In general, a realist psychology would remain beholden to observations of the factors driving human behavior as we find it in our concrete experience rather than grounding its judgments in an ideal conception of our normative or psychological powers generated by transcendental or other a priori means. We will be using the term “realism” more narrowly to refer to the psychology of action licensed by the account of power offered in this chapter.
During our discussion of Robert Pippin’s and Harry Frankfurt’s treatments of commitment, we saw that their accounts rest on versions of the “horizontal” strategy for explaining the possibility and character of erotic agency. We saw that horizontal views account for “foreground,” or discrete activity, in terms of “background” normative patterns in the following way.\footnote{As we saw in Ch. II, “backgrounded” normative standards may be said to ultimately correspond to behavioral regularities such as general patterns of normative judgment or dispositions to judge in various ways, formal evaluative criteria (as in Korsgaard’s case), and so on. Naturally, approaches to this question depend on the philosophical allegiances that unify or divide views within the horizontal camp.} First one establishes a moral psychology or social ontology whereby a background “horizon of intelligibility” is said to constitute the domain of norms governing psychic and social life. One then accounts for the possibility and normative status of discrete acts or practices in terms of such ontologically prior social or psychic structures. Agency as such is held to be possible, and particular actions are normatively intelligible as being good or bad, right or wrong and so on, only by reference to normative standards already available for use by those making the relevant judgments. This is the sense in which the background standards furnishing horizons of meaning are supposedly constitutive of the normativity of acts and practices. We further saw that there are some apparent difficulties with the horizontal strategy but postponed examining these until we had a closer look at how it can be applied systematically to the question of what makes erotic agency possible and what ideal of integrity is licensed by the answer it provides.

So in Chapter II we turned to the task of examining a horizontal account of personal integrity. We chose to investigate Christine Korsgaard’s broadly Kantian theory of agency, which we called Sovereign Idealism (SI), highlighting three salient features of its normative psychology and metaphysics of action. These features qualify SI as
horizontal in a robust sense. We showed this by placing Korsgaard’s conception of agency in the context of Judith Butler’s work on hate speech and linguistic agency, finding that Korsgaard’s doctrine of the will as a “self-determining causality” paradigmatic for human agency rests on an ascription of sovereign power in Butler’s sense. For, according to Korsgaard, finitely rational, human agents are necessarily subject to a volitional requirement of sovereign authority and control whose satisfaction unifies our psyches, rendering us whole and so capable of responsible action.¹³⁴

Personal integrity on this view just is psychic self-unification so construed. Those with integrity normatively govern their own conduct; they can act. Those without cannot. From this standpoint, the Underground Man lacks integrity because he lacks the well-ordered psychic activity required of and for a self. Because he compulsively fails to satisfy the volitional requirements constitutive of unified, self-governing subjects, he literally has no life. He is a mess, a heap of stray impulses instead of a whole person with a coherent identity. In short, SI provides a solution to the problem of integrity that rests on a paradigmatic attribution of sovereign agency to human persons.

This is where we left things at the end of the previous chapter. So what, if anything, is wrong with this solution to the problem of integrity posed by the Underground Man? In Chapter II we saw that Korsgaard’s psychology of action appears to stand on impressive phenomenological credentials. Does it not? If not, why not? Moreover, what would a viable alternate picture “of what you are and of how you

¹³⁴ This supported the conclusion we reached during our discussion of Frankfurt’s notion of wholeheartedness that horizontal views of agency and commitment tend to represent human agents as ideally or paradigmatically sovereign.
work.”\textsuperscript{135} look like? And why, in the end, should we prefer it over Sovereign Idealism, which offers us a conception of ourselves as natively free and so securely empowered, and enjoys a venerable pedigree in the history of modern philosophy?

In the following chapters we will try to answer these questions by exploring a hypothesis about the problem of integrity framing our inquiry. The hypothesis identifies and purports to explain an apparent psychological shortcoming of horizontal views that we identified at the end of Chapter I, and it embraces three related claims. \textit{First}, horizontal views are beset with a common explanatory and phenomenological deficiency—a tendency to underestimate the ontological role and potential value of erotic loss and transformation and so to elide the persistent conflict within psychic and social life which makes it possible. \textit{Second}, this deficiency results from a failure to adequately thematize power as the source and medium of contestation so construed. For our purposes, SI exemplifies the horizonal strategy because of its robust commitments and since it is the view to whose details we have devoted the most attention. \textit{Third} and finally, our account of power will suggest a suitably non-horizontal conception of agency. And with this in hand, we will be able to address our problem of integrity from a broadly Nietzschean standpoint, arguing that there is an ethically and politically worthwhile erotic ideal, a conception of personal integrity, that requires an agonistic style of engagement with the range of differences through which the commitments and attachments orienting one’s own agency are sustained and challenged.

In this chapter, we will defend the first two claims by introducing the domain of power elided by horizontal views and showing why SI is especially hobbled by this

\textsuperscript{135} Korsgaard 2009, pp. 66-67. Korsgaard claims the knowledge which philosophical inquiry is after is self-knowledge, “a self-conscious appreciation of what you are and how you work, which will make you better at being what you are and at working in that way.”
omission. The business of the two remaining chapters will be to defend the third claim by articulating and defending an “agonistic” conception of integrity that is on all fours with the conception of power elaborated below.

2: power’s horizons

In order for us to develop a manageable conception of power suitable for our purposes, our treatment must be selective. Recall that we are in the business of doing erotic psychology. This means that we want resources for addressing conceptual and normative issues of self and agency that arise alongside problems of commitment, albeit ones that depart from SI in taking contestation to be centrally important. We want an account of power since the term suggests division, conflict and the possibility of loss and transformation, and we want an account of their role and potential value within psychic and social life. We will therefore use the term ‘power’ to refer to the range of effects and relations responsible for the enduring role of contestation in psychic and social life.

Moreover, whatever else political life is about, it is about power. So in a sense, we are concerned with political psychology, too. Some of the examples we will consider in this and the remaining chapters will help to illustrate why. However, we are not going to address the question of what power is from a primarily metaphysical standpoint, regarding it as a substance with no essential import for issues of selfhood and agency. Rather, we start with the hypothesis that there is something important to be learned by treating contestation as a key factor in the formation and agency of human subjects. ‘Power’ will be our term for whatever effects and relations play the formative role that enables subjects to contest the meanings and values in whose midst they live and act.
Now, as we have seen, Sovereign Idealism is a comprehensive account of the essential psychological constitution and normative capacities of human agents. Because we want a psychology of self and action that is a genuine alternative to SI, we will start by provisionally isolating three consequences of the character, operation and effects of power which correspond to the three primary elements of SI reviewed in the previous chapter inasmuch as they are negations of them. We therefore start with a negative formulation of power—namely that it precludes the formations and exercises of agency required by SI—from which we will proceed to construct a positive representation.

With these qualifications in mind, we offer our negative characterization and suppose that, for our purposes, power precludes the three key elements of SI, namely:

1. **Wholeness** of the psyche and **self-imposed unity** in the volitional activity of a human subject represented as paradigmatic for human agency

2. **Sovereignty**, in Judith Butler’s sense of immediately efficacious first-personal volition, as realizable within certain limits (Sufficiency of the Will Thesis) and otherwise paradigmatic for the whole of human activity (Deliberative Control Thesis)

3. **Stable**\(^\text{136}\) analytical **horizons** of intelligibility for (i) thematizing the conditions for the possibility of the constitution of the active subject, and so for (ii) normatively regulating the exercise of its agency

\(^\text{136}\) In relation to Korsgaard’s account of SI, “stable” entails *apodictic* in the sense associated with synthetic a priori judgment or judgment rendered from the standpoint of ‘pure reason’ in the sense defended within Kant’s practical philosophy, especially in his *Groundwork of the Metaphysics of Morals* and *Critique of Practical Reason*.
As it happens, these three ways in which power precludes the formations of subjectivity and exercises of agency privileged by SI derive from positive representations of its character, operation and effects to be found in work of Friedrich Nietzsche, Judith Butler, Michel Foucault, and Raymond Geuss. In the first main part of this chapter we will draw upon the resource they provide for articulating a conception of power suitable for our purposes, primarily focusing on Nietzsche’s and Judith Butler’s closely related conceptions of the formation and agency of human subjects. In the second main part of this chapter we will defend our conception by highlighting key deficiencies within the phenomenology of choice and action from which SI supposedly derives its authority and sketch an alternative more supportive of our view. Since power precludes the three core features of SI, our general aim is to show that there is broad support within philosophy and social and political theory for skepticism about its representation of subjectivity and for an alternative approach to erotic agency that takes the formative role of power seriously.

What power is, in nuce

There is a common thread underlying the different perspectives on subject formation and agency considered below. It is that power forms human subjects and so shapes our agency in a twofold fashion, as both a productive and a limiting condition of the normatively salient agency possible for us. So construed, power is an enabling constraint that, in its operation and effects, is **productive, excessive and dynamic**. This has a number of implications for erotic and political psychology. Those of greatest
interest here can be represented in terms of the three features of SI that supported our
negative characterization of power. In its role in subject formation and agency, power is:

1. **Productive** of identity/difference (precludes wholeness-as-unity)
2. **Excessive** (precludes sovereign control)
3. **Dynamic** (precludes final horizons)

Since we are framing our account of the salient relations between power, subject
formation and agency as an alternative to Sovereign Idealism, and since power in our
sense underwrites enduring contestation, we will call it **Agonistic Realism** (AR). In each
of the two main parts of the remainder of this chapter, we will set out one of the first two
distinguishing characteristics of AR just enumerated.

In the first part, we will motivate the first aspect of AR—the ‘productivity’ of
power through contestation—by turning to genealogical and psychoanalytic accounts of
subject formation. We will start by reviewing the part of Nietzsche’s attempt to pursue a
critical history and interrogation of the “value of our values” in the second essay of his
1887 work, *On the Genealogy of Morals*. We’ll supplement and clarify our findings by
turning to Judith Butler’s analysis of subject formation and subordination in *The Psychic
Life of Power*. Then in the second part, we will return to Nietzsche’s *Genealogy*, and
turn again to Butler’s work on linguistic performativity, to shed light on the character of
the agency possible for the self so constituted, explaining the sense in which it is
‘excessive.’ We will postpone discussion of the ‘dynamism’ of power until the next
chapter where we will consider AR’s implications for the notion of a ‘horizon of
intelligibility’ as it pertains to substantive and formal issues within erotic and political psychology.

3: the conception

*Nietzsche’s psychosocial history of “the sovereign individual”* & *“his conscience”*

Nietzsche’s *Genealogy* is a sensible place to start looking for psychological resources for constructing the sort of alternative to Sovereign Idealism we are after. What we want is a conception of the active self that arises out of an account of power as the formative occasion for self and responsible action, and the *Genealogy* is in many ways the *locus classicus* for this sort of account, one that is variously developed by the other writers whose views will occupy us below. We will dwell especially on Nietzsche’s interpretation of the form of subjectivity Judeo-Christian ideals inscribe insofar as it corresponds with the ideal of agency established by Sovereign Idealism. For ease of exposition, we will refer to the view of the subject whose origins and value Nietzsche wants to diagnose as Moral Personhood (MP). We will find that there is a great deal of similarity between MP and SI, and that Nietzsche’s genealogical account helps explain why the notion of the active and responsible self which they inscribe has taken hold within our way of life. In this way the understanding of power we glean from the *Genealogy* will already be indexed to the normative and psychological concerns framing our inquiry.

Yet our aims in the present section are modest in an important sense. Like all of Nietzsche’s mature works, the *Genealogy* is an ambitious and richly ambiguous text. We
will not attempt to offer a systematic exegesis of the work as a whole. Instead we will proceed by moving between its companion essays, highlighting salient perspectives on the human psyche and agency staked out by Nietzsche in his efforts to expose the contingency, artifice and normative ambiguity of the developments which he believes promoted Moral Personhood as an ideological and institutional paradigm.

We just said that the values whose history and value Nietzsche is concerned with are those of Judeo-Christian morality and the view of the human person they require, and that he takes them to be “ours” in an important sense. To that extent he is in accord with Korsgaard, whose phenomenological defense of Sovereign Idealism is supposed to vindicate our everyday consciousness of ourselves as freely self-determining practical deliberators.\footnote{137 Nietzsche likens the phenomenon to be explained as the ripened fruit of an ancient tree, that for whose production the long pre-history of “society and the morality of custom” have been the means: We discover that the ripest fruit is the *sovereign individual*, like only to himself, liberated again from morality of custom, autonomous and supramoral (for “autonomous” and “moral” are mutually exclusive), in short, the man who has his own independent, protracted will and the *right to make promises*—.\footnote{138 What distinguishes Nietzsche’s approach is that he does not take the fact that we tend to regard ourselves as having sovereign jurisdiction over our own psychic and volitional lives to be conclusive. Rather he wants to raise doubts about this self-understanding and about the system of ethical attitudes and practices reliant upon it.}

What distinguishes Nietzsche’s approach is that he does not take the fact that we tend to regard ourselves as having sovereign jurisdiction over our own psychic and volitional lives to be conclusive. Rather he wants to raise doubts about this self-understanding and about the system of ethical attitudes and practices reliant upon it.

\footnote{137 However, we shall see that Nietzsche does not take the “everyday” phenomenology of choice and action to be as uniformly supportive of Kantian intuitions about agency and value as Korsgaard does. See GM I:16.}

\footnote{138 Nietzsche, GM II: 2}
“Morality” is Nietzsche’s name for this ethical system. He describes it as an altruistic mode of evaluation that privileges the “’unegoistic,’ the instincts of pity, self-abnegation, self-sacrifice.” Identifying these attitudes as “our moral prejudices,” he announces the emphatic ‘this-worldliness’ of his method by telling us that he has “ceased to look for the origin of evil behind the world.” Rather, he seeks to prepare the way for a sobering, historically informed assessment of the value of morality by asking “under what conditions did man devise these value judgments good and evil? and what value do they themselves possess?” The Genealogy is thus an inquiry into the origin and value of morality construed as a complex, socially instituted perspective on the ultimate meaning and value of human conduct and on the nature of the subject who acts. In short, Nietzsche’s task in the Genealogy is both diagnostic and critical. He asks, “how did we come to understand ourselves and our lives in this way; and, in the end, is it a healthy way of understanding what is possible and desirable for us?”

Of course, Nietzsche is a moral skeptic. That is, he recognizes the historical scope and social ascendancy of morality’s perspective on value and human nature but denies its transcendental status and so its universalistic claims to unconditional normative authority. Instead he maintains that the moral perspective is an artificial one arising out

139 In what follows, we will use the term “morality” to denote the object of Nietzsche’s investigations in his Genealogy. We thus follow not only Nietzsche’s usage but that of Bernard Williams, who, in his Ethics and the Limits of Philosophy (1985), describes morality as a special sub-system of ethics having many of the same features examined in the Genealogy. In Ethics and the Limits and elsewhere, Williams argues that “the morality system” presupposes a special and particularly demanding notion of obligation as well as a peculiar conception of the self, namely one capable of discharging such obligations. Like Nietzsche, Williams urges skepticism regarding the moral mode of practical evaluation and the view of the human person it requires.
140 GM Preface: 5
141 Ibid., Sec. 2
142 Ibid., Sec. 3
143 Here we cannot venture to say whether morality so construed—the object of Nietzsche’s inquiries in the Genealogy and elsewhere—and Korsgaard’s Sovereign Idealism are necessarily coextensive. But, as we have already indicated, we will be focusing on those of morality’s features for which we find systematic support in SI.
of a series of transformative clashes among different elements of self and society. It is therefore psychologically and historically contingent but has real lasting consequences for how those under its influence understand themselves, their conduct and the social worlds they inhabit.

By Nietzsche’s lights, morality therefore endures as a diffuse social institution with an unfolding history and concrete implications for human values and psychology. Each of the *Genealogy*’s three companion essays focuses upon a different set of phenomena associated with it. When Nietzsche says that he wants to interrogate the value of our values, he means to examine some ways in which human possibilities are shaped by the psychological and institutional impact of the moral standpoint. The *Genealogy*’s First Essay is concerned with the moral mode of evaluation and its insistence upon the distinction between “good and evil,” as opposed to “good and bad,” as the supreme standard for judging the character and actions of agents. The Second Essay is concerned with the character of moral agency itself, namely the conception of the human person and its distinctive powers as required by morality. The Third Essay closely examines “asceticism” as an ethical ideal deeply interwoven with morality’s conception of value and human personhood.

The first two essays of the *Genealogy* thus constitute an account of the internal deepening and structuring of the human psyche in ways that transform human capacities for action (Second Essay), and of the origin and value of discourse representing these new capacities as metaphysically and ethically necessary (First Essay). Nietzsche represents these capacities and the discourse supporting them as historical artifacts in the first instance and so as contingent, not metaphysically transcendent as within SI, and as
constituted by a series of transformative struggles. In short, he constructs a psychologically oriented account of the social transformations through whose unfolding history much of modern experience and institutional life has come to be guided by the presumption that individual human persons as such freely determine and are answerable for their own conduct. But, again, Nietzsche is not only telling us a story of how certain ideas arose and took hold.

The opening lines of the Second Essay plainly state the phenomena to be explained: “To breed an animal with the right to make promises—is not this the paradoxical task that nature has set itself in the case of man? is it not the real problem regarding man?” Nietzsche wants to suggest what it takes, at the bodily and social—the psychic, linguistic and institutional—levels to bring such a subject into existence: “this precisely is the long story of how responsibility originated.” As we saw, to see a subject as self-responsible is to see it as sovereign over its own psychic and volitional life. The sovereign individual is marked by “the proud awareness of the extraordinary privilege of responsibility.” His consciousness of this “rare freedom, this power over oneself and over fate” penetrates “to the profoundest depths and become instinct, the dominating instinct.” In other words, the assumption of our self-determining freedom has become the dominant instinct for modern individuals. It is the second nature of the “sovereign man,” and he “calls it his conscience.”

There is thus a sense in which Nietzsche and Korsgaard are after the same thing. They each want to disclose the conditions for the possibility of there being reflectively self-conscious, responsible subjects; human beings “with the right to make promises;”

---

144 GM II: 1
145 Ibid., Sec. 2
146 Ibid.
sovereign individuals endowed with a conscience. And both Korsgaard’s Sovereign
Idealism and Nietzsche’s genealogy of the sovereign subject underwrite a representation
of the normative powers exercised by agents so construed. For whereas Genealogy II
offers an account of moral subject formation, Essay One’s narrative of the origins of
moral discourse contains an account of the linguistic agency through whose exercise this
discourse arose and won out in competition with rival vocabularies of the self.

Of course the sort of accounts of the constitution and powers of the moral subject
respectively offered by SI and Nietzsche are quite different in spirit. We have already
indicated the essential difference. SI is a teleological account meant to disclose the
conditions for the possibility of responsible agents understood metaphysically.147
Whereas Nietzsche’s genealogical hypotheses primarily function as claims about fluid,
historically contingent social and psychological formations. But we will have more to
say about this in a later section where we discuss the sense of a “horizon” of analysis
operative in our account of power, subject formation and agency.

So what does Nietzsche think it takes to breed such an animal? What is the
formative occasion for self-responsible human subjects, according to the Genealogy? We
will find our answer in the Second Essay. Recall that in this chapter we are offering a
threelfold characterization of power as integral to agency, each of its three elements
corresponding negatively to those in whose terms we defined SI. In Genealogy II,
Nietzsche shows us that the formation of the ‘moral person’ is not a metaphysical fait

147 I think this is a fair characterization of SI’s doctrine of “self-constitution.” For as we saw in Chapter II,
Korsgaard rests her account of the “constitutive principles of volition” on a prior, teleological metaphysics
according to which normative standards confer unity on the phenomena to which they apply, and so the
standards for proper action confer unity on the selves that act: “action is self-constitution.” We will
consider some further implications of this difference between SI and Nietzsche’s account of power at the
end of the present section.
accompli as it would be through all agents’ necessary subjection to formal rational
requirements, but is rather productive of rich, normatively salient differences within and
among subjects in a way that precludes psychic wholeness and normative sovereignty as
elements of a credible erotic ideal.

3.1: productive power: genealogies of subject formation

Nietzsche on “the bad conscience”

In *Genealogy* II, Nietzsche maintains that, prior to the advent of MP as an ethical
and institutional paradigm, human beings were relatively shallow, superficial creatures
not yet capable of taking responsibility for their conduct in the way SI requires. As we
have already seen, in order to streamline his exposition of the normative powers
constitutive of Moral Personhood, Nietzsche takes promise-making for his paradigm.
With the rise of MP, human beings assumed the psychic depth and complexity required
for a reflective awareness of their own perceptions and desires as potential grounds for
belief and action for which they are first-personally responsible. In *Genealogy* II,
Nietzsche suggests that the formation of the ‘moral person’ as “the animal with the right
to make promises” is not a metaphysical fait accompli shaped by a formal requirement
but is instead the result of the repeated application of disciplinary techniques imposed
from material sources both within and without the psyche.

To have standing to “make promises,” or future-oriented commitments, one must
be capable of binding oneself to an as-yet-unrealized end of action, and this requires a
powerful memory. Nietzsche insists that memory is not a passive faculty but “an active
desire not to rid oneself [of an impression], a desire for the continuance of something
desired once, a real memory of the will:
To ordain the future in advance in this way, man must first have learned to distinguish necessary events from chance ones, to think causally, to see and anticipate distant eventualities as if they belonged to the present, to decide with certainty what is the goal and what the means to it, and in general be able to calculate and compute. Man himself must first of all have become calculable, regular, necessary, even in his own image of himself, if he is to be able to stand security for his own future, which is what one who promises does.\footnote{GM II: 1}

It is obvious that sustaining a commitment already made requires resisting the pull of countervailing inclinations tempting us to abandon or compromise the ends we have set. Frankfurt and Korsgaard make much of this fact. But here Nietzsche is reminding us that the ability to set ends of any kind presupposes a basic set of capacities for having sustained, future-oriented desires and a hierarchical ordering of one’s drives. One must be capable of projecting an image of oneself and one’s interests backwards and forwards in time and of sustaining these projections over time against contrary pressures.

Nietzsche insists that breeding an animal psychologically capable of uniform and predictable conduct—an animal with a “real memory of the will”—is extremely difficult because there is a powerful “opposing force, that of forgetfulness.” Like memory, forgetting for Nietzsche is not a passive but “an active and in the strictest sense positive faculty of repression…which is like a doorkeeper, a preserver of psychic order, repose, and etiquette: so that it will be immediately obvious how there could be no happiness, no cheerfulness, no hope, no pride, no \textit{present}, without forgetfulness.”\footnote{Ibid.} So creating a memory requires countering the “active forgetfulness” that constitutes “a form of \textit{robust} health” for the human animal. And this requires a tremendous effort. What effort could suffice?

\footnote{GM II: 1} \footnote{Ibid.}
Nietzsche identifies punishment in all its myriad ritual forms as the mechanism by which forgetfulness is initially overcome and the psychic foundations of memory laid:

One can well believe that the answers and methods for solving this primeval problem were not precisely gentle; perhaps indeed there was nothing more fearful and uncanny in the whole prehistory of man than his *mnemotechnics*. ‘If something is to stay in the memory it must be burned in; only that which never ceases to hurt stays in the memory’—this is a main clause of the oldest (unhappily also the most enduring) psychology on earth.¹⁵⁰

Nietzsche goes on to identify the psychic capacities achieved at length through punitive measures in terms that recall the horizontal vocabulary of “taking seriously” we introduced in Chapter I:

One might even say that wherever on earth solemnity, seriousness, mystery, and gloomy coloring still distinguish the life of man and a people, something of the terror that formerly attended all promises, pledges, and vows on earth is *still effective*; the past, the longest, deepest and sternest past, breathes upon us and rises up in us whenever we become ‘serious.’¹⁵¹

Finally, anticipating Frankfurt’s discussion of the importance for commitment of certain courses of action being subjectively “unthinkable,” Nietzsche anticipates his own later discussion of the ascetic ideal, the whole of which belongs here: a few ideas are to be rendered inextinguishable, ever-present, unforgettable, ‘fixed,’ with the aim of hypnotizing the entire nervous and intellectual system with these ‘fixed ideas’—and ascetic procedures and modes of life and means of freeing these ideas from competition of all other ideas, so as to make them ‘unforgettable’.¹⁵²

So through being passively subjected to punitive measures, the human animal becomes a subject in the active sense implying the capacity for responsibility. It is not difficult to see how behavioral conditioning may result from forcible subjection to punishment. Nietzsche reminds us that repeatedly subjecting individuals to the painful and humiliating ordeal of punishment reinforces a tendency to observe behavioral

¹⁵⁰ Nietzsche, GM II: 3
¹⁵¹ Ibid.
¹⁵² Ibid.
requirements both in those being punished and in those who witness the spectacle. But punitive rituals are also supposed to structurally enrich the psyches of those affected by extending their faculties of desire in the ways just discussed. What is at stake here is the internal hierarchy and differentiation that Frankfurt and Korsgaard refer to as “fragmentation”—the awareness of a perception or drive as an item in one’s own mental life, that is, as a candidate for reflexive identification or rejection. What we want to account for is “conscience,” and we may wonder exactly how it is supposed to result from subjection to punishment.

Nietzsche’s answer is that conscience is brought about by the inherently repressive, involuntary socialization of the human animal. He is referring to a series of incremental changes in the human psyche that unfold in real bodies over time through punitive rituals and other “fearful bulwarks with which the political organization protected itself against the old instincts of freedom.” Two closely related aspects of Nietzsche’s account of this process are of special importance to us. First, its psychic effects are to be explained largely in terms of a key principle of Nietzsche’s psychology, namely that “All instincts that do not discharge themselves outwardly turn inward—this is what I call the internalization of man: thus it was that man first developed what was later called his ‘soul.’” Second, this turning of animal instincts against the animal whose instincts they are produces “the bad conscience,” or the consciousness of guilt.

---

153 Nietzsche, GM II: 16. We have seen that Nietzsche casts the developmental process in historical terms, but his claims are not meant as a purely or even primarily empirical-historical hypothesis. For instance, Judith Butler, whose version of the same sort of view we encounter in the next section, indexes the relevant transformations to the developmental history of discrete individuals, at least under late modern conditions. The reasons for these partly rhetorical differences are subtle and would take us too far afield. Let it suffice to say that Nietzsche and Butler offer two versions of what is essentially the same account of the relevant psychic phenomena. Both are “genealogical” in the sense that the relevant developments are seen as socially and historically contingent but real in their normative and psychological effects.

154 Ibid., Sec. 16, p. 84
Guilt so construed is the subjective experience of one’s own instinctive drives both as one’s own and as illicit. The varied means of compulsory socialization “brought about that all those instincts of wild, free, prowling man turned backward against man himself. Hostility, cruelty, joy in persecuting, in attacking, in change, in destruction—all this turned against the possessors of such instincts: that is the origin of the ‘bad conscience.’”¹⁵⁵

Two points about this account of the origins of conscience should be emphasized.

I. Disciplinary Power Presupposes and Produces Difference

Nietzsche’s account entails that the normative and psychological capacities belonging to human beings, and so the ideals achievable through their exercise, grow out of developmentally prior conditions quite at odds with them. For consider that conscience grows out of the bad conscience, which in turns grows out of a condition of forced confinement “within the walls of society and of peace.”¹⁵⁶ We saw that conscience is Nietzsche’s name for feeling oneself to be ultimately accountable for what one does as one who could have done otherwise. He claims that this feeling can only emerge in a creature first endowed with the capacity for distancing itself from and questioning its own instinctive drives. Since this mistrusting of one’s own drives is the bad conscience, it follows that it conditions reflexive self-consciousness, or the engine of rational agency as understood by SI.

Crucially, Nietzsche calls the bad conscience a “serious illness,” since the capacity for negating one’s own drives is contrary to the human animal’s primitive

¹⁵⁵ Nietzsche, GM II: 16
¹⁵⁶ Ibid.
reliance upon the socially unmediated guidance of its instincts. So the sovereign individual with his consciousness of his own “freedom” and “power” cannot have emerged but through developments which produce and entrench “man’s feeling of shame at man,”

“man’s suffering of man, of himself;” the tendency to find the instincts and drives of our animal natures, and so ourselves, to be repellant and unworthy. All of this despite the fact that Nietzsche regards the “fruit” of the bad conscience, namely the possibility represented by the sovereign individual, to be extremely propitious:

Let us add at once that, on the other hand, the existence on earth of an animal soul turned against itself, taking sides against itself, was something so new, profound, unheard of, enigmatic, contradictory, and pregnant with a future that the aspect of the earth was essentially altered…From now on, man is included among the most unexpected and exciting lucky throws in the dice game of Heraclitus’ ‘great child,’ be he called Zeus or chance; he gives rise to an interest, a tension, a hope, almost a certainty, as if with him something were announcing and preparing itself, as if man were not a goal but only a way, an episode, a bridge, a great promise.—

In short, a mode of subjectivity that we may well find attractive, the sovereign individual, depends for its very possibility on a prior, deeply problematic mode of subjectivity, the bad conscience, which in turn results from the application of disciplinary techniques essentially involving domination. “Conscience,” as the possibility of an ideal of freedom and responsibility, therefore rests upon a sustained institutional and ritual practice of violent subordination: “the welding of a hitherto unchecked and shapeless populace into a firm form was not only instituted by an act of violence but also carried to

---

157 Nietzsche, GM II: 7
158 Ibid., Sec. 16. Nietzsche here reminds us that the advent of the bad conscience and all it portends was accidental, a “lucky throw” in a game of chance played out from human pre-history down to the modern age. He thus hints that further developments, conditions of life beyond the “sickness” of “guilt” or shame at our instinctual drives, are possible, and concludes by assigning a high value to such “a hope.” When we attempt to defend a broadly Nietzschean erotic ideal of “serious play” in our concluding chapter, we will be offering an interpretation of the possibility Nietzsche here identifies as the “great promise” of the bad conscience. Indeed, in the above passage, Nietzsche identifies the spirit of that ideal in terms of recognizing the value of “an animal soul turned against itself, taking sides against itself.”
its conclusion by nothing but acts of violence.”159 If ‘power’ is our term for the effects and relations that enter into the formation of subjects, Nietzsche’s account implies that power is essentially productive of normative and psychological differences, differences which grow out of and into other differences.

So Nietzsche’s account of the origins of conscience suggests a general lesson about the formation of active subjects: any set of normative and psychological capacities belonging to human beings at a particular time grow out of material conditions different from and even hostile to those capacities. Subjects are formed through a kind of violent interplay among differences, and the interplay produces further differences. For our purposes, the most important implication of this view of power as presupposing and producing difference is that wholeness or unity in the psyches of subjects shaped by power is not a viable ideal because it has no paradigmatic connection to the conditions of their constitution. This amounts to a denial of the first core tenet of SI that we identified.

Recall Korsgaard’s rhetorical insistence that rational agents constitute their identities through action, such that who we are is posterior to, and a normative function of, what we do. Her account of why this is and what it means entails that there is a necessarily prior and so an inescapably authoritative standard for effective self-constitution—the moral law, a procedural norm that embodies the “constitutive principles

159 Nietzsche insists that the earliest communities in which the bad conscience arose were “fearful tyrannies” fashioned by conquering groups from populations culled from nomadic tribes with relatively fluid forms of social organization and rule (GM II:17). Contrast Nietzsche’s genealogy of “the sovereign individual” with Korsgaard’s defense of SI, which rests on the claim that, at bottom and in the most philosophically important respect, responsible agency is constituted by a priori formal requirements of reason. The poverty of the latter as an explanatory route to what human agents as we find them are concretely like and why is effectively acknowledged by Korsgaard herself when she claims to offer no plausible story of how “evil” (or disunity in the rational psyche) comes about: “If the categorical imperative is the natural law of a free will, why should a free will, noumenal and pure, that is, uninfluenced by any law of [sic] outside of itself, act on any law other than its own? Why should it allow itself to be governed by inclination? And if justice is the very form of a unified soul, why are there souls that are not just?...I am not trying to explain Plato’s views on how bad action comes about, or to say anything about that question myself.” (Korsgaard 2009, p. 164).
of volition,” namely that rational agents must be autonomous and efficacious. So though particular human identities are not metaphysically fixed in advance of relevant behavior, there is a “universal identity” of rational selfhood to which all individual selves are subject because it defines the universal, necessary parameters for properly constructing any particular, contingent identity. The subjectivity it prescribes therefore has an immutable, albeit purely formal, structural character. In short, SI offers a paradigm of universal subjectivity, an image of a fully formed, or wholly realized, agent. By its very nature, it is a fixed and final template. According to SI, any agent anywhere that is not subject to its requirements cannot act in the strict sense implying self-determination. There is really only one model of responsible action, “the sovereign individual,” and it is timeless because its requirements are constitutive of action itself.

Nietzsche’s genealogy of the sovereign individual and his “conscience” is meant to show why it is not the only game in town. There are other possible forms of human subjectivity. We can see this because the sovereign individual both grows out of them and, as Nietzsche’s remarks about “the great promise” suggest, points beyond itself.160 His account is meant to remind us that human beings as we find them are not shaped by an a priori formal principle but by other human beings through the repeated application

160 It might be thought that, when Nietzsche identifies the advent of the bad conscience with the breeding of an animal endowed “with the right to make promises,” and characterizes the latter in terms of the origin of responsibility, he is effectively conceding to SI that there is really only one model of responsible agency, namely the sovereign individual which grows out of the bad conscience. But this would be a mistake. As we have already indicated, in our concluding chapter we will attempt to defend an erotic ideal—a view of how we should live out our commitments seriously and thereby exercise the sort of responsibility required by integrity—that breaks sharply with the horizontal strategy introduced in Chapter I and examined in a robust form as SI in Chapter II. The erotic ideal in question, one of “serious play,” will be an interpretation of what Nietzsche calls the “great promise” of what lies beyond the bad conscience. So it is an account of responsible agency, but meant as an alternative to SI. This is in keeping with the upshot of Chapter I, where we supposed that there is a post-Nietzschean route to a worthy erotic ideal.
of material techniques like disciplinary “mnemotechnics” that violently brand and ultimately transform real human bodies over time.

Now in Chapter II we saw that SI rests on a vocabulary of transcendence: “reason’s own principle” is nowhere bodily instantiated but deeply conditions human subjectivity wherever it is found. By Nietzsche’s lights, this is psychological nonsense. It is a mistake stemming from historical naivety and reinforced by a quasi-religious philosophical mythology. On the contrary, the “breeding” of human beings with any particular set of capacities is an irreducibly this-worldly process. Proponents of SI have mistaken an ideal they inherited from a tradition, the sovereign individual, for what that tradition tries to pass off as an otherworldly artifact. Given that the tradition that disciplines the sovereign individual is an essentially ascetic one, as Nietzsche explains in the Third Essay of his *Genealogy*, it is little wonder that proponents of SI claim to disclose its basic conditions through the exercise of “pure” reason.

These observations bring us to the second important point about Nietzsche’s account of the origins of conscience.

II. Disciplinary Power Is Contingent

The developments which lead to the bad conscience and so to the sovereign individual’s conscience are deeply contingent. This is a point we have already made, but we can now further develop it as a claim about the contingency of the conditions that shape the powers of human agents. This will further distance our account of the constitution of the active self from the SI’s, which sees the formative occasion for agency as rationally necessary and universal in scope. For Nietzsche, the relevant conditions
express no rational necessity or inexorable teleology. Any claims about “developmental priority” among them are meant as purely hypothetical judgments about the bodily features of human beings as we find them in our experience. Human capacities are contingent in the sense that they are produced by regulatory practices and disciplinary rituals whose purposes are themselves fluid. What a particular practice means to those affected, and the impact it has on the development or exercise of their capacities, is subject to far-reaching, unforeseen transformation. This suggests we are on the right track since we initially proposed to understand power as the domain of effects and relations driving contestation over disputed meanings in a community.

So power produces contingent differences because the conditions of its productivity are fluid. We should try to clarify this important point, since it already points to the second of power’s basic aspects, its “excessive” character, to which we will turn in the next main section. During his discussion of the bad conscience, Nietzsche explains that the whole constellation of ideas surrounding and supporting this phenomenon emerged from a prior institutional context in which those terms had no place. He observes that “It was in…the sphere of legal obligations, that the moral conceptual world of ‘guilt,’ ‘conscience,’ ‘duty,’ ‘sacredness of duty,’ had its origin,” reminding us again that “its beginnings were, like the beginnings of everything great on earth, soaked in blood thoroughly and for a long time.”\(^{161}\) He continues: “The feeling of guilt, of personal obligation, had its origin…in the oldest and most primitive personal relationship, that between buyer and seller, creditor and debtor: it was here that…one person first measured himself against another.”\(^{162}\)

\(^{161}\) Nietzsche, GM II: 6
\(^{162}\) Ibid., Sec. 8
Nietzsche thus traces the earliest origins of the consciousness of guilt (the German Schuld) to “the very material concept Schulden [debts].” He suggests that the earliest punitive rituals, “as requital, evolved quite independently of any presupposition concerning freedom or nonfreedom of the will,” but afforded creditors compensation for their material loss in the form of the “counterbalancing” pleasure obtained by venting their violent animal drives on errant debtors, whether directly or vicariously through appointed third-parties: “To what extent can suffering balance debts or guilt? To the extent that to make suffer was in the highest degree pleasurable.”\textsuperscript{163} The notion that the errant debtor was an irreducibly responsible party bearing an internal burden of guilt corresponding to her material debt came much later, in Nietzsche’s view. Initially there was simply the primitive assumption that “every injury has its equivalent and can actually be paid back, even if only through the pain of the culprit.”\textsuperscript{164}

This line of thought leads to Nietzsche’s colorful descriptions of the public punitive rituals of primitive humanity in which disciplinarians and spectators performed and beheld execution, torture and other violent spectacles of cruelty in a joyful, ecstatic atmosphere:

\textsuperscript{163} Nietzsche, GM II: 6
\textsuperscript{164} Ibid., Sec. 4. Nietzsche describes the transition from early communities, in which punishment against lawbreakers was severe and unmediated by judicial impartiality, to more secure and populous societies whose punitive measures are more lenient and impartially administered. He suggests a general rule: the less threatened a community is by the deviance of its lawbreakers, the less severe its punitive measures are likely to be. He regards the practice of “mercy,” or the remission of legal offenses, as an important index of the strength of a community: “The justice which began with, ‘everything is dischargeable, everything must be discharged,’ ends by winking and letting those incapable of discharging their debt go free: it ends, as does every good thing on earth, by overcoming itself. This self-overcoming of justice: one knows the beautiful name it has given itself—mercy.” (GII 10, pp. 72-72). Nietzsche held parallel views about the capacity for sympathy and even forgiveness in individuals. He hints at this at the end of this passage: “…it goes without saying that mercy remains the privilege of the most powerful man, or better, his—beyond the law.” Finally, Nietzsche suggests that with a greater social resiliency and punitive leniency in the face of lawbreaking, there arose a practice of institutionally distinguishing the criminal and her crime. This is a further stage of the Second Essay’s genealogy of “the sovereign individual,” albeit an easily overlooked one since Nietzsche remarks on it only in passing.
To see others suffer does one good, to make others suffer even more: this is a hard saying but an ancient, mighty, human, all-too-human principle...Without cruelty there is no festival: thus the longest and most ancient part of human history teaches—and in punishment there is so much that is festive!—

These are the lowly origins of the bad conscience, the consciousness of guilt as a first-personal burden of blameworthiness, and of conscience as the experience of oneself as a free and responsible agent. Punitive rituals which eventually came to be understood in terms of holding guilty parties responsible were before devised for the cruder purposes just described. Accordingly, the primitive creatures whose ritual practices sowed the seed of conscience were quite unlike the psychologically robust individuals who eventually bore the fruit. The general lesson here is that changes in what is possible for human subjects track changes in how they are shaped by the involuntary rituals of socialization to which they are subject.

Nietzsche introduces his notion of “the will to power” to explain how this can be. He begins with a distinction between the origin and the purpose or meaning of any technique, practice or instrument of use:

…The cause of the origin of a thing and its eventual utility, its actual employment in a system of purposes, lie worlds apart; whatever exists, having somehow come into being, is again and again reinterpreted to new ends, taken over, transformed, and redirected by some power superior to it…purposes and utilities are only signs that a will to power has become master of something less powerful and imposed upon it the character of a function; and the entire history of a “thing,” an organ, a custom can in this way be a continuous sign-chain of ever new interpretations and adaptations whose causes do not even have to be related to one another but, on the contrary, in some cases succeed and alternate with one another in a purely chance fashion.

For Nietzsche, power is thus the fundamental principle of change and transformation in living systems. As such, it is responsible for the dynamic life of any set of capacities to which one can assign an interpretation in terms of an immanent meaning or purpose.

---

165 GM II : 6
166 Ibid., Sec. 12
draws our attention to the distinction between what is “relatively enduring, the custom, the act...a certain strict sequence of procedures,” and “that in it which is fluid, the meaning, the purpose, the expectation associated with the performance of such procedures.” ¹⁶⁷ Both aspects are fluid, but “the ‘meaning’ is even more so.” ¹⁶⁸

Moreover, Nietzsche insists that the historical sedimentation of the changing “meanings” of ritual practices make the resulting phenomena “a kind of unity that is hard to disentangle, hard to analyze, and, as must be emphasized especially, totally indefinable.” Recall SI’s insistence that the possibility of rational agency can only be explicated in terms of a fixed set of formal principles. Nietzsche’s terse rejoinder distills the guiding insight of his genealogical method: “only that which has no history is definable.” ¹⁶⁹ In short, by overlooking the role of power in constituting any set of ritual practices and bodily capacities, “one overlooks the essential priority of the spontaneous, aggressive, expansive, form-giving forces that give new interpretations and directions” to the practices or capacities in question. ¹⁷⁰ The fluid contingency of power’s productive conditions makes any form of human subjectivity an essentially precarious achievement.

We proposed to understand power in terms of the effects and relations that enter into the formation of human capacities for action. In this section, we reviewed its productive character and saw why it precludes the psychic wholeness or unity that SI represents as the sole structural paradigm for responsible agents. By doing so, we have already begun to articulate power’s second primary feature, its inherent tendency to

¹⁶⁷ GM II: 13
¹⁶⁸ Ibid., Sec. 12. As Nietzsche puts it in the Genealogy’s Third Essay: “All good things were formerly bad things; every original sin has turned into an original virtue.” (GM III: 9).
¹⁶⁹ GM II: 13
¹⁷⁰ Foucault’s work on disciplinary society and sexuality follows Nietzsche’s by bringing the body into focus as the site for the application and exercise of social technologies of power.
‘exceed’ discrete acts. We will see shortly what this means and why it precludes SI’s second core tenet, the ideal ascription of sovereign control to responsible agents. But first we must complete our review of what it means to say that power presupposes and produces difference such that conditions of alterity, not unity, fundamentally shape subjects. To do so we turn to Judith Butler’s psychoanalytic and genealogical account of subject formation through relations of subordination and attachment.

Butler on subjection & subordination

We have seen that Nietzsche’s account of the “bad conscience” constitutes a psychosocial history of the formation of “the sovereign individual.” We care about Nietzsche’s account because the sovereign self invites comparison with the self-governing moral subject championed by Sovereign Idealism. The sovereign self is so constituted as to regard itself as radically detachable from its occurrent desires and perceptions and therefore capable of reflectively self-determined conduct. For Nietzsche, the structuring of the human psyche which makes this possible is the result of a process of disciplinary conditioning aimed at making individuals agents of the social reproduction of normative standards. This is accomplished through the internalization of social expectations in the form of “conscience,” or the voice of socially instituted authority now found within the subject’s psyche.

By Nietzsche’s lights, this process is not to be understood as the colonization of an already existing internal domain of consciousness by an external agency, but as the production of a set of bodily capacities that would not otherwise exist. Agents are subjects in that they come and continue to exist insofar as they are passively subject to
productive disciplinary power. To be capable of incorporating normative demands, they
must themselves be inaugurated through involuntary subjection to pressures towards
behavioral conformity that are manifested through rituals of socialization directed at the
body. Subjection to these pressures turns drives against themselves, anchoring power in
the body through conditioned mechanisms of self-monitoring that are themselves subject
to disciplinary oversight.

But this cannot be the whole story for our purposes. For what explains the sense
in which subjects are active and first-personally responsible for the norms they
incorporate? What grounds an appeal to erotic ideals whereby human beings take a stand
on their commitments to the norms they incorporate? Recall that we ultimately want an
account of the possibility and value of taking our commitments seriously in light of their
susceptibility to loss and transformation, which Nietzsche would describe as their “fluid
contingency.” To do this, we need an account of the power exercised by subjects, not just
an account of the power exercised through them.

In The Psychic Life of Power, Judith Butler addresses these questions by
developing an account of the formation of responsible human subjects that extends
Nietzsche’s genealogy of conscience. What she wants to explain is how subjects formed
through power come to have powers of their own, capacities for whose exercise they are
responsible. She describes the formation of subjects as a socially conditioned process
whereby relations of domination and subordination play an integral role in the production
of the psyche. “Power” is Butler’s shorthand term for the field of psychic and bodily
effects and relations through which subjects are constituted and sustained.
Butler’s account recalls Nietzsche’s genealogy of conscience by highlighting power’s critical role in “subjection” understood both in terms of subordination—“a setting under” a prior agency—and the process through which subjects are brought into existence. Indeed there is a sense in which Butler offers an account of the conditions for the possibility of responsible agency. But Butler’s position differs significantly from Korsgaard’s. According to Butler, thematizing social power, not formulating rational principles, is the key to understanding the constitution of subjects. Since subjects are radically conditioned by power, power provides the formative occasion for the exercise of agency. For according to Butler, agency presupposes a subject that acts, though she hastens to add that ‘subject’ so construed “ought to be designated as a linguistic category, a placeholder, a structure in formation” only.\(^{171}\)

For Korsgaard, the supreme principle of practical reason, the categorical imperative, is the ultimate source of rational agency and so the ultimate reference point for thematizing it. Since the categorical imperative is supposed to entail a substantive commitment to moral constraints, morality alone is said to fundamentally condition the nature and possibility of agency. Whereas on Butler’s view, as on Nietzsche’s, anything can shape a subject’s capacity to act, including violence.\(^{172}\) There is a Freudian dimension to Butler’s account whereby unchosen attachments deriving from early, intimate personal dependencies—forms of “subordination”—embed patterns of evaluation and affect in the psyche of the dependent subject.

Moreover, Butler argues that to the extent that subjects are distinguished by the exercise of their normative powers, their capacities to act draw upon the power by which

\(^{171}\) Butler1997(2), p. 10
\(^{172}\) Ibid., p. 25 et passim
they are formed since those capacities are produced through prior operations of power. That is, a subject’s own power is conditioned by an “external” power already embodied in the subject that it has formed. However, Butler insists that there is no unequivocal way of disentangling these two moments in the operation of power. She represents the distinction as incorrigibly “ambivalent.”

It is difficult to see what this is supposed to mean. As Butler puts it, subjection is marked by “an irresolvable ambiguity [that] arises when one attempts to distinguish between the power that (transitively) enacts the subject, and the power enacted by the subject, that is, between the power that forms the subject and the subject’s ‘own’ power:”

What or who is doing the ‘enacting’ here? Is it a power prior to the subject or that of the subject itself? At some point, a reversal and concealment occurs, and power emerges as what belongs exclusively to the subject (making the subject appear as if it belonged to no prior operation of power). Moreover, what is enacted by the subject is enabled but not finally constrained by the prior working of power. Agency exceeds the power by which it is enabled. One might say that the purposes of power are not always the purposes of agency. To the extent that the latter diverge from the former, agency is the assumption of a purpose unintended by power, one that could not have been derived logically or historically, that operates in a relation of contingency and reversal to the power that makes it possible, to which it nevertheless belongs. This is, as it were, the ambivalent scene of agency, constrained by no teleological necessity.¹⁷³

Butler’s reference to “the subject’s ‘own’ power” recalls Korsgaard’s contrasting discussion of the necessary first-personal authority, or “mineness,” of agents’ actions. But Butler’s analysis complicates matters by following Nietzsche’s discussion of the fluid meaning of an act or process that disciplines subjects. As that which both shapes and is exercised by subjects, power is “ambivalent” in that it has an irreducibly dual aspect. It “appears in two incommensurable temporal modalities:” what precedes and conditions the subject’s capacity to act, and the “willed effect” of the subject’s capacity to act:

¹⁷³ Butler 1997 (2), p. 15
...the subject is itself a site of this ambivalence in which the subject emerges both as the effect of a prior power and as the condition of possibility for a radically conditioned form of agency. A theory of the subject should take into account the full ambivalence of the conditions of its operation.\footnote{Butler 1997 (2), p. 15}

Power is thus both prior and posterior to the subject as the first-personal site of agency. The subject is passive (subject-to) in relation to power that “acts on” or is external and prior to it; yet it is active (subject-for) in relation to power that is “acted by” or constitutive and posterior to it. Both aspects are integral to the formation of subjects and to their functioning as agents, yet SI only recognizes one aspect by representing subjects as strictly active \textit{qua} agents. Even SI’s treatment of the subject as potentially subject \textit{to} norms represents it as irreducibly and unequivocally active. This is problematic for reasons we can begin to articulate now that we understand power as equivocally productive.

This feature of SI is reflected in an account of normative discourse sketched by Korsgaard in order to illustrate the implications of her view of integrity for social and political theory. Unsurprisingly, her view of “interaction” between agents is a Kantian one. To be subject to a norm is to be addressed or held accountable by others in normative discourse. The sort of discourse taken to be relevant is one in which normative practical questions arise and in which participants hold themselves and one another responsible for the stands they take. Further, such discourse requires its participants to address one another as self-unifying sovereigns in keeping with SI’s core requirements. We will return to SI’s theory of discourse in the following section on the “excessiveness” of power. For now we simply notice that SI’s view of practices of holding responsible rests on its view of participants as necessarily sovereign, self-unifying agents, and that

\footnote{Butler 1997 (2), p. 15}
this flouts the view of power and subjectivity we have been reviewing. In the next section, we will explain why SI’s model of “sovereign discourse” distorts our understanding of practices of holding responsible. Further exposing SI’s deficiencies in this way will bring us one step closer to articulating the agonistic erotic ideal we are after.

For now let us conclude our review of Butler’s characterization of power’s irreducibly dual aspect and the highly qualified form of agency it conditions. We are indeed defined by our roles, and by our identities insofar as these express values to which we are attached—Korsgaard is right about this much—but we outstrip them and the role played by power in psychic life explains why. Butler maintains that there is an “irresolvable ambiguity” at the heart of the relation “between the power that (transitively) enacts the subject, and the power enacted by the subject…the subject’s ‘own’ power.”\textsuperscript{175} In other words, power is equally the condition and the effect of active subjects. The subject cannot be reduced to the power which occasions it, nor can the power which forms the subject be reduced to the subject who wields it as an agent. The prior, formative power which inaugurates the subject and agency understood as the subject’s exercise of its “own” power are mutually enabling yet mutually transcending conditions. Each dialectical element depends on but outstrips the other. The subject has a power of its own, but it is never simply “mine,” contrary to SI’s insistence.

So what ultimately distinguishes the subject from the power through which it is formed and sustained? Butler’s answer is that, \textit{ultimately}, nothing does. The subject is a “dissimulation” of power, the effect of a “reversal and concealment” of its prior, formative operation, but, owing to the “irresolvable ambiguity” we have just identified, the effect is precisely \textit{not a fait accompli}. This is to say that the boundary between the

\textsuperscript{175} Butler 1997 (2), p. 15
formative working of power and the subject so formed is ambiguous, marked by excess, by what Nietzsche would call a surfeit or ecstatic “overflowing” of “the will to power” as the form-giving, self-overcoming basic drive or fund of proliferating differences ontologically responsible for this dynamic cycle of being and becoming. To return to Butler’s Foucaultian idiom, the subject is not the self-enclosed center of control and authority of Korsgaard’s psychological imaginary, but an active point of intersection in social space, an unstable crossroads of multiple circuits of psychic and bodily power:

If conditions of power are to persist, they must be reiterated; the subject is precisely the site of such reiteration, a repetition that is never merely mechanical. As the appearance of power shifts from the condition of the subject to its effects, the conditions of power (prior and external) assume a present and futural form. But power assumes this present character through a reversal of its direction, one that performs a break with what has come before and dissimulates as a self-inaugurating agency. The reiteration of power not only temporalizes the conditions of subordination but shows these conditions to be, not static structures, but temporalized—active and productive.176

As a self-dissimulating formation of power, an opaque, reflexive turning of power onto itself, Butler’s subject is thus a self-overcoming of social and psychic power. The interior domain of reflective distance so privileged and invested with sovereignty and paradigmatic wholeness within SI is, for Butler, really a site of contention given depth and made active through a kind of psychic and social violence, that of “primary” subordination, dependency and concealment.

3.2: equivocal acts, excessive agency: Butler on “excitable speech”

We concluded Chapter II’s discussion of Sovereign Idealism by comparing its representation of human agency with the idea of sovereign power central to Judith Butler’s work on linguistic performativity and hate speech regulation. We saw that SI

176 Butler 1997 (2), p. 16
asserts the paradigmatic sovereignty of human agency. In Korsgaard’s words, “…as long as you’re in charge, so long as nothing happens to derail you, you must act. You have no choice but to choose, and to act on your choice.” We then disambiguated SI’s ascription of sovereignty to human agents, finding that it represents us as sovereign under two distinct but related aspects.

First, where our ‘outward’ or causally and interpersonally implicated conduct is concerned, we are merely subject to an ideal of (unequivocal and immediate, that is, sovereign) efficacy; “merely” subject means there is no reason to expect that we will ever be capable of fully realizing the ideal in concrete practice. The reason for this is that our efforts to realize the aims we set are plainly subject to causal interference from the world and practical interference from other agents whose aims or conduct may frustrate ours in lots of ways. Call this the weak but wide ascription. Second, there is an internal domain of conduct in which we finite, concrete human agents are always already fully realized, unequivocally authoritative sovereigns. For according to SI, when what are at issue are the ends we set as such and our capacity to initiate the outward effort to realize them, as agents we are in charge, period. That is, our will understood as a “self-determining causality” is necessarily sufficient to set the ends we elect to set in the fashion we envision, and so too with respect to at least the initiating exercise of volition aimed at realizing our ends through overt, or worldly and interpersonally implicated, conduct. Call this the strong but narrow ascription.

Notice that SI is committed to both ascriptions. It cannot withhold the strong but narrow ascription of fully realized sovereignty to concrete human subjects, for unless we were already at least nascent sovereigns, we could not now recognize ourselves as

---

177 Korsgaard 2009, p. 1
unavoidably bound by the ideal of “wholly” free and effective action, of being “perfectly rational” beings, as Kant put it. The ideal of wholly realized sovereign authority in the causal world and among other agents could only be constitutive for agents with one foot already in that world, each of whom is always inherently oriented in that direction by virtue of already participating in what Korsgaard calls “the pure form.” The necessary point of contact between the ideal world of uniformly, fully realized rational agents, the Kingdom of Ends, and our own world where interpersonal difference, disruption and dislocation abound, is the internal domain of each individual’s will. The existence of a private volitional domain is presupposed by the two sub-theses of the Sufficiency of the Will Thesis which, as we saw in Chapter II, supplies one of three main elements of SI’s ascription of sovereignty. SI insists that, as self-determining causality, I cannot fail either to make some end my own if I so choose (SWT-a), or to commit myself to the realization of an end through my action if I so will it (SWT-b). This is, as we have said, the privileged, indeed robustly private, sphere within which subjects are ultimately “in charge” and so genuine agents even in the world as we find it where freedom and effective control is always circumscribed. As Korsgaard is fond of saying, “There’s a touch of the divine in being an agent,” and concrete human beings are agents on her view. We are simply as-yet-imperfect ones.

However, we postponed discussion of whether this dual ascription of sovereignty weakens SI. Now we may address that question, since a major aim of this chapter is to show why both ascriptions of sovereignty, the weak but wide and the strong but narrow, are psychologically untenable. In this section, we further explicate our working
conception of power by returning to Butler’s treatment of “the sovereign conceit” in hate speech discourse.

Recall that our general aim is to defend a conception of agency that precludes, rather than requires, the human exercise of sovereign power, whether as a realized capacity or as a paradigm. So we started by articulating it negatively: it precludes sovereignty. But the resources for such a conception we find in Butler’s work on linguistic performativity have a constructive implication: that agents so constituted as to be incapable of sovereign power may develop other sorts of normatively salient capacities which ascriptions of sovereignty preclude. The potential ethical and political value associated with the exercise of these non-sovereign capacities is an issue we take up in the following chapters. For now we want to further the case already made for an alternative to SI’s conception of agency by focusing on the issue of sovereign power.

We will do so in two steps. First, we will return to Butler’s discussion of linguistic agency to recall why the pervasively “excitable” character of speech precludes sovereignty even as a paradigm for human linguistic subjects. Second, we will use Butler’s account to expose a confusion in Korsgaard’s theory of reasoned discourse which we briefly addressed in the previous section. Now, as we saw in Chapter II, Korsgaard tries to vindicate SI’s basic conceptual vocabulary by establishing a phenomenology of practical deliberation that requires it. But her strategy plays into our hands. For our criticism of Korsgaard’s account of discourse will help explain why her phenomenological account of our experience as practical deliberators is deficient, too. But first we should clear the way by addressing a possible objection to our characterization of SI in terms of SWT (a-b).
There seems an obvious objection to ascribing SWT (a-b) to SI. Surely an individual’s will is not sufficient to establish any end of action, and surely charity requires that we take Korsgaard to recognize this fact. After all, it seems obvious that some ends of action, like being a samurai, are simply unavailable because the ritual and institutional support they require no longer exists. The samurai way of life is strictly unavailable today because the feudal Japanese society that produced and enforced bushido, or the samurai code of conduct, collapsed long ago. This observation echoes Nietzsche’s historicist account of subject formation and agency. Can Korsgaard accommodate this much of the account of power examined so far?

She might try by insisting that there is a middle position between the robust sovereignty ascription we have been considering and one so weak it no longer entails an ascription of sovereignty. She may thus hold that any concrete individual’s end-setting powers are constrained by social and historical factors, but that no normative demand can bind her without her reflective endorsement of it. In Chapter II, we saw that Korsgaard is indeed committed to the latter claim; we called it the Necessity of the Will Thesis (NWT). The idea is that agents cannot be bound by a requirement unless they lay it down for themselves as a law. One’s whims, impulses, and so on are not manifestations of who we really are unless we make them such through an irreducibly first-personal exercise of our will.178

But a little reflection shows that this is really no “middle position” at all, but reintroduces the robust version of SWT through the back door by construing the

---

178 We also saw that Korsgaard believes NWT follows from the third component of SI’s general requirement of sovereignty, namely the claim that agents are aware and therefore in control of their perceptions, beliefs and desires as potential grounds for action. We called this the Reflective Distance Thesis (RDT).
normative valence of sovereign power as essentially reactive. On this reading, the will is a kind of guardian of the citadel of the psyche; a defensive, all-seeing monitor in the opaque hub of an internal panopticon, or a minotaur tirelessly patrolling the fragmented labyrinth of reflective consciousness. Whether a monitor or a minotaur, the will so construed remains robustly sovereign. For even when its constructive power are supposedly limited, its powers of effective surveillance and resistance are not. The will’s power to withhold endorsement from perceptions, desires and other incentives as potential grounds of belief and action must be immediately efficacious in order for the will to be irreducibly self-determining in the sense Korsgaard requires if she is not to abandon her primary aim of showing that rational beings are essentially distinguished by their capacity for authoritative, autonomous self-legislation. Since self-determination in this sense, whether primarily constructive or “defensive,” is precisely what the concept of sovereignty requires, Korsgaard cannot hope to evade the dubious implications of SWT (a-b) by modifying the valence of the will’s sovereign power.

179 Compare this image not only with the central structure of the symbolic panopticon of Foucault’s Discipline and Punish representing regulatory power, but also with Kant’s famous figure of the “opaque” will. Also compare the hub/will’s supposed opacity with Butler’s claim, discussed below, that the subject is a scene of agency created by the self-dissimulation (or, in Nietzsche’s terms, a self-overcoming) of a prior, disciplinary power.

180 I am mixing metaphors and images a little here, but the “bull-headed” figure of the minotaur is apt in light of the Nietzschean criticisms of SI’s horizontal approach to integrity we introduced in Chapter I and will return to in Chapter V. There is a provocative passage in Nietzsche’s Beyond Good and Evil that invites comparison with this way of framing what SI is committed to by suggesting an antipodal, agonistic erotic ideal ironically set out in its terms: “Independence is for the very few; it is a privilege of the strong. And whoever attempts it even with the best right but without inner constraint proves that he is probably not only strong, but also daring to the point of recklessness. He enters into a labyrinth, he multiplies a thousandfold the dangers which life brings with it in any case, not the least of which is that no one can see how and where he loses his way, becomes lonely, and is torn piecemeal by some minotaur of conscience. Supposing one like that comes to grief, this happens so far from the comprehension of men that they neither feel it nor sympathize. And he cannot go back any longer. Nor can he go back to the pity of men.——” (BGE, Part Two, Sec. 29). The association of the psychic enforcer (the minotaur) with “conscience” recalls our comparison of Nietzsche’s genealogy of the conscience of “the sovereign individual” with SI’s account of Moral Personhood.
After all, we can simply reframe the objection to keep pace with the modification. What the “samurai objection” presupposes is that one cannot set an end, let alone live out a commitment in one way rather than another, in the absence of a set of ritual and institutional practices that raise the question of what doing so properly requires. So if robust sovereign control (whether constructive or reactive) is a real possibility (let alone the supreme paradigm) for human agents, the samurai objection’s presumption requires that something be said about its genealogy: what ritual tradition of use disciplines the sovereign capacity? What essentially bodily forms of domination and violence have inaugurated this contingent possibility and account for its continuing incorporation by subjects? But SI eschews genealogy in favor of a metaphysics of normativity and an a priori rational psychology of action. The human animal’s capacity for sovereign control is supposedly vouchsafed by its inescapable subjection to a formal rational principle, not by its contingent subjection to “an originary violence” rooted in social sanctions and somatic attachments that are as essentially unchosen as they are subject to reflective appraisal. And, as we will see in a moment, a genealogy of the subject directly precludes ascriptions of sovereignty by showing that the power that shapes agency makes an ideal of mastery a chimera. For these reasons, Korsgaard cannot diffuse the worry that sovereignty in whatever form entails a “magical efficacy.”

In Chapter II, we saw that Butler’s treatment of sovereignty in Excitable Speech is focused exclusively on the sort of agency human beings exercise through speech. Nevertheless, her discussion of hate speech sheds light on the present question of what

---

181 Butler 1997 (2), p. 25
182 Ibid., p. 21. Contrast with Korsgaard’s claim, discussed in Chapter II, that “not every action is physical,” and her attempt to clarify what is involved in self-determination by appealing to what we suppose happens “when we imagine cases in which someone produces an effect magically” (Korsgaard 2009, pp. 95-96).
power is, how it functions, and its relation to human powers of action, for she thinks the
capacity for speech distinguishes human beings as such:

Could language injure us if we were not, in some sense, linguistic beings, beings who require language in order to be?...To be called a name is one of the first forms of linguistic injury that one learns. But not all name-calling is injurious. Being called a name is also one of the conditions by which a subject is constituted in language. ¹⁸³

One might suppose that this pre-aligns her view with Sovereign Idealism, which regards “rationality,” or the capacity for reflection upon and consequent control over potential grounds of belief and action, as the distinguishing feature of human agency. But this would be a mistake.

According to Butler, the human capacity for speech does introduce new possibilities of agency, but they do not include prospects for masterful control over the ultimate grounds, content and character of one’s conduct. On the contrary, the leitmotif of *Excitable Speech* is that language has an agency of its own, one that outstrips the conscious, intentional oversight of individual speakers. Speech empowers speakers; it broadens and enlarges their capacity to make a normative difference to the worlds of speech they inhabit. But when speakers speak, their speech takes on a life of its own. This means that my speech is not really mine, after all, but a riotous and elusive force over which no discrete agent has supreme authority or control. And to the extent that my capacity for speech is the measure of my “own” agency, then in the robust sense required by Sovereign Idealism, my own agency is not really mine, either.

What’s even more surprising is that if our agency in some sense produces our identities, perhaps by establishing our ends and commitments, then none of these things really belong to us either, at least in the sense of authoritative, irreducibly first-personal

¹⁸³ Butler 1997 (1), pp. 1-2
ownership required by SI. Neither our actions and practical stances nor our commitments and identities exhaust us. There is always an excessive “remainder” that opens the circuit of the self out onto an unpredictable future, persistently deferring its completion. As *agents*, subjects are not whole or in control either as we experience them or in the paradigm case. This is Butler’s core thesis about linguistic agency, and it clearly flouts SI’s first two core commitments, psychic wholeness and sovereignty. But it sounds strange. So let’s take a closer look.

The basic idea is that though speech endows human beings with distinctive forms of agency, the effects of what is said can come back to haunt speakers in ways that cannot be reliably foreseen or controlled. This is a consequence of the fact that speech is everywhere subject to disruption and failure. Butler’s “presumption is that speech,” and so the agency exercised through it, “is always in some ways out of our control.”¹⁸⁴ This is what is meant by the claim that power is *excessive*. Sovereign speech is speech which is not paradigmatically subject to deficiency, excess, disruption or failure. Butler’s claim is that no speech is like this. No speech, and so no speaker’s agency, is subject to masterful control.

We can revisit our discussion of the productivity of power to explain why. We saw that, on Butler’s view, subjects cannot be reduced to the power which forms them, nor can the power which forms subjects be reduced to the acts they perform. Subjects are brought into existence by being “set under” a prior power that disciplines their psyches in concrete ways. Power that is prior to subjects thus shapes and limits their capacities to act:

¹⁸⁴ Butler 1997(1), p. 15
...the subject is neither a sovereign agent with a purely instrumental relation to language, nor a mere effect whose agency is pure complicity with prior operations of power. The vulnerability to the Other constituted by that prior address is never overcome in the assumption of agency (one reason that ‘agency’ is not the same as ‘mastery’).  

However, subjects’ acts outstrip the power that makes their actions possible. This is because the subject and her acts are an accumulation as well as a “dissimulation” of power. Disciplinary power anchors itself in human bodies by conditioning normative behaviors. But bodies can transgress normative boundaries, reverse or reconstitute traditions of use or practice, enact new possibilities by enacting novel or unexpected performances. So by showing up as the power of a subject that it has disciplined, previously prior, productive power conceals itself, and this concealment anticipates an overcoming of the patterns of action it formerly enacted. Power is transformed by circulating productively and emerging concealed in the actions it helps make possible. So the ambivalence of power’s operation is what explains the excessive character of the speech that it makes possible. Subjects are both enacted by and enacting of power. Power is both prior and posterior to the subject, which is therefore both passive and active. And the “purposes” of power under one of these dual aspects are not necessarily those of power under another. In Excitable Speech, Butler frames these themes in terms of naming understood as both constitutive and potentially injurious:

...by being called a name, one is also, paradoxically, given a certain possibility for social existence, initiated into a temporal life of language that exceeds the prior purposes that animate that call. Thus the injurious address may appear to fix or paralyze the one it hails, but it may also produce an unexpected and enabling response.

---

185 Butler 1997 (1), p. 26. Butler claims that “That linguistic domain over which the subject has no control becomes the condition of possibility for whatever domain of control is exercised by the speaking subject” (Ibid., p. 28). The domain over which the subject supposedly has “no” control presumably includes past ritual practice. But since it is difficult to see why Butler is entitled to the sort of realism about past enactments apparently presupposed by this remark, it seems overdrawn.

186 Ibid., p. 2
…to be addressed is not merely to be recognized for what one already is, but to have the very term conferred by which the recognition of existence becomes possible. One comes to ‘exist’ by virtue of this fundamental dependency on the address of the Other.\textsuperscript{187}

Moreover, the reason why speakers can act through speech, and the reason why their linguistic agency and therefore the commitments and identities they fashion through speech are vulnerable to disruption and failure, are really one and the same reason. It is that, since linguistic practices are social conventions with contingent, unfolding histories, the agency possible through speech is heavily ritualized and “citational.”\textsuperscript{188} “Making community with a history of speakers,” or reenacting past semantic and performative practice, is what gives speech whatever efficacy it has:

If a performative provisionally succeeds...then it is not because an intention successfully governs the action of speech, but only because that action echoes prior actions, and \textit{accumulates the force of authority through the repetition or citation of a prior and authoritative set of practices}.\textsuperscript{189}

However, the citationality of speech makes the power exercised through its use subject to the “ambivalence” we addressed in the previous section on power’s productive operation. As the above passage indicates, the import, force and efficacy of speech is never exhaustively governed by speaker intention—an implication that directly flouts SI’s core commitments to the Necessity as well as the Sufficiency of the Will (a-b). For the ritual patterns of past action “cited” to authorize and animate present speech are by their very nature unstable and prone to transformation: no two enactments may be reduced one to the other. The normative import of present speech is in excess, and \textit{an excess}, of the past practice out of which it grows:

\textsuperscript{187}Butler 1997 (1), p. 5  
\textsuperscript{188}Ibid., p. 27  
\textsuperscript{189}Ibid., p. 51
Indeed, one of the chief lessons of Nietzsche’s *Genealogy* I is that evaluative signifiers like “good,” “strength,” “freedom” and “mastery” are fluid and ritually subject to semantic and performative reversal and resignification over time. Linguistic rituals survive through repetition, but there is always a gap between the ritual history of performance and present and future enactments. This is what Butler calls “the open temporality of the speech act.”\(^1\) Given this, the very traditions permitting certain speech acts to exert a normative or performative force can be challenged and subverted from within. This sort of subversive action can help to disable the relevant acts and create space for alternative oppositional or constructive practices. The general lesson drawn by Butler is that the unstable basis for the power of speech makes that power exceed the conscious control of speaking subjects:

If the subject who speaks is also constituted by the language that she or he speaks, then language is the condition of possibility for the speaking subject, and not merely its instrument of expression. This means that the subject has its own ‘existence’ implicated in a language that precedes and exceeds the subject, a language whose historicity includes a past and future that exceeds that of the subject who speaks. And yet, this ‘excess’ is what makes possible the speech of the subject.\(^2\)

For one thing, there are always multiple competing “traditions of use” within any particular speech community. A genealogy of an ideal or a value is therefore a history of its conflicts and compromises with rival values, or interpretations thereof, within a tradition. Again, as Nietzsche’s account of “the slave revolt in morality” in the First Essay of the *Genealogy* illustrates, traditions of use can be undermined from within. Subordinate groups can come to dominate their former masters, contesting and upending “the lordly right of giving names.” The persistent institution of new practices guided by

---

\(^1\) Butler 1997 (1), p. 15
\(^2\) *Ibid.*, p. 28
fresh ideals nonetheless rooted in past practice shows speech to be markedly un-lordly, un-sovereign, in its efficacy.\footnote{192}

In \textit{Excitable Speech}, Butler is interested in the performativity of hate speech. She appeals to the foregoing considerations to explain why language designed to threaten or wound, like racial slurs, can be counter-appropriated by their intended targets, causing the relevant acts to misfire by compromising the injurious trajectory of the words through which they are performed. This is done either by refusing to “make community” with the relevant tradition of use or by instituting new uses: “If to be addressed is to be interpellated [or called into existence as a speaking subject capable of response], then the offensive call runs the risk of inaugurating a subject in speech who comes to use language to counter the offensive call.”\footnote{193} The counter-appropriation of racially inflammatory language by black comedians and hip-hop artists, as well as “the revaluation of such terms as ‘queer’” by those potentially harmed by its derogatory use furnish contemporary examples of what such efforts might look like.\footnote{194}

\footnote{192} Nietzsche, GM I: 2. A modern example of this understanding of power put to work rhetorically in a socially and politically revolutionary context can be found in the Preamble to the Constitution of the Industrial Workers of the World: “It is the historic mission of the working class to do away with capitalism. The army of production must be organized, not only for everyday struggle with capitalists, but also to carry on production when capitalism shall have been overthrown. By organizing industrially we are forming the structure of the new society within the shell of the old.” (http://www.iww.org/en/culture/official/preamble.shtml). Industrial laboring classes, once called into existence, can unexpectedly come to assume ends quite at variance with those for whose sake they were ostensibly designed by the early “architects” of capitalism. Thanks to Bryce Huebner for calling my attention to this provocative passage. Another example, flagged by Butler, is “the recent reappropriation of ‘civil rights’ discourse to oppose affirmative action in California” at the time \textit{Excitable Speech} was written (Butler 1997 (1), p. 93). Contemporary examples of reactionary, conservative counter-appropriation of the rhetoric and imagery associated with past progressive movements can be found in the Tea Party’s September 12, 2009 “Taxpayer March on Washington” and the August 28\textsuperscript{th}, 2010 “Restoring Honor Rally” held on the 47\textsuperscript{th} anniversary of Martin Luther King, Jr.’s “I Have a Dream” speech, in which Glenn Beck is substituted for King as a unifying symbol of American civil rights advocacy.\footnote{193} Butler 1997 (1), p. 2 \footnote{194} \textit{Ibid.}, p. 14
What this means is that SI’s representation of human agents as sovereign rests on confusion about the way that the exercise of agency is conditioned by the power of speech. Even our capacities to set ends as having a particular character, and our capacities to endeavor to realize them through overt action in a particular manner, are vulnerable to disruption and failure. Since our capacities to act are shaped within contested linguistic traditions—ritual accretions of contingent enactments over time by an indefinite range of different speakers—their exercise is not ultimately under anyone’s final control as long as the ritual practice opens onto a future. No part of the meaning of a speech act is univocally fixed by individual, internal exercises of will performed under sovereign jurisdiction. Rather, the sort of agency made available by the subject’s “linguistic vulnerability” is “something we might still call agency,” though it essentially involves “the repetition of an originary subordination for another purpose, one whose future is partially open.”

Through Butler’s account of linguistic performativity, we thus learn that “A speech act can be an act without necessarily being an efficacious act.” Hate speech may be seen as intended by its users to constitute its targets as socially subordinate by reenacting a ritual history of domination. But a performative utterance designed to injure in this way can fail to produce the intended effects “or initiate a set of consequences” for the addressee. Following J. L. Austin’s work, Butler describes failed performatives as infelicitous. Performative infelicity is possible because of the gap that we noted earlier exists between the ritual history of performance and present and future enactments:

195 Butler 1997 (1), p. 38
196 Ibid., p. 16
…an ‘act’ is not a momentary happening, but a certain nexus of temporal horizons, the condensation of an iterability that exceeds the moment it occasions. The possibility for a speech act to resignify a prior context depends, in part, upon the gap between the originating context or intention by which an utterance is animated and the effects it produces. For the threat, for instance, to have a future it never intended, for it to be returned to its speaker in a different form, and defused through that return, the meanings the speech act acquires and the effects it performs must exceed those by which it was intended, and the contexts it assumes must not be quite the same as the ones in which it originates (if such an origin is to be found).¹⁹⁷

This notion of a “gap” or “faultline” in speech is what explains its susceptibility to disruption and failure, and what undermines the sovereign ideal of establishing an immediate and automatic relation between:

- Intention and utterance: one says only what one means
- Utterance and action: one does only what one says
- Intention and action: one does only what one means to do

Butler’s contention is that the performative faultlines in citational rituals preclude fixing the relationships between these temporally distinct moments of speech.

If agency is derivative in the way Butler suggests, sovereignty does not make sense even as a paradigm, a worthy ideal but one which speaking subjects cannot fully realize in the causal and linguistic worlds they inhabit (the “weak but wide ascription” of sovereignty). For, as we observed earlier, SI is not even disposed to offer a genealogy of the sovereign ideal. It is a purely formal standard of complete and immediate efficacy. Since the standard is supposedly constitutive for agency as such, it is not supposed to have a history. We are simply expected to recognize ourselves as bound by it when we reflect upon the experience of choice and deliberation and see that our consciousness of ourselves as in control of what we do and why entails recognizing the authority of that standard.

But Butler’s account of speech suggests an alternate phenomenology of action, one whereby speakers are often painfully aware of their vulnerability to injury through speech and of their associated dependency upon the address of other speakers for their own standing to speak. For Butler reminds us that subjects have the ability to produce various effects through the use of language, but that doing so requires the deployment of social authority to institute and enforce appropriate ceremonial practices. Examples familiar from J.L. Austin’s work on performativity include the conventions surrounding the establishment of mutually binding agreements, rituals of naming and of adjudication, and so on. For an individual speaker to have standing to issue a linguistic command in a military context, for instance, she must have the relevant authority. More interestingly, Butler presumes that we often encounter our words and even our own identities in painful or even unfamiliar guises:

There is no way to protect against that primary vulnerability and susceptibility to the call of recognition that solicits existence, to that primary dependency on a language we never made in order to acquire a tentative ontological status. Thus we sometimes cling to the terms that pain us because, at a minimum, they offer us some form of social and discursive existence.

All of this despite the fact that “the address that inaugurates that possibility of agency, in a single stroke, forecloses the possibility of radical autonomy.” We are conscious of our linguistic vulnerability by being conscious of having been named by a power that emanates from elsewhere than one’s own consciousness, a productive subordination that counts as “the condition of all of us at the beginning and even,

---

198 What this precisely requires will vary depending on the context of use, and is an issue far beyond our present scope.
199 Butler 1997 (1), p. 26
sometimes, prior to the beginning.” As Butler puts it in *The Psychic Life of Power*, “the vulnerability of the subject to a power not of its own making is unavoidable:”

That vulnerability qualifies the subject as an exploitable kind of being…Bound to seek recognition of its existence in categories, terms, and names that are not of its own making, the subject seeks the sign of its own existence outside itself…

Here it is evident that Butler’s phenomenology is more faithful than SI’s to the experience of one’s own vulnerability to exploitation as a speaker of an unchosen language. For we do not seek affirmation from an *a priori* formal law, but from other vulnerable people “in a discourse that is at once dominant and indifferent.” To be sure, in modern liberal societies, this discourse often deploys terms such as “freedom” and “choice,” but that fact alone hardly vindicates SI’s metaphysically robust, proceduralist account of what these contested terms necessarily entail for any human subject. Butler’s phenomenology of subjection suggests instead that,

Precisely at the moment in which choice is impossible, the subject pursues subordination as the promise of existence. This pursuit is not choice, but neither is it necessity. Subjection exploits the desire for existence, where existence is always conferred from elsewhere; it marks a primary vulnerability to the Other in order to be.

So speaking subjects come to be by occupying space within a ritual practice through an unchosen address that constitutes that very space. When an infant is incorporated into gendering rituals by caregivers and the broader community, it comes to occupy a named space in Butler’s sense. Like other disciplinary regimes, gendering rituals are not established in the fashion considered paradigmatic by SI: by the conscious formation of an intention by appeal to stable, pregiven criteria. They were never *established simpliciter*. Power does not work that way. The assumption that there must

---

200 Butler 1997 (1), p. 31
201 Butler 1997 (2), p. 20
be a sovereign, first-personal agent from whose pure will norms originate is precisely what genealogical inquiry undermines. Rituals do not operate by the conscious deployment of discrete criteria. This is partly what differentiates them from the formal procedures seen as constitutive by SI. They circulate within and between bodies through repetition and amidst the “faultlines” in repetition we have discussed. These gaps are such that

one may well imagine oneself in ways that are quite to the contrary of how one is socially constituted; one may, as it were, meet that socially constituted self by surprise, with alarm or pleasure, even with shock. And such an encounter underscores the way in which the name wields a linguistic power of constitution in ways that are indifferent to the one who bears the name.\textsuperscript{203}

We may conclude that, if Nietzsche and Butler are right about the excessive and productive operation of disciplinary power, SWT (a-b) is false. For speaking subjects are brought into existence by practices of naming and calling arising from elsewhere and exceeding their own history and control. Indeed, this “fundamental dependency on the address of the Other”\textsuperscript{204} and the “vulnerability to the Other constituted by that prior address”\textsuperscript{205} endow speaking subjects with their capacities to act through speech.

Nevertheless, Butler insists that

untethering the speech act from the sovereign subject founds an alternative notion of agency and, ultimately, of responsibility, one that more fully acknowledges the way in which the subject is constituted in language, how what it creates is also what it derives from elsewhere. Whereas some critics mistake the critique of sovereignty for the demolition of agency, I propose that agency beings where sovereignty wanes.\textsuperscript{206}

We will have to postpone our statement and defense of an “alternative notion of…responsibility” until our concluding chapter, when we sketch an erotic ideal of

\textsuperscript{203} Butler 1997 (1), p. 31
\textsuperscript{204} Ibid., p. 5
\textsuperscript{205} Ibid., p. 26
\textsuperscript{206} Ibid., pp. 15-16
integrity based on the account of power and subject formation introduced in this chapter, Agonistic Realism (AR). For now, we will conclude our statement of AR by contrasting Butler’s “citational” account of speech with Korsgaard’s own theory of discourse. Doing so will help us to fortify our account of AR by addressing a likely objection to the criticism of SI supported by it.

_Private will, public speech: against Korsgaard’s theory of discourse_

We have seen that Korsgaard invests subjects with the “sovereign” power of irresistibly calling into existence an intended effect by an irreducibly first-personal exercise of the will. The two sub-theses of her Sufficiency of the Will Thesis (SWT) specify this commitment. Moreover, our discussion of AR helped us to see that SI represents sovereign power and psychic wholeness as complementary regulative ideals for human subjects by ignoring the dependency of their powers of action on contingent linguistic and disciplinary rituals whose excessive “historicity” prevents them from coming under anyone’s sovereign control.

But we might ask: why think that proponents of SI should take this to be a serious criticism? Perhaps they will remind us that, by our own admission, they are committed to a conception of the will as a power of self-determination exercised by agents over and within a private domain of reflexive self-consciousness. To be sure, once willed effects enter the causal and intersubjective world, they may misfire in all the ways Butler enumerates. But SI gives a metaphysical account of action that purports to disclose the _a priori_ conditions of possibility for being a responsible agent, namely subjection to formal standards of efficacy and autonomy. So we must take ourselves to be subject to the same
requirements if we are to be entitled to our practically indispensable conception of
ourselves as responsible agents where “responsibility” presupposes our autonomy and
efficacy. SI concludes that since single unified subjects conscious and in control of the
grounds of their conduct are constitutively subject to these standards, our conception of
ourselves as free and responsible subjects requires us to take ourselves to be such entities,
at least in the paradigm case. It may therefore seem that SI can acknowledge subjects’
integral ties to communities of speakers in a way that does not threaten the argument they
wish to make. A proponent of SI can thus deny the relevance of Butler’s analysis of
power since it discloses the contingent features of speech, whereas SI is concerned
instead with disclosing constitutive, that is to say, inescapably necessary, features of the
agency for which speech is a vehicle.

It turns out that this strategy of response is not available to Korsgaard, for if we
press the question of how agents on her view actually exercise their sovereign capacity to
set and pursue ends, she will answer that, as deliberative reasoners, we do this through
speech. There is therefore a pronounced, unacknowledged tension in her defense of SI:
between her commitment to a notion of volitional privacy that we have already discussed
and her commitment to an ostensibly Wittgensteinian account of the publicity of the
deliberative reasoning through which agents articulate and justify the normative
commitments underlying their practical identities.207 The problem is not just that these
theoretical commitments are incompatible, but that her commitment to the spoken
publicity of reason makes Korsgaard’s defense of SI vulnerable to the unfriendly
implications of Butler’s “citational” analysis.

207 She highlights the supposed parallels between her defense of the publicity of reason and Wittgenstein’s
“private language argument” in The Sources of Normativity, Ch. 4.
Throughout her published work on normativity since at least the mid-1990s, Korsgaard has relied on the same general account of practical deliberation. It is that reasons for belief and action, indeed all genuinely normative states of affairs such as meanings and values, are essentially shared because the speech through which they are articulated is public in the first instance. Practical deliberation is thus to be understood in terms of a view of rational discourse, where “rational” means oriented towards responsibly articulating and evaluating normative demands. We have treated SI as a robust form of the “horizontal” strategy we introduced in Chapter I, and indeed, Korsgaard’s theory of discourse is a version of that view of the basis of normative authority, namely that legitimate authority derives from an ideally unanimous normative demand. We called the assumption that legitimate authority presupposes an ideal consensus the consensus premise (CP) of the horizontal strategy. We also noted that presuming the ideal unanimity of a demand does not necessarily involve presuming its universality. But within SI, it does: according to Korsgaard, legitimately authoritative norms are *ipso facto* objects of an ideally universal normative consensus: Call SI’s universalist version of CP the *universality premise* (UP).²⁰⁸

Korsgaard’s attempt to demonstrate UP founders on a question-begging appeal to her presumption that human agents cannot fail to take themselves to be paradigmatically autonomous and efficacious—that is, sovereignly empowered—*qua* responsible agents. Moreover, Butler’s analysis helps us to see that Korsgaard’s appeal to this presumption leads her to crucially misrepresent the performativity of speech and how it is encountered

---

²⁰⁸ Korsgaard 2009, p. 80: “...we must establish that the reasons embodied in universal maxims must be understood as public, or shareable reasons: reasons that have normative force for all rational beings.”
by speaking subjects. To see why, consider the argument she gives in *The Sources of Normativity* for the claim that meanings are public.\(^{209}\)

She refers to the argument in Wittgenstein’s *Philosophical Investigations* for the impossibility of a “private language” containing meanings that are “essentially private and incommunicable, say for instance a sensation that is yours alone, and cannot be described in any other way than by a name that you give it.”\(^{210}\) Korsgaard takes Wittgenstein’s conclusion to be that meanings are instead inherently public and communicable. She thinks the argument is sound and is applicable, *mutatis mutandis*, to the question of the normativity of reasons. She presents the argument roughly in the following steps:

1. Normative notions are inherently relational because they must admit “the possibility of misunderstanding or mistake”
2. Therefore, “Meaning is relational because it is a *normative* notion: to say that X means Y is to say that one ought to take X for Y; and this requires two, a legislator to lay it down that one must take X for Y, and a citizen to obey.”\(^{211}\)
3. The notion of a reason is relevantly like the notion of a meaning: “to say that R is a reason for A is to say that one should do A because of R; and this requires two, a legislator to lay it down, and a citizen to obey.”\(^{212}\)
4. Relational states of affairs are “inherently shareable”\(^{213}\)
5. Therefore, *qua* relational, meanings and reasons are inherently shareable (from 1 through 4)

\(^{209}\) Chapter 9 of Korsgaard’s *Self-Constiution*, “Integrity and Interaction,” is partially devoted to advancing the same argument from *The Sources of Normativity* we are considering here. We focus on the version of the argument in *Sources* because it is more compressed and perspicuous.

\(^{210}\) Korsgaard 1996, pp. 136-137

\(^{211}\) *Ibid.*, p. 137


\(^{213}\) *Ibid.*, p. 135
Now, it is unclear whether Butler’s analysis of speech precludes “the private
language argument” so construed. This is a difficult question that we must lay aside.
What interests us is how Korsgaard uses the argument to motivate a separate claim about
the nature of the meaningfulness or intelligibility of our deliberative practices. For we
should notice at once that step 4 in the argument involves a non sequitur. It plays a
critical role in Korsgaard’s argumentative strategy but she nowhere defends it.

The argument begins (in step 1) with the notion that normative states of affairs are
relational in that they are by their very nature defeasible or capable of being either
properly or improperly taken-as by responsible parties. That is, normative states of
affairs essentially require the involvement of a responsible subject capable of giving
normatively salient uptake. Notice that this already frames the issue of normativity as in
the first instance one of first-personal engagement and not second-personal response to
prior address. This already militates in favor of SI over AR, for according to the latter
subjects are called into a fundamentally dependent and vulnerable existence by a power
not of their own choice or making.

The picture already beginning to form is familiar from our discussion of SI:
however things stand in the causal world, in the paradigm case disclosed by
transcendental analysis agents encounter norms by encountering one another already fully
formed and secure in their powers of self-determination. Normative issues arise in
context of interaction between individual wills encountering one another qua self-
determining legislators making claims on one another. This explains the appearance of
the familiar figure of legislative authority in steps 2 and 3 of the argument. However, the
need for the possibility of responsive uptake of proprietary salience underdetermines a
reliance upon the figure of essentially first-personal, legislative self-determination. The availability of an alternate vocabulary of subject formation and responsible agency, namely AR, explains why.

But Korsgaard does more to prejudice the outcome of her analysis in favor of SI. She appeals to the reflective character of deliberation to reinforce her claim that we must take the reasons we deploy when articulating and justifying normative demands to be inherently shareable (step 4). She claims that whereas “people suppose that practical reasons are private because they suppose that reflection is a private activity,” in fact reflection occurs within shared languages furnishing terms and concepts whose meanings are mutually accessible by different users. Misunderstandings occur, to be sure, but mutual understanding is the default mode:

> It is nearly impossible to hear the words of a language you know as mere noise. And this…means that I can always intrude myself into your consciousness. All I have to do is talk to you in the words of a language you know, and in that way I can force you to think. The space of linguistic consciousness is essentially public, like a town square.

This rhetoric of inherently stable and transparent, default intelligibility gets to the heart of what is wrong with Korsgaard’s interpretation of the publicity of reason(s) in terms of any version of CP. Its distorting effect can be seen in her discussion of the significance of being named or called by another:

> If I call out your name, I make you stop in your tracks…[Or] if you walk on, you will be ignoring me and slighting me. It will probably be difficult for you, and you will have to muster a certain active resistance, a sense of rebellion. But why should you have to rebel against me? It is because I am a law to you. By calling out your name, I have obligated you. I have given you a reason to stop.

---

214 Korsgaard 1996, p. 136
215 Ibid., pp. 139-140
216 Ibid., p. 140
Recalling Butler’s discussion of naming, we see that Korsgaard gets much of what this entails right but falters when she attributes an automatic performative felicity to the call. A second-personal address is an act of speech, to be sure, but we saw earlier why the act is never sufficient by itself to produce a determinate set of consequences, such as imposing an obligation or stopping someone dead in their tracks. The idea that speech functions this way is the now-familiar “sovereign conceit” concerning its efficacy. Korsgaard’s tacit reliance upon the conceit explains her automatic insertion of the figure of the self-determining legislator into the primordial dynamics of normativity in steps 2 and 3 of her version of the private language argument.

The interpellative call initiates a scene of agency, a context in which questions of responsible agency arise, but Butler’s analysis shows why it is a mistake to assimilate “social interpellation to the divine performative” whereby “the authority of the ‘voice’ of ideology, the ‘voice’ of interpellation, is figured as a voice almost impossible to refuse.”217 Butler’s account of speech underscores the way in which it, and so the agency it makes possible, always in some ways exceeds the control of those whose agency it helps make possible. Contestation and “normative ambivalence” is therefore the hallmark of discourse where issues of meaning and value arise, not ideally unanimous consensus secured through the paradigmatically authoritative pronouncement of binding law (à la some version of CP). As Butler puts it, “…the utterance has become a scene of conflict.”218 AR is hardly committed to the privacy of speech. But it offers an account of its essentially contested publicity that casts Korsgaard’s appeal to the “town square” in an ironic light.

217 Butler 1997 (1), p. 31
218 Ibid., p. 91
In short, Korsgaard distorts the performativity of the interpellative call because she fails to notice that taking speech to be the medium of reflection conflicts with her insistence upon the sovereignty of agency. She represents speech as a static and transparent medium of shared thought, the common coin of shared meanings. This overlooks the fact that speech is a temporally extended ritual conditioning but exceeding the power of the subject, and as such it is beset with fractures. There are performative gaps between intention and utterance, between utterance and action, and between an action and its effects. What a speaker aims to do through what she says can misfire or come back to haunt her by taking on meanings or purposes she neither foresees nor intends: “the utterance does not have the same meaning everywhere.”

Korsgaard denies this almost in these terms when discussing “what makes you take my reasons into account, or bridges the gap between your reasons and mine,” insisting that “there is no gap to bridge.” If there were, Korsgaard proposes that “I would have to hear your words as mere noise, not as intelligible speech. And it is impossible to hear the words of a language you know as mere noise.” Korsgaard thereby poses a false dilemma. Untrammeled sovereignty and utter inarticulacy are not realistic, let alone exhaustive, theoretical alternatives. Since the speaking subject’s power is a linguistic one, that is, a capacity to initiate consequences through speech, the faultlines in speech are faultlines of agency. Denying the “inherent shareability” of meanings and reasons does not commit one to affirming existence of purely private meanings and reasons. On the contrary, once we jettison the sovereign conceit concerning agency, we may abandon

---

219 Butler 1997 (1), p. 91  
220 Korsgaard 1996, p. 143
that binary opposition in favor of a more nuanced understanding of the equivocal relationship between the power of speech and the power of speaking subjects.

4: conclusion: against shared horizons

Butler’s analysis of speech helps us to see that meanings are not possessed in common by speaking subjects. This is not because meanings are private but because they don’t belong to anyone in the first place. The account of productive power and excessive agency sketched in this chapter, Agonistic Realism, helps explain what this strange expression means and why it is credible after all. Korsgaard’s theory of rational discourse fails to account for the power of speech because it rests on the question-begging assumption that a binary opposition between “shared” and “private” exhausts the relevant possibilities. With this in mind, we proceed to Ch. IV where we will see that many contemporary social theorists who rely upon the figure of the “shared horizon of intelligibility” to account for normativity and agency fail to do so well because they make the same sort of the mistake.
CHAPTER FOUR

BEYOND SHARED HORIZONS:
CONTESTED PRACTICES, PRACTICES OF CONTESTATION

I mistrust all systematists and avoid them. The will to system is a lack of integrity.\textsuperscript{221}

1: introduction

We have seen that speech has a life of its own because the power that produces subjects capable of acting through it is excessive. No speaking subject wields immediate, unequivocal control, even of the sort over its “internal” psychic domain required by SI. So even \textit{qua paradigm}, the ideal of immediately efficacious volitional control is a non-starter according to the view of power and agency we are defending, Agonistic Realism (AR). AR shows how, by virtue of the ritual and institutional conditions of its efficacy, our speech can take on meanings and produce effects both beyond and contrary to those we mean to produce through our conduct.\textsuperscript{222} AR therefore entails that by its very nature, or \textit{in the paradigm case}, the power we exercise as linguistic agents assumes purposes other than those we intend. So, in short, power precludes sovereign agency because there

\textsuperscript{221} Friedrich Nietzsche, TI I: 26
\textsuperscript{222} Another vivid class of illustrations considered by Judith Butler surrounds the ostensibly sovereign speech of legal authorities, particularly the judiciary exercise of state power when interpreting the First and Fourteenth Amendments. Butler considers a number of controversial Supreme Court rulings on cases involving the adjudication of hate speech, showing that judiciary speech can enact forms of violence quite contrary to the general purposes for which it is presumably designed as well as those intended by justices in particular cases. See especially Chapter 2 of \textit{Excitable Speech}, “Burning Acts, Injurious Speech.”
are no sovereign powers. This means that power is not well understood on the model of
sovereignty but rather in terms of the Nietzschean ideas of a founding subordination,
unchosen attachments, excess and productivity set out by Judith Butler. These relations
preclude sovereign control, and so insofar as they enter into the formation of human
agents, human agents are shaped, enabled and limited by a multiplicity of contending
pressures and influences, all of them concretely embodied and in tension, rather than by a
single formal principle somehow aloof of the causal order. However, the power of
speech conditions the exercise of power through speech. This is why the psychosocial
reality of power is not a counsel of despair for a discourse of responsibility. On the
contrary, Butler’s account of the contestation endemic to speech communities shows that
it creates space for resistance and social transformation. These tactical openings
correspond to fractures or “faultlines” in the performative terrain of discourse about
contested meanings and values.

All this entails that final horizons are unavailable to human speakers, since there
is no final act of speech. That is, speech as non-sovereign and excessive precludes fixing
the essence of things through speech, e.g. articulating constitutive standards once and for
all. Only finite horizons are possible for finite agents, and the virtue of a Nietzschean
psychosocial genealogy of ‘power’ is that it is marked by an uptake of its own finitude,
that is, it embraces its own provisional status by recognizing the dynamism and excess of
the language in which it is set out. It does this by being set out in precisely these terms.
Hence we try to avoid the “family failing of philosophers” identified by Nietzsche, and
hope instead to exhibit “the virtue of modesty” he recommends.
In this chapter we will conclude our presentation of AR by turning to its third core feature, a denial of what we identified in Chapter II as the third core feature of SI: stable analytical horizons of intelligibility for (i) thematizing the conditions for the possibility of the constitution of the active subject, and so for (ii) normatively regulating the exercise of its agency. We will therefore revisit the figure of the “horizon” that we encountered in our first two chapters, evaluating it in light of the priority assigned to contestation within AR. Our aim is to challenge it directly, keeping in mind the other two features of AR laid out in Chapter III. We will find that a commitment to AR does not preclude recourse to “horizons” for analytical purposes, but only as they are seen within SI. Normative practices are intersubjective and temporally extended but not essentially shared. As we began to see at the end of the previous chapter, their publicity primarily consists not in their inherent, paradigmatic susceptibility to consensus, but in the fact that they are persistently contested.

We will begin by reviewing some primary themes in the accounts of normativity and agency developed by two Heideggerian philosophers, John Haugeland and Jonathan Lear. This will help to show that the figure of the horizon plays an important analytical role within social philosophy and not just moral or erotic psychology. We will then criticize the horizontal strategy in social philosophy by recurring to some themes already introduced in our discussion of AR. We will conclude by drawing some general implications for how appeal to the horizon should generally figure in theoretical analysis of agency. Since we are addressing vexed questions of philosophical methodology, this chapter’s analysis is offered as a provisional guide to what more should be said in this area. It is more important for our present purposes that, having cast doubt upon the
notion of a “shared horizon” in light of the significance of contestation, we may proceed in the following chapter to sketch an alternative account of integrity as an erotic ideal for human agents as conceived by AR.

2: social horizons in Charles Taylor’s *Ethics of Authenticity*

In Chapter I, we first encountered the figure of the horizon as a key concept for analyzing social practices whose participants have and work out motivating and orienting ties to values, ideals and other candidate erotic objects. We saw that Robert Pippin effectively relies on the notion of the horizon in his social theoretical work on commitment. Charles Taylor has relied on it perhaps more heavily and overtly than any other contemporary social philosopher. He offers a representative statement of his view of the significance of social horizons of significance in *The Ethics of Authenticity*.

The book’s aim is to rescue the ideal of authentic existence from the trivialization and justificatory “inarticulacy” ostensibly threatened by its contemporary association with a robust subjectivism about value. In Chapter IV, “Inescapable Horizons,” Taylor seeks “certain general features of human life that condition the fulfillment of this or any other ideal,” in which he may ground his defense of authenticity as a worthwhile goal for human beings. He claims to find a suitably general and necessary feature of human life in “its fundamentally *dialogical* character,” asserting that “We become full human agents, capable of understanding ourselves, and hence of defining an identity, through our acquisition of rich human languages of expression” into which we are inducted by others with whom we share them throughout our lives. Such languages include “not
only the words we speak but also other modes of expression whereby we define
ourselves, including the ‘languages’ of art, of gesture, of love, and the like.”

Taylor believes that recognizing the dialogical character of human life will erode
our commitment to the subjectivist thesis that “things have significance not of themselves
but because people deem them to have it—as though people could determine what is
significant, either by decision, or perhaps unwittingly and unwillingly by just feeling that
way.” Dialogicality instead entails that “Things take on importance against a
background of intelligibility.” Taylor proposes, “Let us call this a horizon,” adding
that “horizons are given,” or “pre-existing.” The idea is that social frameworks of
meaning and value determine what can and cannot count as important for individuals
because those frameworks are relatively stable, or fixed in advance of the self-defining
activity of individuals. Since my effort to define my identity is an effort to determine
what matters to me and why, my efforts to do so can only be “inescapably” constrained
by horizons of meaning if they determine what can count as significant for me in a way
that I cannot resist through my effort. This explains Taylor’s insistence that “Which
issues are significant, I do not determine.” He even goes so far as to assert that “It

---

223 Taylor 1991, p. 33
224 Ibid., p. 36. By “people,” Taylor should probably be taken to mean individuals considered in isolation
from others and from cultural traditions. For his insistence that individuals can only define their identities
against “a background of things that matter” (p. 40) in fact seems to require that people acting together over
time to maintain cultural traditions collectively and perhaps indirectly invest things with significance. This
would seem to follow unless there is a transcendent source of value that confers significance upon things
that matter to people. But though Taylor never expressly endorses this idea, he sometimes comes close, as
when he claims that practices “that opt for self-fulfillment without regard…to demands of any kind
emanating from something more or other than human desires or aspirations are self-defeating” (p. 35). But
the reference here to “human desires or aspirations” contains the same ambiguity that we find in his
reference to “people” above. How we should resolve it on Taylor’s behalf is a question we must leave
aside.
225 Ibid., p. 37
226 Ibid., p. 39
227 Ibid., p. 38
228 Ibid., p. 39
follows that one of the things we can’t do, if we are to define ourselves significantly, is suppress *or deny* the horizons against which things take on significance *for us.*”

Taylor thus provides a clear statement of the “horizonal” strategy we identified at the outset of our inquiry. On this view, horizons are frameworks of value and significance that determine what can and cannot count as important for individuals. They determine what persons, objects or ideals are legitimate candidates for the erotic attachment of individual participants within social practices in which such questions arise. Horizons thus enjoy absolute priority over individual action in theoretical analysis of what makes normatively intelligible action possible in the first place, just as they have priority in discrete, context-specific efforts to assign a meaning and value to the actions of particular individuals within traditions of value. We have devoted a great deal of time to considering SI as a robustly neo-Kantian theory of the horizon. It turns out that the overall strategy is also shared not just by Charles Taylor but by social philosophers whose theoretical commitments are robustly Heideggerian.

3: Heideggerian horizons: “the odd stray” and normative breakdown

In his essay “Heidegger on Being a Person,” John Haugeland offers a sympathetic reconstruction of Martin Heidegger’s view of personhood. The view assigns high priority to sociality and derives many differentia of personhood (such as speech,

---

229 Taylor 1991, p. 37 (my emphasis)
230 Though as we saw in Chapter II, the formal norms constituting and governing action *as such* enjoy absolute priority under SI, whereas in Taylor’s formulation, horizons of intelligibility are essentially socio-historical and supply conditions for meaningfully defining one’s identity, not the conditions for acting as such. As we saw in Chapter I, what makes a view of normative practice horizonal in our sense is its absolute privileging of either identity *or* action as supplying the conditions of possibility for normatively intelligible agency.
231 Haugeland’s discussion is restricted to an exposition of the account of personhood he claims to find in Heidegger’s *Being and Time* (1927).
reasoning or concept-use) from “our basic communal nature.” Haugeland begins his account with a sketch of a community of creatures hard-wired with the “conformist” disposition to aggregate, imitate and censor one another’s behavior, with the resulting formation of “herds—distinguishable, reidentifiable clusters of animals, separated by clear gaps where there are no animals (save the odd stray).” To be a person is to be a self-responsible member of some such herd.

Let us clarify. For Heidegger, according to Haugeland, everything with which we meaningfully deal in everyday life is constituted as whatever it is by being caught up in a network of social practices structured by norms governing the relevant takings-as: to be a baseball is just to be the sort of thing which is taken as an appropriate object for players to hit when at bat, and so on. Objects so construed, “paraphernalia” in Heideggerian parlance, cannot be individuated independently of the network of mutually referring concepts in which they are embedded and in whose terms their significance is to be articulated:

The totality of all paraphernalia *cum* referral relations is called the “referral nexus of significance;” but since paraphernalia is taken broadly enough to include practically everything with which we ordinarily work, cope, or bother (except other people), this totality is tantamount, in fact, to the everyday world. The everyday world…is…the ‘world’ of daily life and affairs…It is essentially a cultural product, given determinate character by— instituted by—the norms of the conformists who live in it.

The anyone itself…articulates the referral nexus of significance (p. 129 *B&T*).

---

232 Haugeland 1982, p. 15
233 Ibid., p. 16
234 The centrality of practices of “taking-X-as-Y” in Haugeland’s reconstruction of Heidegger’s theory of social normativity recalls Korsgaard’s reconstruction in *The Sources of Normativity* of Wittgenstein’s “private language argument” for the publicity of meaning and reasons. This suggests a salient interpretive bias common to Korsgaard and Haugeland that we must postpone examining further, except to say that our reading both SI and Heideggerian social philosophy as horizontal provides one formulation of the shared bias. The important question of its “genealogy” is beyond our scope.
235 Haugeland (1982), p. 18
“The anyone” is the norm-wielding community writ large whose ongoing takings-as behavior constitutes the world, the meaningful horizon, within which cultural and social practices are possible.

Now, persons are no less essentially dependent upon being socially instituted by being taken up as significant by practitioners, but differ from mere objects (indeed from everything else) by virtue of being taken up by participants in social life as a “units of accountability,” patterns “of normal dispositions and social roles…a subpattern of Dasein—an institution.”236 People differ from other instituted types because we ourselves are appropriate objects of censure when our behavior runs afoul of relevant norms. And, crucially for us, the way of being of persons has a kind of ontological priority on the Heideggerian picture: “Moreover, institutions of this kind are essential for all others; for without accountability there would be no censorship, hence no norms, no anyone, no Dasein at all. Thus, accountable cases [of Dasein] are ‘primordial’ institutions.”237 This fundamental structure of Dasein Heidegger calls “in-each-case-mineness.”

Haugeland translates the case of the hypothetical community of conformists and odd strays into an account of the sources of normative force: “When behavioral dispositions aggregate under the force of conformism, it isn’t herds that coalesce, but norms…distinct, enduring clusters of dispositions in behavioral feasibility space, separated in that space by clear gaps where there are no dispositions (save the odd stray).”238 In his book Radical Hope, Jonathan Lear’s aim is to articulate more precisely what takes place in such penumbral gaps where norms have begun to lose, or in the

236 Haugeland 1982, p. 21
237 Ibid.
238 Ibid., p. 16
extreme case of severe cultural collapse, have all but lost, their grip on those who inhabit them. How does one prefigure and begin realizing alternate normative possibilities when the concepts and skillful practices heretofore responsible for making one’s conduct intelligible have broken down, perhaps to the vanishing point as in the case of cultural devastation for communities and, arguably, in the case of severe, disabling depression for individuals?  

I think the Heideggerian view outlined by Haugeland, shared in large part by Lear and endorsed in spirit by Taylor relies too heavily on an explanatory appeal to conformism. I want to suggest that it tends to misrepresent erotic practice as ideally a fait accompli, and that this entails a psychologically distorted characterization of what we do and are like. After all, there is good reason to suppose that nonconformists integrally contribute to the dynamic normative life of any social community, and our social ontology should countenance this. This is the case, first, in the weak sense of “non-conformity” attaching to mere deviance, however inadvertent or unconscious. One may say that opportunities for mere non-compliance with going norms are built into the pragmatic structure of the interactive episodes which those norms make intelligible, hence performatively “available,” for participants. After all, the terms in which their normative import may be parsed by participants—linguistic, conceptual, and other terms

---


240 My use of the expression fait accompli in this context for these purposes is broadly indebted to the argument against what we are calling “identity first” versions of the horizontal strategy advanced by Patchen Markell in his 2003 book Bound by Recognition. Though Markell does not use a vocabulary of horizons, is not concerned with issues of commitment or integrity as such, and addresses normative problems that are more overtly political than ours tend to be here, his criticism of discourse about the moral import of intersubjective recognition is similar in spirit to our worries about the horizon strategy generally and SI in particular.

241 In our next and concluding chapter, I will suggest why it legitimates an ethically and politically irresponsible erotic ideal as part of my defense of an alternative ideal of integrity based not on SI or any other horizontal theory of agency but on AR.
of meaningful bodily interaction—are pragmatically and semantically porous and inexhaustible within the horizon of any such episode [or possible set thereof?]. Butler’s account of the “excitability,” or excessive character, of linguistic agency goes some way towards showing why.

Consider for instance [provide examples: flirtatious behavior, funeral orations, dining etiquette, sharing confidences, accepting gifts, rendering assistance, and virtually any other norm-governed interpersonal practice]. Without constitutive room for deviance of this routine sort, pockets of ambiguity in “feasibility space” built into the behavioral dispositions which Haugeland says define its boundaries, one is stuck with an account of norm-governed life that is structurally static. The resulting conceptual picture of life is lifeless, a poor model of what people do and are like. If norms can ever lose their grip on us—something whose possibility Haugeland and Lear’s Heideggerian views of social practice are supposed to illuminate—then they can always lose their grip. That is, their grip can be loosened at any time. Whether this happens and when are entirely dependent upon the way things fall out interpersonally, and it is part and parcel of the historicist picture Heideggerians appear to want to offer that this is always a socially emergent and historically contingent affair. To this extent they are in accord with the emphasis upon contingency in genealogical inquiries such as Nietzsche’s and Butler’s.

I do not claim that Lear and Haugeland deny this. Lear explicitly affirms it in *Radical Hope* when discussing the “peculiar vulnerability” to which all human beings are subject, namely our constitutive dependency as agents upon the intelligibility furnished by a set of background practices, and hence our necessary exposure to the risk of damage wrought on our capacity to act by their actual, historical collapse:
What is this possibility of things’ ceasing to happen? If this is a possibility, it is a possibility we all must live with—even when our culture is robust, even if we never have to face its becoming actual. It is a possibility that marks us as human. 242

Lear thus recognizes that the possibility of collapse is ever-present, ontologically speaking. 243 But our account of AR so far suggests that this is not enough. It suggests that the psychological and linguistic conditions of deep contingency which make severe collapse always possible fuel contestation that is itself always actual as long as speaking subjects remain dynamically alive.

We suggested that there is a weak sense of non-conformity—“mere” deviance from normative constraints—that is a structurally essential part of the dynamic life of any social practice. There is another, stronger sense in which non-compliance with norms otherwise taken to be healthy and robust is a structural precondition for the real concrete instantiation of a way of life understood in Heideggerian terms as a rich and co-referring set of norm-governed social practices. This is the sort of non-conformity that is relatively conscious and articulate, that aims to expand the boundaries of existing understandings or modify the scope or force of reigning norm-governed activities. Lear presents the interpretation of dreams as traditionally carried out by the Crow, the Native American society he focuses on in Radical Hope, as an example of “an established practice for pushing at the limits of their understanding…[which] gave the tribe resources for thought—for practical reasoning—that would not have been available to them in any other way.” 244

242 Lear 2006, p. 9
243 There is obviously a great deal of historically contingent variation in the degree of threat of collapse from one community to the next.
244 Ibid., p. 66
Dream interpretation among the Crow is thus a cultural practice designed to draw upon Crow “traditions in novel ways in the face of novel challenges.” Lear insists that the process is radical and even “enigmatic.” It is not as though norms governing the use of concepts from one corner of Crow life were sought for enlistment in the service of illuminating problems concerning the normative deployment of concepts in another corner. In the historical circumstances Lear is concerned with in *Radical Hope,* this apparently routine form of internal cultural self-correction would not have been sufficient. For Lear describes severe conditions of impending cultural collapse in which traditional Crow life was rapidly losing its intelligibility in the face of massive upheaval. The problem the Crow faced was one of finding ways of understanding the demise of their way of understanding the world. Lear’s claim is thatCrow life *itself* was therefore increasingly unable to provide the Crow with adequate resources for understanding what was happening to them as they were driven from their traditional nomadic-hunter-warrior way of life onto squalid, sedentary reservations.

Notice that, by Lear’s account of the “enigmatic,” prefigurative Crow practice of dream interpretation, existing norms are nevertheless seen as capable of supplying resources for expanding the field for the deployment of those very resources. But if Crow concepts have really run out, this looks odd. How can existing norms which are seen as shallowing out and requiring supplementation or adaptation in light of genuinely “novel” circumstances underwrite such a process on their own behalf? We need an account of what makes such “enigmatic” prefigurative practice so different from “routine” cultural self-interpretation, an account of what made the circumstances of Crow collapse so “novel” and ordinary cultural practices (especially those of self-interpretation

---

Lear 2006, pp. 65-66
and internal revision) so “radically” different. In short, we need an explanation of what ontological distinction underwrites the rhetorical differentiation Lear makes between ordinary Crow life (and by extension, the up and running, non-collapse condition of any way of life) and Crow life on the verge of the sort of major transformation Lear characterizes as “collapse.”

Part of any sensible answer to the questions just asked would seem to have to be that no system of norms, no Heideggerian “way of life,” forms either a closed system or a fixed pattern. This is precisely what AR requires, and it seems that Haugeland and Lear do not take its consequences seriously enough. One notable consequence, especially relevant here, is that non-collapse social practice is not categorically different from the sort that takes place in conditions of impending collapse. Lear overdraws this distinction to great rhetorical effect, but distorts the social ontology in doing so.

We will return to the two senses of non-conformity distinguished above when we proceed to a positive articulation of integrity as a virtue of practical engagement, for as a species of non-conformity, its deployment is keyed to precisely the sort of ontological structures nonconformists (of both sorts) characteristically exploit. But first, to see how Haugeland’s appeal to conformism goes too far, we must return to his treatment of the

---

246 This “by extension” is one Lear himself explicitly authorizes in the first few pages of *Radical Hope*: “The Crow actually did endure a cultural catastrophe, and looking at their actual experience may make it somewhat easier to grasp this elusive possibility of things’ ceasing to happen. But the possibility that concerns me is not the special province of this or any other culture: it is a vulnerability that we all share simply in virtue of being human,” Lear 2006, p. 8.

247 One danger of failing to take this implication seriously is succumbing to rhetorical hyperbole when addressing real historical cases offered as illustrations of one’s social ontology. For instance, to some extent Lear’s rhetorical strategy in *Radical Hope* involves portraying Crow society as a closed system, shielded from penetration by the influence of non-Crow social practices and cultural meanings. And this is despite his concession that this was not strictly the case, as the Crow had many dealings with Western European, largely French Canadian, traders in the American Northwest dating from the 18th century. However, if Crow society were not seen as a closed system, it would seem far more difficult to license the hard interpretive line Lear takes regarding the “last great” Crow chief Plenty Coups’s provocative remark that “...when the buffalo went away the hearts of my people fell to the ground, and they could not lift them up again. *After this nothing happened,*” (my emphasis), Lear 2006, p. 2.
ontologically distinctive character of persons. Haugeland quotes Heidegger as saying that “The ‘essence’ of Dasein lies in its being extant.”248 “Extant,” which translates the German word “Existenz,” is to be contrasted with “real;” reality is the mode of being of traditional res, an independent thing or substance. Dasein, by contrast, is not a substance; rather, it is the character of beings which concretely embody, or live out, an interpretation of their own being, one which shows up in the network of co-referring social practices composing their idiosyncratic “form of life.” To be extant, then, is to be that which is its own efforts to understand that which it is, where understanding is equated with competence or know-how rather than detached, thematic cognition.249

In Aristotelian fashion, for Heidegger this conception of what it is to be a person goes hand in hand with a conception of how one excels as such an entity. What is involved in being good at being (knowing-how to be) oneself, or being the individual who one is? According to Haugeland, it’s a matter of practically negotiating the potentially competing proprieties associated with the various roles one inhabits at any particular time. Haugeland distinguishes two ways of dealing with this:250

1. “Sliding:” dealing with the challenges and decisions ad hoc, choosing the path of least resistance; “forgetting” about the temporarily less conspicuous demands pressing upon one; “this is to remain dispersed in the worldly”

2. “The opposite possibility is to confront the conflicts, and resolve them: that is, to make up one’s mind;” this is the “exercise of a higher-level disposition which we might call ‘self-criticism’”; Heidegger calls it “conscience”

---

248 Haugeland 1982, p. 21
249 This is my attempt to paraphrase Haugeland’s version of Heidegger’s unlovely formulation, which is the closest Haugeland says he comes to defining being extant.
250 Ibid., p. 23
Notice that Haugeland contrasts consciously, and conscientiously, confronting conflict in order to “positively” adjudicate them, on the one hand, with “passively” letting one’s dispositions take the lead on the other. But this distinction seems misleading. After all, one may seldom consciously deliberate, but in the sort of cases Haugeland appears to have in mind, we cannot evade responsibility for the policies we carry out in practice, however well or poorly (or rarely) we have deliberated about them at any stage along the way. And on the other hand, even Haugeland presents the path of least resistance as something in some sense chosen. It seems reasonable to doubt both that we are ever as passive as Haugeland’s “sliders” would seem to have to be, and that we need be as self-consciously active as he suggests to be properly held responsible for the many sorts of acts and omissions which characteristically pertain to the roles we inhabit.

Perhaps the difference between sliding and self-criticism is supposed to lie in the mode, or performative character, of engaged activity peculiar to each type. This is a promising thought. Haugeland’s way of spelling out just what “conscientious” self-criticism looks like, however, is troubling: “…as a unit of self-accountability… I find and root out an inconsistency in my overall self-understanding…the only end is self-constancy—a clearer, more coherent self-understanding ability to be me.”

Strikingly, Haugeland here ascribes an erotic ideal to Heidegger’s philosophy that is, at least for our purposes, strictly equivalent to SI’s view of integrity as requiring wholeheartedness or the ideal absence of psychic ambivalence brought on by subjection to normative conflict. We should know by now why this way of framing an ideal of self-responsibility is psychologically dubious. The self-guided quest for a maximally consistent, conflict-free ordering of one’s dispositions is a delusional errand of masterful sovereignty. It can only

---

251 Haugeland 1982, pp. 23-24
rely on our allegiance if we ignore or denigrate the integral role played by the
transformation and loss fueled by persistent contestation over values and ideals. If AR is
right, then, the Heideggerian account of “conscience,” no less than SI’s a priori theory of
Moral Personhood, is a false solution to the problem posed by the Underground Man.
The Underground Man is pathologically subject to “vacillating unwittingly between one
‘me’ and another,”252 but we cannot hope to properly diagnose and overcome this
pathology by wishing away the conditions which fuel it or any other mode of dynamic
erotic activity.

Haugeland identifies “resoluteness” as the Heideggerian term for the skillful
capacity to maximally purge one’s life of normative tensions: “As the resolution of
conflicts that lead to wavering inconsistency, [this virtue] is also resoluteness.”253 As a
Heideggerian version of SI’s erotic ideal of wholeheartedness, Haugeland’s account of
“resoluteness” supplies a useful foil for the agonistic account of integrity we will provide
in the next chapter. But beforehand we must correct some of the ontological
oversimplifications on which Haugeland and Lear rely. Once we begin to better
understand the condition of breakdown itself, we will be in a good position to say what
erotic subjects who take it seriously as agents are characteristically like.

4: power’s perspectives: flexible, dynamic horizons

We have come to the last feature of power to be explained, that it is dynamic in a
sense that precludes the horizonal analytical strategy pursued so robustly by SI. In
Chapter II, we saw that horizonal views establish a relationship of absolute priority

252 Haugeland 1982, p. 23
253 Ibid., p. 24
between identity and action as constitutive dimensions of human agency in order to explain what makes norm-guided action possible as such. We can shed light on the dynamic character of power by reference to its “excess,” which we discussed in the previous chapter. Doing so will help explain why we do not wish to supply a metaphysically robust theory of power. In fact, we do not aim to provide any sort of exhaustive account of power, subject formation or agency.

So what do the objects and the method of our analysis have to do with one another? Well, one thing meant by our saying that power is excessive is that it outstrips any particular exercise of the agency inaugurated by it. It therefore partially eludes any determinate characterization since efforts to exhaustively specify or articulate it are themselves exercises of agency conditioned by power always partly arising “from elsewhere.” This is to say that power is what makes any thematic exposition of agency, too—precisely as an exercise of agency—incorrigibly subject to deficiency, excess, disruption and failure in the sense that we saw Butler believes precludes sovereignty as a suitable model of human agency. Power thus compromises any attempt to exhaustively articulate the conditions or structures responsible for any domain or feature of human life. This naturally includes our analysandum, namely power as encompassing the range of effects and relations responsible for the enduring contestation in psychic and social life.

This is what is meant by the third feature of power, that it is *dynamic*. Since power is what generates contestation, (a great deal of?) normatively salient activity takes the form of contestation, and subjects acquire and exercise their normative powers in and through contestations over meanings and other resources necessary for psychic and social life, power constitutes selves by underwriting the acquisition and exercise of our

---

254 I do not yet know which claim my position minimally requires.
manifold capacities to make a difference as participants within psychic and social practices where things are at stake.

The positive characterization of power which underwrites Agonistic Realism is therefore not meant to be exhaustive. It is not even meant to be part of an exhaustive theory insofar as theory ideally aims at a perfectly general, universally valid, systematic exposition of its object domain. Here we emphatically part ways with Korsgaard, who wants to disclose once and for all the conditions for the possibility of rational agency, including first-personal responsibility, personal integrity, and so on. Our ambitions are more modest here and in principle.

Instead we follow Raymond Geuss in thinking that, “It is probably a mistake to treat ‘power’ as if it referred to a single, uniform substance or relation wherever it was found. It makes more sense to distinguish a variety of qualitatively distinct kinds of powers.” Geuss gestures at some of the different phenomena one can designate with the term ‘power’, including the variety of techniques and relations involved in coercion, persuasion, diplomatic or rhetorical acumen, and so on. We will be satisfied if we exhibit some characteristics of power which facilitate a suitable comparison with corresponding features of Sovereign Idealism and underwrite an alternative erotic psychology.

The three positive features of power advanced within AR are offered as provisional descriptions, “foreground” or “perspective estimates” in Nietzsche’s sense, which represent it by exhibiting salient ‘family resemblance’-style relations rather than through systematically establishing necessary and sufficient conditions. This means that our account of power has implications for questions about the proper scope and tenor of

---

255 Geuss 2008, p. 27
256 Nietzsche, BGE Part Two: 34
inquiry about agency. For if power is a critical dimension of human activity, and it precludes sovereign agency by its very nature, then there can be no sovereign discourse about it or anything else. And, again, insofar as “theory” ideally aims at exhaustive, unequivocal articulation of the possible range of effects and relations fitting some description—such as, in our case, “those responsible for contestation”—power precludes theory, too.

So, to return to the Heideggarian deployment of “the horizon,” what does AR entail regarding the social ontological structures responsible for normative breakdown or collapse, and therefore for loss and transformation? We should note that we have so far presented AR as a psychological, social-theoretic and linguistic account of the relationship between contestation and agency. And we have just said that the first two features of AR should make us wary of offering general, systematic claims about what sorts of things exist. But if we were to translate its commitments into a social-ontological idiom, we may say that the structures responsible for breakdown or collapse are pervasive, that they don’t just emerge or engage in circumstances of catastrophic failures of intelligibility, as with the demise of the Crow way of life discussed by Lear. Rather, they pervade and partially delimit the field of engagement between agents. One relevant consequence of this claim is that breakdown dynamics are not merely possible but, in an important sense, everywhere instantiated. We have already seen one way in which this can be shown when we reviewed Butler’s account of the performative and normative gaps or “faultlines” in speech seen as a ritual practice whose historicity exceeds but is exceeded by the speaking subjects whose agency it supports.
Now, it may be pedagogically expedient to dwell upon patent examples of breakdown, macro-level contexts of cultural crisis or radical and abrupt transition where many of the terms in which a practice or a way of life makes sense to those who live it out come under external threat more or less all at once. This is surely Lear’s expository strategy in *Radical Hope*. But the strategy can impart an excessively narrow conception of the range of phenomena worth characterizing in terms of breakdown. It tends to suggest that explanatory or prescriptive appeal to strategies of ‘radical’ agency are especially called for when an entire way of life, or some set of macro-level structures composing it, is crucially endangered, threatened with imminent collapse or pervasive transformation.

An important part of this task is developing a better vocabulary for capturing the pervasiveness and constructive potential of these dimensions of social ontology than the terms in which breakdown is commonly described. Social life and the normative resources comprising them are sometimes represented as a *fait accompli*, as obtaining in the *background*, or as supplying a fund of determinate meanings antecedently available to subjects. Placed in the foreground, subjects’ activities are supposed to be understood as taking place against the larger, horizontal order receding in all directions like a vast threshold embracing and enabling discrete action. Here is Charles Taylor, in “The Politics of Recognition:”

---

257 Alongside Lear’s work on “radical hope,” consider for instance Donald Davidson’s account of “radical interpretation” and Mark Norris Lance’s recent (May 2010) analysis of “radical change” and the role and import of retraining in the skills we deploy in interpersonal practices such as flirting that are structured by continuing legacies of oppression that partially corrupt the practical competences of participants. My reading of Davidson on this score relies on Rorty’s account of his view of language in *Contingency, Irony, and Solidarity* (Chapter 1, “The Contingency of Language”).
Consider what we mean by identity. It is who we are, ‘where we’re coming from.’ As such it is the background against which our tastes and desires and opinions and aspirations make sense.258

Here Taylor offers a concise statement of the ‘fixed background’ view of personal identity. It parallels his view in *The Ethics of Authenticity* of the possibility of social life as resting on pregiven horizons of meaning that function just as he says here “identity” does for individual selves, namely by supplying a stable of committed points of evaluative focus against which one may understand and negotiate the engaged flux of foreground activity. Taylor’s and Lear’s pictures are gainsaid by the extent to which differences from within and without the self compromise the fixity of these fixed points.

Without their presumptive fixity, one cannot appeal to them to ground an account of what makes agency as such possible. It would seem to follow, moreover, that one cannot theoretically apply the horizional strategy to address narrower normative questions without distorting the dynamic character of the phenomena at issue. To see how this can happen, consider Mark Lance’s recent work on rebuilding competences surrounding interpersonal transactions that have been corrupted by a legacy of oppression.

In “Walking the Walk and Talking the Talk,” Lance wants to shed light on what is at stake in the possibility of “rebuilding competences” regarding flirting in light of patriarchy’s continuing corrupting influence on the skills men develop and exercise for flirtatiously interacting with women. Here is Lance:

…the linguistic enters into the practice of flirting not merely through the fact that most flirting is talk, for an absolutely crucial dimension of flirting…lies in the ability of participants to recognize legitimate challenges to ongoing patterns of behavior, regardless of whether that behavior is itself linguistic…[T]he difference between decent and piggish behavior rests precisely on how smoothly one switches back and forth between this fundamental involvement in the practice of flirting, and the related critical mode of listening, attending to objections, seeing what the other person is getting at, noticing emotional uptake,

---

258 Taylor 1994, pp. 33-34
and modifying one’s behavior in light of all this. Morally competent flirting requires, that is, a smooth ability to move back and forth between flirting and discourse about flirting and to integrate the results of that discourse into the practice of flirting itself.\footnote{Lance 2010, from “§3: Structural aspects of power and oppression,” p. 7}

A few pages later, Lance draws the first of three general lessons:

…when we are facing a habituated inability to give skillful uptake to a genuine normative status, we are thrown out of the element of skillful involvement. We are faced with a type of Heideggerian breakdown in which we need to assess our practice from a detached and theoretical perspective. We must rebuild our competence on the basis of that perspective, in the face of, rather than using, the resources of our background contextual skill.\footnote{Ibid.}

I want to suggest that Lance’s account of what “rebuilding competences” involves is essentially right but (as it were) flirts with the misleading figure of the horizon we have ascribed to Taylor, Haugeland, et al. Consider Lance’s claim that properly rebuilding competences must be carried out from the detached, theoretical perspective. One wonders how this can be when, as he puts it, “the problem is not one of adopting the right principles…but rather one of a failure of perception that arises out of the complex interaction of our ongoing practical engagements with one another” (13); and, more concisely, that “[i]n none of this is the problem at the level of principle” (15). Unless the detached, theoretical perspective is one from which we can not only do more than assess the validity or contextual applicability of guiding principles but can specifically “retrain ourselves” in the way Lance thinks is sometimes necessary in cases of breakdown, we are going to have to rely upon some relevant “background contextual skill” (14) in the course of the sort of “rote practice” (12) involved in such retraining. After all, we may presume that the engaged interactive skill in need of revision, whatever it is, will be densely tied to a host of others with which it interrelates.
In the case of flirting, these include a score of contextually determinate interactive and linguistic skills involving body language, face reading, proper tone of voice, and so on, that are integrally tied to the narrower range of interactive skills under scrutiny. Those within the narrower range being reconstructed will themselves involve inflections or permutations of the broader interactive skills: we must talk with and respond to those whom we are retraining ourselves to better talk with and respond to in the specific way that is required, and we cannot avoid doing this all at the same time. It therefore seems too strong to claim that a general lesson of the discussion is that rebuilding competence is carried out on the basis of detached reflection “rather than [by] using the resources of [the relevant] background contextual skill.” This oversimplification finds expression again on p. 16: “Living, as we are, in a class-stratified, racist, sexist, imperialist and authoritarian culture, we simply bring no skills to living in an egalitarian, democratic, anti-racist, post-revolutionary world.” Surely the distinction between the world we now inhabit and the one towards which we may prefiguratively grope is not really so starkly cast, or else genuinely “radical” action seems to require bridging an unbridgeable gap.

We are now in a better position to see what it is that AR gives us reason to take issue with in the social-theoretic deployments of the horizonal strategy we have been discussing. The conception of social horizons of meaning variously endorsed by Pippin, Taylor, Haugeland and Lear embraces two related ideas.

First, theorists of the social horizon maintain that norms existing in the “background,” whether as patterned behavior or as general commitments or demands of some kind, are what make foreground activity normatively intelligible and tractable both transcendentally speaking (as a condition for the possibility of meaning) and in particular
cases where contextual judgments are required. So, in general and in cases, background norms condition the intelligibility of foreground activity.

Second, and as a result of the first claim, theorists of the social horizon maintain a functional dualism of normative intelligibility. “Intelligibility” is understood according to a social-ontological version of the post-Kantian Consensus Premise (CP) whereby legitimate authority derives from an ideally unanimous normative demand. We may call the social-horizonal version of CP the Intelligibility Premise (IP). IP holds that intelligibility is the constitutive paradigm for norm-governed agency.\(^{261}\) Though the paradigm may never be instantiated by any finite exercise of agency, it nevertheless furnishes the standard by which we judge their success or failure as such. Breakdown in the authority of norms is to be understood as a deviation from the paradigm case of “smooth” or well-functioning norms. This is meant as a claim about the relation of ontological priority between “intelligible,” or tractable and consensually robust, functioning and “breakdown,” or opaque and erratic functioning. IP therefore follows from the opposition between foreground and background phenomena and entails a strict privileging of fully realized normative intelligibility,

Butler’s work on the persistently equivocal functioning and non-sovereign exercise of power through linguistic performativity helps to clarify why accounts of social practice which rely upon IP ontologically misleading and politically near-sighted. Consider her claim in *Excitable Speech* that “if agency is not derived from the sovereignty of the speaker, then the force of the speech act is not sovereign force. The ‘force’ of the speech act is, however incongruously, related to the body whose force is

\(^{261}\) For neo-pragmatic theorists such as Haugeland and Lear, the paradigm of agency tends to be understood in terms of the ideal realization of some form of practical mastery involving an embodied skill such as the use of a concept by linguistic means in a social practice of holding responsible.
deflected and conveyed through speech. As excitable, such speech is at once the
deliberate and undeliberate effect of a speaker.” (39).

The idea is a familiar one by now. It is that because no speaker can ever exercise
sovereign power, no speech act is sovereign in its effects. One thing this precludes is any
strict or reliable proportionality between the efficacy of a speech act and the power of the
speaker who deploys it. Sometimes, for certain particular observational or rhetorical
purposes, one may assert that there is a generally reliable measure of defeasible parity of
power between a speaker’s social position and the force of a speech act she performs, but
Butler’s thesis is that, strictly speaking, there is always a gap between the two in
particular cases, one which other speakers can, with contextually varying degrees of
success, strategically exploit in various ways. Thus there is no “unbridgeable” gap
between the world we inhabit and the one we may vaguely envision, no primary
ontological divide between the paradigm deviant cases. The normative terrain is
pervasively fractured, shot through with gaps. If we really want to represent things
ontologically, we would do better to say that crisis is not a “radical” or deficient
condition but a persistent one with which speaking subjects are always confronted.

5: ambiguous horizons of action & identity: Butler on gay self-definition in
the United States military

In Chapter 3 of *Excitable Speech*, Judith Butler addresses questions about the
relationship between speech and power raised by discourse over the issue of gay self-
definition in the military. Both the Department of Defense regulations and the ensuing
public debate over their fairness and value proliferate the term meant to be constrained,
“homosexual,” by inscribing it in official speech and fueling discourse about policies affecting those who self-identify in its terms.\textsuperscript{262} This is because

the term ‘homosexual’ was [initially] disallowed as part of a self-ascription or self-definition on the part of the military personnel. The term itself was not banished, but only its utterance within the context of self-definition. The very regulation in question, must utter the term in order to perform the circumscription of its usage…Indeed, the regulations might be held accountable, paradoxically, for the apparent fact that the word has become more speakable rather than less.\textsuperscript{263}

Butler calls attention to some curious ambiguities associated with the status of the DOD regulations both on their own terms and considered as performative speech. The regulations in question treat speech as at the same time conduct when it involves homosexual self-definition. It does this in order to license prohibiting military personnel from self-identifying as homosexual:

…stating that one is a homosexual, that is, making reference to one’s status is reasonably construed as homosexual conduct itself. In the Department of Defense Policy, statements are themselves conduct: according to the more recent Congressional Statute, statements present evidence of a homosexual ‘propensity’ that poses an unacceptable risk for the military…[and] there may be other ‘signs’—affiliations, gestures, nuances, all of which equally point in the same direction.\textsuperscript{264}

This regulatory strategy thus ascribes “a natural teleology to homosexual status, whereby we are asked to understand such status as always almost culminating in an act.”\textsuperscript{265} The state’s speech, expressed in the policy, thus conjures a figure of homosexual utterance as invested with a radically efficacious performative power of self-assertion through self-definition. Of course the state, through its official regulatory speech, claims for itself the capacity to wield this sovereign power: it presumes to authoritatively govern homosexual action \textit{qua} superior power. But it aims to do so by officially constituting a

\textsuperscript{262} Butler observes the parity between her account and Foucault’s analysis of “the incitement to discourse” in his \textit{History of Sexuality, Vol. 1}, Part Two, Ch. 1, “The Incitement to Discourse.”
\textsuperscript{263} Butler 1997 (1), p. 104
\textsuperscript{264} \textit{Ibid.}, p. 106
\textsuperscript{265} \textit{Ibid.}
discursive space within which homosexuals are invested with one and the same power it
reserves for itself. So the official DOD policy collapses the distinction between
(homosexual) identity or “status” and (homosexual) speech, action or, in the language of
the policy, “conduct.”

What is the relevance of these considerations for our present task? Butler points
to it when she remarks that

It seems clear, as Janet Halley has shown, that arguments that seek to restrict the
prosecution of homosexuality to either status or conduct are bound to produce
ambiguities that threaten the coherence of either legal basis.\textsuperscript{266}

According to Butler, these “threatening ambiguities” pervade speech as a consequence of
the fact that it is persistently excitable in the sense reviewed in the previous chapter.

Speech is never fully governed or determined by the intentions which animate an
utterance. And this gives rise to a condition of incommensurability between intention
and utterance, utterance and action, and intention and action that Nietzsche perceived and
diagnosed as an integral component of Moral Personhood according to which “it is no
longer the consequences but the origin of an action that one allows to decide its value:”

To be sure, a calamitous new superstition, an odd narrowness of interpretation,
thus became dominant: the origin of an action was interpreted in the most
definite sense as origin in an intention; one came to agree that the value of an
action lay in the value of the intention. The intention as the whole origin and
prehistory of an action—almost to the present day this prejudice dominated
moral praise, blame, judgment, and philosophy on earth. But today…at least we
immoralists have the suspicion that the decisive value of an action lies precisely
in what is unintentional in it, while everything about it that is intentional,
everything about it that can be seen, known, ‘conscious,’ still belongs to its
surface and skin—which, like every skin, betrays something but conceals even
more. In short, we believe that the intention is merely a sign and symptom that
still requires interpretation—moreover, a sign that means too much and therefore,
taken by itself alone, almost nothing.\textsuperscript{267}

\textsuperscript{266} Butler 1997 (1), p. 106
\textsuperscript{267} Nietzsche, BGE Part Two: 32
Nietzsche’s insight and Butler’s more systematic analysis suggests that the notion of pregiven horizons of intelligibility is linguistically naïve. This is because the threatening ambiguity of speech casts doubt upon the effort to frame the possibility of norm-governed agency in terms of a fixed relation of priority between identity and action. After all, if a discourse of identity and action is going to help to clarify the operation of power in contemporary life, it cannot set one term over against the other univocally. The DOD and congressional policies to which Butler refers define homosexuality in terms of a readily discernible, irrepressible continuity between sexual orientation as a fact about one’s identity and the sort of expressive behavior in which homosexual action supposedly consists. Notice that this is a political problem, the problem of how homosexuals are defined and administered by public authorities, and it is indeed a problem of identity. I do not wish to deny that. After all, the regulations defining self-declaration as expressive conduct function in such a way that:

The term [“homosexual”] is to remain a term used to describe others, but the term is not to be used by those who might use it for the purposes of self-description; to describe oneself by the term is to be prohibited from its use, except in order to deny or qualify the description. The term ‘homosexual’ thus comes to describe a class of persons who are to remain prohibited from defining themselves; the term is to be attributed always from elsewhere…A homosexual is one whose definition is to be left to others…one whose self-denial is a prerequisite for military service.\(^{268}\)

However, this political problem of identity has its very origin in the juridical elision of the distinction between act and identity, conduct and status, by agencies of state power. Further, as we have already seen, Butler suggests that no credible solution is to be had by attempts to re-establish the distinction, for “arguments that seek to restrict the prosecution

\(^{268}\) Butler 1997 (1), p. 105

190
of homosexuality to either status or conduct are bound to produce ambiguities that threaten the coherence of either legal basis.”

So, the vocabulary of identity and action functions in a peculiar way in the context of the issue of gay self-declaration in the military. Its deployment embodies a strategy for establishing the relation between the two terms, namely by defining one’s status (as a homosexual), through the mechanism of overt declaration of that status, in terms of one’s propensity to act in ways expressive of it. And this strategy serves to undermine any univocal ordering of identity over against action. It poses the two dimensions of (homosexual) agency as incorrigibly continuous in order to justify policies that constitute spaces for self-identification while undermining the agency of those inhabiting them—as Butler observes, gays are to be defined precisely as those who may not define themselves.

The attempt by the state to authoritatively establish the background norms concerning homosexual identity and action, no less than the ensuing contestations over its efforts, therefore generates ambiguities that undermine the very effort. The case suggests that purportedly sovereign articulations of background norms founder on the fractures inherent in the speech on which they rely. In this case, the effort generates not only ambiguity and performative excess but a kind of discursive violence through which subjects are perversely assigned a status that is both impossibly godlike and unconscionably degraded.
6: love and loss

We will conclude our discussion of the conception of horizons compatible with AR by considering some salient aspects of loving attachments. For obvious reasons, our discussion of love will have to be radically provisional and incomplete, but even a cursory review of some commonly acknowledged general features of love will suggest why SI is poorly equipped to account for it without essentially denigrating it. The reason is simply that the constraints of love are deeply contingent and emphatically not self-imposed. The sort of erotic phenomena constituting love therefore flout the core volitional and normative requirements identified by SI as constitutive of agency as such. Advocates of SI are therefore committed to regarding love as an essentially deficient mode of erotic agency. And so apart from being psychologically and phenomenologically inarticulate with regard to love’s erotic prospects, SI’s diagnosis of love is ethically odious.

When one loves another, one precisely lacks the sort of reflectively guided, sovereign volitional wherewithal and self-possession which SI’s Deliberative Control Thesis paradigmatically requires of all rational agents. To love a person, an institution or an ideal is *inter alia* to open oneself to losing oneself. It is to be subject to a relation in which substantial elements of one’s sense of self and worth are bound up with the disposition and fate of the object of love, and no resort to formal procedures can suffice to resolve the questions which press hard upon the lover. It is tempting to say that, where human relationships of this sort are concerned, this much is common knowledge if anything is. Yet SI requires one to regard this condition as at best a deficient, severely compromised one for exercising agency in the highest and truest sense.
I think we should reject this attitude. It strikes me as not only tendentious but perverse, especially in light of how pervasive and integral to human life is the normative ambiguity, deliberative opacity and potential for thoroughgoing disorientation endemic to intimate relationships. Since the sovereign control SI requires of agents is perhaps most conspicuously absent within intimate relationships, as a systematic account of the conditions for the possibility of agency, we should be skeptical of SI’s claim to illuminate the dynamics of one of the most central domains of human life.

Now, in light of SI’s apparent difficulty illuminating the sphere of intimacy, it should come as no surprise that the normative practical principle central to SI has a dubious authority over our conduct within intimate relationships. Indeed, it is doubtful whether morality, in the sense represented by Korsgaard’s Kantian imperatives ostensibly identical with “reason’s own principle,” properly governs one’s conduct within them. Something along these lines is a familiar refrain among virtue theorists and ethical perfectionists within a broadly Aristotelian tradition. Nietzsche too recognized this:

Whatever is done from love always occurs beyond good and evil.269

Of course, proponents of SI will surely insist that, however limited may be its prescriptive scope within intimate relationships where partners must in a sense improvise as they go, morality nevertheless imposes fundamental constraints on the sort of conduct permissible within these or any other interpersonal relationships. This indeed follows the supreme authority over all norm-governed practice which Kantian morality claims for itself. In the previous chapter we called this the Moral Sovereignty Thesis (MST).270 But

---

269 Nietzsche, BGE Part Four:153
270 See Ch. III, Section 3.2.
once we cast doubt upon SI’s core psychic and volitional doctrines, namely the two elements of the Psychic Wholeness Thesis and the three claims composing the Deliberative Control Thesis, MST is called into question.

Here’s why. According to SI, the supreme authority of the categorical imperative derives from its being the source of the re-unification at which the inescapable task of practical deliberation necessarily aims. Through its assertion that reflective distance (RDT) and so the necessity and sufficiency of the will (NWT & SWT) are prerequisites for agency, DCT tells us that we can and must find sound reasons for thought and action. As a constitutive principle for volition, the categorical imperative is said to supply us with the standard we require for distinguishing legitimate practical reasons from bad ones and justifiable actions from ones that cannot be justified. Good actions are ones which unify our psychic lives and constitute us as agents; bad ones fail to do this. So the sovereignty, or supreme, universal authority, of morality rests on the sovereignty, or self-determination, of agents themselves. In short, impugning the doctrine of the will on which SI’s endorsement of MST relies is to impugn MST itself by undermining one of its primary bases within SI.271

SI tells us that our identity is determined by what we value as this is constituted and expressed through our deliberative and practical activity. Butler contends instead

---

271 The other pillar that operates in conjunction with DCT to support MST is the teleological metaphysics of action which Korsgaard introduces independently of the phenomenological case she makes for DCT. These are separate sources of support but neither entails MST without the other’s contribution. Since MST is underdetermined by a teleological metaphysics of action alone, which we called the Action Unity Thesis (AUT), the objection we are considering loses its force in the absence of compelling arguments for both DCT and AUT. Nevertheless, the third feature of the account of power defended in this chapter—its preclusion of “final horizons” of analysis—casts doubt on precisely the teleological rationale for AUT. To put the objection in terms of the account of integrity to be defended in the next chapter, the resort to final horizons of analysis involved in AUT represents a failure of integrity, a failure to take seriously the role of power in shaping and conditioning human agency, including the theoretical activity carried out by any philosopher.
that what we value is largely a function of our primary attachments and dependencies, which are robustly contingent and established through unchosen relations of subordination. She draws on Freud, Nietzsche and Foucault to argue that loss and ambivalence enter into the very formation of subjects and condition the agency of which they are capable. This suggests a methodological parity between AR and horizontal views, particularly Sovereign Idealism: the condition figured as psychically primordial is held up as erotically ideal or ethically paradigmatic. For horizonal theory, it is a condition of primal psychic wholeness, as we saw in our discussion of Frankfurt’s ideal of wholeheartedness, where it restores the psychic unity compromised by psychic fragmentation. Curiously, we saw that Frankfurt maintains that psychic fragmentation constitutes the subject by affording its nascent power of reflexivity with a range of possible objects of identification. Butler follows Nietzsche and Freud in asserting something similar, though she embraces the implications for the lives of subjects so formed: their agency remains persistently but ambiguously tethered to the prior mechanisms by which they are formed.

Consider what Butler means when she claims that developing capacities for forming particular erotic attachments involves foreclosing certain sorts of attachments. The idea is that there is a sense of loss that precedes subjects capable of attachment by shaping their capacities in determinate ways, disposing them to acknowledge certain desires and foreclose others:

Freud distinguishes between repression and foreclosure, suggesting that a repressed desire might once have lived apart from its prohibition, but that foreclosed desire is rigorously barred, constituting the subject through a certain kind of preemptive loss. Elsewhere I have suggested that the foreclosure of homosexuality appears to be foundational to a certain heterosexual version of the subject. The formula ‘I have never loved’ someone of similar gender and ‘I have never lost’ any such person predicates the ‘I’ on the ‘never-never’ of that love
and loss. Indeed, the ontological accomplishment of heterosexual ‘being’ is traced to this double negation, which forms its constitutive melancholia, an emphatic and irreversible loss that forms the tenuous basis of that ‘being.’

It therefore makes sense within AR to refer to erotic horizons, but not as stable or paradigmatically seamless, unified or whole. Horizons of intelligibility are not paradigmatically intelligible; they are paradigmatically fractured due to the dynamic and excessive character of speech, and they are paradigmatically unchosen due to the foreclosure that enters into the formation subjects’ erotic capacities. Since this is what explains the fact that erotic “horizons” are persistently subject to crisis and collapse, it shows why they are always in a precarious state requiring their continual renegotiation. Thus the conditions of that Lear, Haugeland et al. represent as extreme are instead in a way routine.

I acknowledge that the claim that crisis is paradigmatic sounds paradoxical. But AR is not the only perspective on questions of value and agency that has this sort of implication. The German political theorist Carl Schmitt adopts a similar approach when he represents politics as fundamentally determined by relations of existential antagonism between friend and enemy, where the latter is, “in a specially intense way, existentially something different and alien, so that in the extreme case conflicts with him are possible.” Whatever is to be made of Schmitt’s account, his strategy is to understand politics from the point of view of what appears to be the extreme case, guided by the thought that doing so reveals something of enduring importance about the conditions of human life. AR’s treatment of erotic horizons as paradigmatically fractured and persistently in-crisis is offered in the same spirit.

272 Butler 1997 (2), p. 23
273 Schmitt 1996, p. 27. Thanks to Gerald Mara for calling my attention to this parallel.
Recall that, in his discussion of horizons of intelligibility, Charles Taylor denies that “people could determine what is significant, either by decision, or perhaps unwittingly and unwillingly by just feeling that way.”\textsuperscript{274} Though he erroneously infers that the question of what can matter to people is settled by paradigmatically stable, pregiven social horizons of meaning, we now see what truth there is in his assertion. Also recall our discussion in Chapter I of Frankfurt’s account of caring, particularly his discussion of “volitional necessities,” the deeply orienting and motivating erotic attachments that define what we are willing to do by determining what we are subjectively incapable of bringing ourselves to do. Volitional necessities supposedly help constitute practical rationality itself by making certain courses of action otherwise available to an agent nevertheless subjectively “unthinkable” for her. In a way, then, Frankfurt’s own version of the horizontal strategy \textit{does} entail that people (indeed individuals) can determine what matters to them “unwittingly and unwillingly by just feeling that way.” And to the extent that it involves the tacit ascription of sovereignty we identified in Chapter I, Frankfurt’s account tacitly denies (or denigrates as deficient) the unchosen precarity of erotic ties. If love is an extreme case of erotic attachment, then it does seem that dwelling on what appear to be extremes can show us something important about the “routine” conditions of human life—by undermining the rhetorically or pedagogically strict distinction between them.

In short, AR licenses appeal to horizons as substantially unchosen and precarious, not fundamentally self-imposed and inescapable. It underscores vulnerability, not sovereignty. And we shall see in the next chapter that it supports a celebration of the prospects of overcoming held out by experiences of loss and transformation brought on

\textsuperscript{274} Taylor 1991, p. 36
by encountering difference. Prospects of overcoming and self-overcoming thus take on a
greater potential allure than those of invulnerable self-mastery and wholeness so rigidly
privileged by SI. Nietzsche again suggests an apt erotic phenomenology:

*Love and duality.* -- What is love but understanding and rejoicing at the fact
that another lives, feels and acts in a way different from and opposite to ours? If
love is to bridge these antitheses through joy it may not deny or seek to abolish
them. -- Even self-love presupposes an unblendable duality (or multiplicity) in
one person. 275

7: conclusion: paradigmatic crisis

Now that we have presented AR as an alternate account of the psychological and
normative character of erotic agency, we can expect advocates of SI to say something of
the following sort:

“Your genealogical account of subject formation may well be right, so far as it
goes. This is essentially an empirical question best left to the social sciences. However
that turns out, it is does nothing to address the first-order normative question of how
human life, whether psychically or institutionally speaking, *ought* to be structured or
governed. After all, claims with normative content cannot be derived from observations
of fact, such as those pertaining to human developmental psychology, without further
ado.

Our view, Sovereign Idealism, is a theory of *rational psychology* and so is
robustly normative—the role of the normative is built right into its characterization of
human powers and human experience from the very outset. Our core claim about the
character of the human psyche is that insofar as it is capable of normative self-
government, it is constitutively, or by virtue of its reflectively self-conscious structure,

275 Nietzsche, HATH: II, Part One: 75
subject to standards of efficacy and autonomy. Because this is a purely formal claim about fundamental requirements, the question of its validity transcends essentially causal stories about the material constitution of the psyche.

Why should one take this approach? The answer is that the best phenomenology of deliberation bears out a vocabulary of agency that requires this formal conception of the psyche. In a word, the assumption that we are free to decide what to think and to do, that it is up to us in the end to evaluate the normative status of our reasons for belief and action, is indispensable because the activity of seeking out such reasons is inescapable for us. Because we must regard ourselves as self-governing, we must regard ourselves as subject to the basic requirements for governing oneself successfully. It turns out that these correspond to the Kantian categorical and hypothetical imperatives, so morality is vindicated after all. Power may play a large role in determining how things are, but we need moral standards if we are to properly determine how they should be.”

The proponent of SI asserts a firm boundary between morality as the deep, necessary source of agency and power as the contingent source of the material circumstances over which morality has proper authority. What can we say in response? Our discussion of Butler’s conception of the “excitable,” or persistently out-of-control, character of the agency exercised through speech provides us with the resources we need to meet this challenge. For at the end of the previous chapter we saw that the phenomenological case needed to defend SI’s claim that we cannot do without its vocabulary of first-personal deliberative control itself implicitly rests on a theory of discourse, a supposedly Wittgensteinian account of agency exercised through speech. Korsgaard has relied on some version of this view for some time in her efforts to
vindicate the universal scope of the authority of substantive, Kantian morality. It is a view according to which the publicity of reasons (and meanings) makes them essentially interpersonally shareable due to the default “intelligibility,” or smooth and transparent semantic and performative operation, of the languages used to articulate and give them uptake. And in the past two chapters, we have seen how Butler’s analysis of some public contestations in and through speech helps show that advocates of SI are right to insist that engagement with norms is essentially public but wrong to insist that participants in practices where they are at stake have to take themselves to be subject to a requirement of mastery and subjective wholeness.

AR entails that to say that human beings are subjects means that they are capable of wielding normative power, but this in turn means that they are subject to powers over which they lack control. SI recognizes that subjects are subject to norms, but insists that this is so—that normative demands are legitimately authoritative—only inssofar as they are ultimately self-imposed. It therefore takes back with one hand the concession to their material subjection that it extends with the other. It maintains instead that in the sense required for an understanding of what makes the capacity for responsible agency as such possible, subject formation is formal. It is accomplished by an individual’s subjection to a necessary a priori constraint that is already an expression of her own identity as a rational being, not by her subordination to contingent regulatory powers that she neither chooses nor masters. The message of SI is that agents are active “all the way down,” at the end of the day, since even their (legitimate) passiveness is paradigmatically chosen. Our discussion of AR has shown why this is psychologically naïve and ontologically
confused. This has helped prepare the way for an alternative erotic ideal of integrity. In
the next chapter, we finally take up the task of articulating its essential features.

We saw in Chapter I that erotic attachments, whether “deep” or “thin,” are those
subjective ties that normatively orient and psychologically motivate subjects. By denying
that human agents, \textit{qua} speaking subjects formed through disciplinary power, are
sovereigns in the sense required by SI, \textit{AR} avoids the mistakes made by both Taylor’s
unequivocal denial that individuals can determine what matters and Frankfurt’s insistence
that they can. This is because, for both Taylor and Frankfurt, as well as for the
Heideggarian social theorists we have just encountered, it is true as a matter of
ontological or psychological necessity that “horizons are given.” And this is so despite
the disagreement that exists within the horizontal camp about the precise nature and
source of the content and authority of horizons of significance.

Here is how things shake out, given what we have seen. For Pippin, Taylor,
Haugeland and Lear, normative standards governing action are supplied by the rich social
practices of \textit{inter alia} holding responsible that all agents (or “persons,” for Haugeland) as
such inhabit. For these folks, \textit{identity} in the broad sense denoting antecedent
commitments fundamentally conditions normative practice as such. Norms must
“already be” in force and have determinate content in order for it to make sense to take
one to be guided by them. This “already be” is said to be matter of strict ontological (not
historical) priority.

For Korsgaard, on the other hand, horizons of value correspond to formal rational
requirements and they specify constraints that constitute and govern action as such.

Because for Korsgaard there is really only one necessary horizon of meaning where

\textsuperscript{276} Taylor 1991, p. 39
normativity and agency is concerned, the analytical horizons she appeals to while arguing for SI are said to be the very same which govern non-theoretical activity. According to SI, *action*—that is, the formal constraints that fundamentally distinguish action from non-action—fundamentally conditions normative practice as such. The standards that define and govern action as such must “already be” in force and have determinate content in order for it to make sense for one’s movements to be recognizable as the responsible action of a single agent. This “already be” is said to be matter of strict ontological (not empirically psychological) priority.

Lastly, Frankfurt’s view does not fit neatly within either the Heideggerian “identity first” or SI’s “action first” camp. His is an “identity first” view but is methodologically individualist and, as we argued in Chapter I, essentially committed to SI’s brand of voluntarism to the extent that it relies on an ideal ascription of sovereign powers to individual subjects. Given these qualifications, Frankfurt’s story reads much like Korsgaard’s.

Given all this, consider the following remark made by John Wild in his forward to Merleau-Ponty’s *The Structure of Behavior*:

> Our worlds of meaning . . . are constantly falling apart; our unity is never automatically guaranteed. We maintain some modicum of human integrity only at the cost of constant struggle and reinterpretation.277

Now, the first part of Wild’s remark is apt and concisely states the main thrust of the case we have made against the figure of the ‘horizon’ as it is often deployed by social philosophers like Taylor and Haugeland.278 But the second part of Wild’s remark, where

---

277 Merleau-Ponty 2008, xvi. Thanks to James Olsen for making me aware of this provocative remark.
278 The central claim of this chapter, the idea that our “worlds of meaning” are in a constant state of crisis, bears a certain resemblance to a thesis of Giorgio Agamben’s referred to by Judith Butler in *Excitable Speech*, namely that “the state has become a protracted ‘state of emergency’” (Butler 1997 (1), p. 105).
he invokes the language of cost, is ambiguous. Surely he is right to suggest that struggling to reinvent ourselves and our worlds of meaning can be costly, but in what sense of ‘cost’ is this true? This question depends upon another: if our integrity is deeply connected to our struggles for reinventing the meanings in whose terms we live, what is the nature of this connection? Must the ambivalence, disorientation and other symptoms of struggling to reinvent ourselves be seen as obstacles to integrity? Wild’s remark that unity is never automatic seems to imply an identification of integrity as an achievement or restoration of unity, and that integrity so construed is threatened by the collapse of our worlds of meaning. On such a view, the costs of struggle are ultimately to be reckoned as losses.

Recall that advocates of Sovereign Idealism also maintain that integrity is hard-won through struggle. Indeed, Wild’s remark splits the difference almost exactly between the two rival views of personal integrity between which our inquiry adjudicates. In Chapter II we saw that Korsgaard characterizes the view she is offering as one about the nature of the felt constraint of normative demands, insofar as they operate as a psychological force within us, which she follows Kant in calling “necessitation.” She distinguishes the sort of view of necessitation she is offering from other influential ones in modern moral psychology by insisting that our experience of necessitation shows that struggle and effort is a crucial feature of our psychic lives as agents. But Korsgaard’s view is that what an agent must take herself to be struggling for in her efforts to live up to the normative demands to which she is subject is the psychic unity necessary to render her whole and so a fit locus of first-personal, and a coherent object of second-personal, responsibility:

279 Cf. Sartre, Nietzsche et al. on “nausea.”
There is work and effort—a kind of struggle—involved in the moral life, and those who struggle successfully are the ones whom we call ‘rational’ or ‘good.’ But it is not the struggle to be rational or to be good. It is, instead, the ongoing struggle for integrity, the struggle for psychic unity, the struggle to be, in the face of psychic complexity, a single unified agent.  

So Sovereign Idealism identifies an agent’s integrity with her psychic unity, and sees unification as something achieved through a sort of constant struggle—not the struggle to reinvent or perpetually re-constitute herself, but the struggle to constitute herself as “single unified agent.” The erotic telos of agency is thus integrity as something to be maintained through the struggle against the disunity arising from multiplicity within the self. So, apart from his remark about the way in which our worlds of meaning are constantly in crisis, which may be too strong for Korsgaard, Wild’s remarks suggest a view of integrity much like the one she offers.

So in what sense is Wild’s remark about winning integrity through constant struggle for reinvention ambiguous? In our final chapter, we will see that there is another sense of integrity that is strongly tied to the struggle for reinventing the meanings in whose terms we live but just as strongly opposed to the Sovereign Idealist conception of integrity as unity. For it is a sense of integrity defensible in terms of AR, one in which conflict and contestation play a crucial role, not as sources of obstacles to be overcome in the name of the horizon of one’s identity but instead as sites for the overcoming of those very horizons. From the point of view of this conception of integrity, self-overcoming is the guiding normative priority, not ‘self-determination’ in the Sovereign Idealist sense Korsgaard just defined above, namely as the establishment of stable boundaries for the self. Prospects for integrity construed agonistically are to be sought not in the overcoming of conflict in order to become a unified agent with a single identity, but in

---

280 Korsgaard 2009, p. 7
overcoming oneself—self-overcoming—by challenging the appearance of one as one
self: that is, in overcoming socially or psychically enforced semblances of self-unity in
order to cultivate opportunities to excel at contestation itself. ‘Agonistic integrity’ would
thus be not an ideal of self-legislation and obedience but of self-overcoming and
excellence in psychic and normative “warfare:”

Unconcerned, mocking, violent—thus wisdom wants us; she is a woman and
always loves only a warrior.  

281 Nietzsche, Z: “On Reading and Writing”
CHAPTER FIVE

“BEYOND YOUR HORIZON”:

TOWARDS AN EROTIC PSYCHOLOGY OF AGONISTIC INTEGRITY

Above all, one should not want to divest existence of its rich ambiguity; that is a dictate of good taste, gentlemen, the taste of reverence for everything that lies beyond your horizon. 282

1: introduction: integrity’s renewed prospects

In the last two chapters, we presented an account of power as a productive, excessive and dynamic range of effects and relations that enter into the formation of human subjects and so condition our capacities as agents. We called this view of subject formation and agency Agonistic Realism (AR) and contrasted it with Korsgaard’s Sovereign Idealism (SI). Both are erotic ideals in the sense introduced in Chapter I—conceptions of how we ought to live out the deeply felt attachments, or erotic ties, which motivate and orient our conduct. But we saw that AR is distinguished by its emphasis upon the integral role played by contestation in the dynamic psychic and social life of agents. It is worth recalling why we have spent so much time contrasting SI and AR as rival accounts of subject formation and agency.

We concluded our first chapter with three related hypotheses: first, that erotic ideals that presuppose the figure of the horizon as a pre-given or “background” source of

282 Nietzsche, GS Book Five: 373, “‘Science’ as a prejudice.—“
normative authority, and so tend to construe integrity as steadfast adherence to demands already established, tend to elide and undervalue the role of loss and transformation in erotic life; second, that recognizing the role and value of loss and transformation is a condition of adequacy for an erotic ideal; and third, that an erotic ideal grounded in an account of power’s integral role in subject formation and agency would be able to better satisfy this condition. Whereas we could only initially motivate the first two hypotheses in Chapter I, it was our task in the following three chapters to more fully vindicate them. With AR in hand as an account of power’s integral role in subjection and agency, we may now finally attempt to vindicate our third hypothesis that such a view can ground a compelling account of integrity by folding its recognition of the role and value of loss and transformation, and of the contestation that makes it possible, into that account. We will see that we were right to suspect in Chapter I that there is a compelling way of understanding how to live out one’s commitments in an agonistic spirit of “serious play,” and that Nietzsche’s rich conception of the “free-spirit” supplies resources for provisionally articulating what this involves.

In short, in this chapter we finally return to the issue with which we began in Chapter I, how to understand integrity as an erotic ideal, but we will approach it from the direction of the agonistic conception of subjection and agency that our account of power has made available. In the intervening chapters, we followed Charles Taylor, John Haugeland, Christine Korsgaard and others in regarding the problem of integrity as best understood as one of social ontology and moral psychology. Now we return to the problem of integrity but reframe it in a way suggested by the conception of power underlying AR. We will cast it explicitly as a problem of political psychology.
understood as encompassing efforts to identify and articulate the normative and psychological resources available to human agents who are engaged in *concrete practices of contestation* over the meanings and values at stake in their ethical and political lives.

This way of understanding what ‘political psychology’ is all about is indebted to the conception of it advanced by Robert Pippin in his recent literary and philosophical analysis of the dialectical relationship existing between seminal American Western films and popular political mythology.\(^{283}\) There Pippin addresses his findings to participants within a specific, historically conditioned political society, namely American popular political culture in the 20\(^{th}\) century. He wants to enrich our society’s political self-understanding by exploring how we as actors, in light of our historically conditioned cognitive and affective dispositions, accommodate ourselves to the social and political institutions in whose midst we are situated. To what extent the approach to political psychology pursued in this chapter significantly departs in spirit from the approach found in Pippin’s recent work is not especially important for present purposes. We simply note that our attempt to think through the normative and psychological contours of erotic agency brings dynamics of contestation into center stage.

In the previous chapter, we revisited the idea of “shared horizons” of meaning and analysis and found that it runs afoul of the priority assigned to contestation within AR. We thus challenged social theoretic assertions of ‘horizontality’ in the way that we criticized Korsgaard’s conception of subject constitution and agency in Chapter III. Having thus called into question a range of horizontal views in light of the significance of contestation, we will wrap things up in this chapter by sketching an alternative political psychology of integrity as a political-ethical ideal for human agents as we now

\(^{283}\) Pippin 2010
understand them. This broadly Nietzschean conception of integrity will constitute a suitably agonistic erotic ideal for actors engaged in contestations amidst the tensioned play of differences which AR tells us constitutes the dynamic psychic and social life of speaking subjects.

2: beyond one’s horizons: integrity as an agonistic ideal

In the introductory chapter of *Self-Constitution*, Korsgaard partially frames her theory of agency as an explanation of the operation of normativity within us as a felt psychological force constraining our behavior. Following Kant, she calls the phenomenon to be explained “necessitation.” She argues that what distinguishes her broadly Kantian view from “dogmatic rationalist” and Humean sentimentalist theories is its insistence that the role played by necessitation in our psychic lives means that:

There is work and effort—a kind of struggle—involved in the moral life, and those who struggle successfully are the ones whom we call ‘rational’ or ‘good.’ But it is not the struggle to be rational or to be good. It is, instead, the ongoing struggle for integrity, the struggle for psychic unity, the struggle to be, in the face of psychic complexity, a single unified agent...The work of achieving psychic unity, the work that we experience as necessitation, is what I am going to call *self-constitution*.”

As we saw at the end of the previous chapter, this remark aligns Korsgaard’s Sovereign Idealist conception of integrity with our agonistic one in one important respect and distances it from it in another. On the one hand, we also hold that prospects for integrity are to be found in “a kind of struggle” in relation to the constellation of drives, perceptions, performances, ideals, attitudes and affects making up the psychic and social relations in which agents are implicated. But on the other hand, the struggle for integrity

---

284 Korsgaard 2009, p. 3
in our sense is not anyone’s struggle for psychic unity, “the struggle to be, in the face of psychic complexity, a single unified agent.” The reason is that our defense of AR has shown that the intuitive conception of psychic wholeness underlying SI’s appeals to the necessity of self-unification is psychologically and linguistically naïve.

This matters because AR provides the psychological rationale for the view of integrity to be defended below, just as SI’s metaphysics and moral psychology of action underwrite its conception of integrity as the struggle to unify the self. We are not offering a general theory of the way that normative requirements make their demands felt within agents through the operation of psychological mechanisms of enforcement. But as an account of power’s role in subjection and agency, AR has implications for the question of how to interpret the significance of the experience of “necessitation” in Korsgaard’s sense.

Recall that, by Korsgaard’s lights, the paradigm experience of necessitation occurs in the presence of internal psychic conflict where one feels pulled in different directions. One feels the constraint of a norm most saliently when one is subject to conflicting motives regarding some willed end or proposed course of action. This much of the phenomenological story is faithful to the experience of internal conflict. But we should question SI’s view of the ethical significance of this experience. For since the possibility of experiencing internal conflict derives from what Korsgaard calls the fragmentation of the psyche, SI’s identification of unity as the constitutive aim of human agency commits it to regarding conflict as essentially threatening, as endangering the active self’s prospects for reunification through practical deliberation. And this constitutes a denigration of conflict as a defining feature of our psychic lives. That is,
SI’s representation of agency, precisely by virtue of eliding the formative role of contestation, entails an underestimation of its potential value.

Or so we will argue. In Chapter III, we explained how power plays a crucial formative role in human identity and action. The time has come to offer an account of the potential value of the contestation which power makes possible. In what follows we will see that much of Nietzsche’s mature work suggests a compelling way of articulating a conception of integrity as a suitably agonistic ideal.

3: against one self: the soul’s plurality

In Chapter III, we drew upon the work of Nietzsche and Judith Butler in order to challenge SI’s representation of the self as paradigmatically whole and its connected claim that pursuing self-unification is unconditionally required of every agent. In Friedrich Nietzsche and the Politics of the Soul, Leslie Paul Thiele offers a commentary on aspects of Nietzsche’s thought that relates to these concerns. His view is that Nietzsche held “the primary task of life…to be the heroic struggle of individuation,” or one’s hard-won growth into oneself amidst the multiple personae and conflicting drives within one’s psychic economy. He sees Nietzsche as attempting a revival of heroic ethics for a godless, post-metaphysical age by pointing the way towards the overcoming of nihilism through the struggle of individuation, or soul-integration, aimed at cultivating and maintaining a “well-ordered soul.”

Since Thiele regards the core Nietzschean ethical ideal to be one of self-cultivation, he contests a strong deconstructionist reading of Nietzsche’s psychology of

\[286\] Thiele 1990, p. 3
the self according to which there is no worthwhile sense in which it can be one or whole, and wherein Nietzsche champions the dispersal of the subject rather than its unification. So far this all sounds compatible in principle with Sovereign Idealism. But Thiele explains why it is not when he insists at the outset that Nietzsche offers no theory of the individual’s formation and ethical value:

This word [‘theory’] connotes a synthetic systematization that is foreign to Nietzsche’s style and antithetical to his understanding of the individual. A theory of individualism, in this sense, is oxymoronic. The individual is precisely that for which no general formulas are applicable. Politics, on the other hand, conveys most of the meanings inherent in the development of individuality: the struggle, the ambiguity and ambivalence, the will to power, the compromises and coalitions, domination and rule, plurality and rank, the search for organic unity.  

What this means for us is that, by Thiele’s lights, whether or not Nietzsche is up to political psychology in our sense, his is a **politicized** psychology of the self and its ethical prospects; it is a “spiritual politics” where “the role of the individual in politics should be subservient to the role of politics within the individual.” Since our defense of Agonistic Realism brought power to the fore by treating contestation as integral to the formation and agency of subjects, the political psychology we now have in hand is politicized in this sense. Nevertheless, Thiele’s talk of “organic unity” raises the question whether his view of Nietzsche’s psychology and ethics comes closer to Sovereign Idealism than AR allows. There is reason to doubt that this is so, for as Thiele explains:

Nietzsche also held the individual to be an abstraction. He rejected the assumption that there was something called ‘the individual’ which had a certain fixity, continuity, and duration. Nietzsche stepped beyond subjectivism to attack the idea of an enduring, unified subject. The individual was conceived as a multiplicity. If one may speak of the uniqueness of the individual, it is because

---

287 Thiele 1990, pp. 3-4  
288 *Ibid.*, p. 4
of the unique composition of drives, the individual’s particular yet pluralistic internal regime.

Nevertheless, we will lay this question aside for now and focus instead on how Thiele’s interpretation of Nietzsche’s ethical project helps us defend a conception of agonistic integrity under AR.

Thiele contends that Nietzsche represents the soul as a plurality, or as William Connolly has put it when discussing the advantages of “‘post-Nietzschean’ ethical sensibility,” “a microsocial structure of voices, replete with foreign relations, implicated in complex relations with the macrosocial structure in which it participates.” On this view the psyche is a tensioned manifold of drives contending for supremacy, a “chaos” out of which the task of self-cultivation is to construct a “cosmos,” or a unique and well-ordered internal regime. The idea that the psyche is a manifold of competing drives recalls our discussion of Nietzsche’s and Butler’s genealogical arguments for seeing power as productive of normatively salient differences through subjection. But what is new about Thiele’s exposition of this familiar idea? His discussion of Nietzsche’s view of the structure of the psyche is not so much new as especially useful for our purposes. The reason is that it sets out ideas we have encountered before in terms that translate effectively into a statement of an agonistic erotic ideal.

4: against oneself: agonistic self-cultivation as an erotic ideal

Thiele’s exposition of Nietzsche’s conception of the figure of the hero is a good place to start indicating what “agonistic integrity” would involve, psychologically and

---

289 Thiele 1990, p. 37. Nevertheless, Thiele later asserts that, for Nietzsche, “The heroic task, assumed only by the few, is to become a sovereign individual” (p. 45).

normatively speaking. Thiele claims Nietzsche’s ethical ideal is embodied by a “higher” type of individual who achieves self-mastery by embracing the prospects for transformation and overcoming inherent in competitive and adversarial relations.

Nietzsche thus regarded the figure of the hero as “the agonal spirit incarnate:”

The hero above all is a lover of life…Competition and battle are not simply the means to winning glory, but constitute ends in themselves. For there is a joy to be found in the midst of struggle which surpasses that of the subsequent accolades…The hero is he who courageously struggles for preeminence. Arete is to be displayed; struggle is the means of its attainment. Only the love of struggle provides the stimulus for self-overcoming, the drive to reach beyond oneself so as to achieve excellence.291

The projection of the hero as both an antinomian and a creator of new regimes is not so much a contradiction as a recognition of his archetypical vocation. The feat of founding a state, culture, or folk is subsequent to warfare, conquest, and destruction. The great man remains essentially a creative force. But before his genius may manifest itself in an enduring work, the existing structures—be they constitutive of an aesthetic, ethical, or political regime—must be demolished and the rubble cleared…The Nietzschean hero, like his classical counterpart, is a breaker of taboos and custom…He blasphemes, he violates, and he destroys. Out of the ashes emerges a new regime.292

So Nietzsche’s ethical ideal is anchored to “the spirit of the agon,” that is, deeply tied to prospects for overcoming and self-overcoming through struggle, conquest, and feats of strength amidst adversarial relations. Moreover, by Nietzsche’s lights, certain forms of struggle have an intrinsic ethical significance. It is one’s performance within such contests of strength that determines one’s “order of rank” among adversaries. We want to develop this insight into a conception of integrity that is on all fours with AR. Since such a view of integrity would rest on an account of subject formation and agency that gives contestation pride of place, it should come as no surprise that it privileges contestation itself. But what sort of struggle does the struggle for integrity involve, if we

291 Thiele 1990, p. 13
292 Ibid., p. 17
are in accord with Sovereign Idealism at least in thinking that integrity is to be seen as something achieved through sustained work and effort?

Thiele and many other commentators realize that any recognizably post-Nietzschean ethic of overcoming will be one of self-cultivation set out in terms of a substantive conception of what makes for a well-lived life. Nietzsche’s ethical writings operate at a number of levels.  For our purposes, we note their resemblance to Aristotle’s in that they do not advance universal prescriptions but instead provide rich descriptive portraits of privileged types by exhibiting what is supposed to be attractive about their characteristic styles of activity. There is one type celebrated in Nietzsche’s mature writings that sheds a great deal of light on the struggle for agonistic integrity: “the free spirit.” Nietzsche portrays the free spirit as a rarefied achievement embodying salient aspects of the “highest” or most valuable mode of existence.

Nietzsche’s figure of the free spirit is a resource for seeing integrity as won not through the struggle to consolidate oneself as a singular, psychically whole locus of univocal responsibility, but rather, at least in part, by struggling against one’s own constitution under any single normative aspect. Such an undertaking is directly at variance with the metaphysical and ethical telos of rational agency as construed by SI: “The task of self-constitution involves finding some roles and fulfilling them with integrity and dedication. It also involves integrating those roles into a single identity, into a coherent life.”

Agonistic integrity would instead involve striving to overcome the contingent horizons of one’s own identity by challenging one’s commitments through transformative encounters with differences within and without one’s psyche. The aim is

---

293 Thiele’s stated aim, for instance, is to reckon with the didactic implications of the fact that Nietzsche tells us that his books are written in his own blood.
294 Korsgaard 2009, p. 25
therefore not to overcome the contingency of one’s erotic ties for the sake of universally necessary demands, but to give one’s sense of self over to the play of differences through which one’s erotic life remains dynamically alive.

Moreover, the adversarial relations one inhabits in order to secure one’s integrity are to be found at least in part within one’s own divided mind. If Nietzsche exalts the heroic type who craves a worthy adversary,\textsuperscript{295} then the type we are interested in, the agent with integrity, strives to be worthy of the rival she finds in herself. This is part of what is suggested Nietzsche’s famous injunction to “live dangerously:”

For believe me: the secret for harvesting from existence the greatest fruitfulness and the greatest enjoyment is—to live dangerously! Build your cities on the slopes of Vesuvius! Send your ships into uncharted seas! Live at war with your peers and yourselves!\textsuperscript{296}

But what is Nietzsche really on about here? What could it mean to live at war with oneself?

In his essay, “A Critique of Utilitarianism,” Bernard Williams argues that a person’s integrity depends upon her ability to make and maintain commitments by pursuing the projects that flow out of them, where such commitments are personal in the sense of not being shared by everyone.\textsuperscript{297} Thus someone’s commitments compose her “practical identity” in Korsgaard’s sense: a conception “under which you value yourself and find your life worth living and your actions to be worth undertaking.”\textsuperscript{298} Indeed, Williams’s language suggests that he believes taking commitments seriously requires adherence to normative demands that are already determined by the relevant “projects and attitudes which in some cases [an agent] takes seriously at the deepest level, as what

\textsuperscript{295} Nietzsche, GM I: 10
\textsuperscript{296} Nietzsche, GS Book Four: 283
\textsuperscript{297} Williams 1973
\textsuperscript{298} Korsgaard 2009, p. 20
his life is about.” This aligns his view with the “identity first” version of the horizontal view that integrity is all about conformity with antecedent norms.

Naturally advocates of the horizontal strategy will be the first to admit that any particular agent will value herself under more than one aspect. She may find that the plight of the poor, or the sanctity of the natural world, impose certain normative demands on her and so she will care about being a charitable or ecologically responsible person. To the extent that her concerns about alleviating poverty or preserving the environment are genuine, we say that she is committed to being such a person in practice. And we identify her commitment in terms of her convictions, her practical sense of what is worthwhile and why. Convictions thus express the commitments that make up our practical identities. They orient our conduct by making certain available acts or policies and not others show up to us as tempting, distasteful, permissible, required, forbidden, and so on. As we saw in Chapter I, the normative demands associated with our convictions are typically seen as already established for the purposes of deliberation on any particular occasion. And where they are not, it is commonly supposed that one is to work out what they entail in such a case by appeal to higher-order principles whose normative requirements have the specificity contextually lacking in the first-order conviction. Framed this way, the role of a conviction captures much of what SI entails about how our erotic lives ought to go.\footnote{Williams 1973, p. 116}

\footnote{Of course, there is at least one major qualification, namely that according to SI, all rational agents share the commitment to be moral whether or not they expressly acknowledge that they are so committed. In Chapter II we called this aspect of SI the Moral Sovereignty Thesis and saw that it rests on the claim that agents as such must take themselves to be subject to the requirements of the constitutive principles of volition which make their agency possible in the first place.}
With these ideas in mind, we may return to Nietzsche’s injunction to live dangerously by waging war against one’s peers and oneself. We were wondering what we should take this to mean, ethically speaking. Nietzsche has a number of sayings about the appropriate way of relating to one’s own convictions and they suggest a very different conception of their role in a well-governed life than SI’s:

A very popular error: having the courage of one’s convictions; rather it is a matter of having the courage for an attack on one’s convictions!!!

Enemies of truth. – Convictions are more dangerous enemies of truth than lies.

Convictions are prisons.

Nietzsche appears to be suggesting that, understood in a certain way, personal commitments can endanger the self. This is not the relatively weak claim that finite and fallible, perhaps well-meaning, human agents might hit upon the wrong commitments by mistake. After all, Sovereign Idealists will be the first to agree and hasten to remind us that they have a standard for resolving such issues, namely “the supreme principle of pure practical reason,” or Kantian morality.

Nietzsche is instead making a much more radical proposal: convictions as such are ethically problematic. A number of worries immediately arise. To begin with: why suppose so? If Nietzsche’s call to wage war on ourselves and our peers has something to do with attacking one’s own convictions, why should we be enticed by this prospect?

---

302 Nietzsche, HATH Part 9: 483
303 Nietzsche, A: 54
304 We have to lay aside the question of how the authority of the so-called moral law relates to the normativity of everyday demands like those of personal loyalty or professional obligation. Korsgaard tells us that her own view is that these demands “are independent [i.e. non-moral] sources of normativity, yet may require moral backing if they are to maintain their normative force in the face of reflection” (Korsgaard 2009, p. 2, footnote 2). Here we will only note that our discussion of the Moral Sovereignty Thesis in Chapter III ought to cast Korsgaard’s relatively conservative claim that non-moral demands “may” require moral backing in a dubious light.
Nietzsche seems to think that doing so requires a form of courage, but why? Moreover, doesn’t this set of evaluative judgments and prescriptive statements constitute an ethical conviction held by Nietzsche himself? Is he engaged in some sort of performative contradiction? Does he care? Should he?

Nietzsche saw the task suggested by these sayings as one suitable for only the most serious and exceptional writer, thinker, artist or creator. His lofty, aristocratic rhetoric about the existence and importance of an “order of rank” between individuals and his apparent disdain for “the herd” suggest a strongly elitist dimension to his thought. The same thing may be supposed to follow from his habitual association of free spirited activity with the philosophical life properly construed:

…the philosopher…has always found himself, and had to find himself, in contradiction to his today: his enemy was ever the ideal of today. So far all these extraordinary furtherers of man…have found their task, their hard, unwanted, inescapable task, but eventually also the greatness of their task, in being the bad conscience of their time…Facing a world of ‘modern ideas’ that would banish everything into a corner and ‘specialty,’ a philosopher…would be compelled to find the greatness of man, the concept of ‘greatness,’ precisely in his range and multiplicity, in his wholeness in manifoldness. He would even determine value and rank in accordance with how much and how many things one could bear and take upon himself, how far one could extend his responsibility.305

These considerations might seem to make Nietzsche’s position less attractive still. But I want to suggest that his ethos of agonistic self-cultivation has broader scope and value than many commentators have supposed. To see why, let’s take a closer look at what sort of attitude and ethical practice distinguish Nietzsche’s “the free spirit.”

To begin with, we note that if Nietzsche’s free-spirit is serious, and her seriousness is a kind of maturity, then it is a peculiar sort that is at the same time a form of play:

---

305 Nietzsche, BGE Part Six:212, p. 137
A man’s maturity—consists in having found again the seriousness one had as a child, at play.\textsuperscript{306}

Moreover, as we have already indicated, what the free-spirit is serious about are prospects for experimentation with her own ideals, her own “virtues:”

You shall become master over yourself, master also over your virtues. Formerly they were your masters; but they must be only your instruments beside other instruments…You shall learn to grasp the sense of perspective in every value judgment—the displacement, distortion and merely apparent teleology of horizons and whatever else pertains to perspectivism; also the quantum of stupidity that resides in antitheses of value and the whole intellectual loss which every For, every Against costs us. You shall learn to grasp the necessary injustice in every For and Against, injustice as inseparable from life, life itself as conditioned by the sense of perspective and its injustice.\textsuperscript{307}

So the free spirit is playful and ironic about the identities and horizons of value she inhabits. She recognizes the “quantum of stupidity” and measure of “injustice” always associated with the taking of a determinate stand on an issue or in devoting oneself to a person, an ideal or any other object of commitment. Accordingly, she refuses to treat any such lived “For” or “Against” as sacred or beyond interrogation, and appreciates the value of carrying out such interrogations in practice. In a section of Part Two of \textit{Beyond Good and Evil}, entitled “The Free Spirit,” Nietzsche frames his position in terms of the horizons of identity he struggled with throughout his life as a spurned would-be lover, a conscientious self-exile from his country and culture of origin, a former pupil, friend and adoptee into the family of Richard Wagner, and a student and professor of philology who outgrew the confines of authorized scholarship and an academic lifestyle:

Not to remain stuck to a person—not even the most loved—every person is a prison, also a nook. Not to remain stuck to a fatherland—not even if it suffers most and needs help most…Not to remain stuck to some pity—not even for higher men into whose rare torture and helplessness some accident allowed us to

\textsuperscript{306} Nietzsche, BGE Part Four: 94
\textsuperscript{307} Nietzsche, HATH I: Preface, Sec. 6
look. Not to remain stuck to a science—even if it should lure us with the most precious finds that seem to have been saved up precisely for us.\textsuperscript{308}

The lesson is that convictions, and the commitments from which they spring, can endanger the self by inhibiting its growth. But, in customary fashion, Nietzsche immediately nuances this characterization of the free spirit’s signature ethos by qualifying his endorsement of what may seem to be an excessively anti-social ideal suitable only for the most solitary:

Not to remain stuck to one’s own detachment, to that voluptuous remoteness and strangeness of the bird who flees ever higher to see ever more below him—the danger of the flier.\textsuperscript{309}

And Nietzsche concludes this section with a final word of warning about taking one’s practice of radical questioning too far in the direction of self-alienation, making of it yet another ossified “virtue” to which one clings:

Not to remain stuck to our own virtues and become as a whole the victim of some detail in us...One must know how to conserve oneself: the hardest test of independence.\textsuperscript{310}

Thus we return to Nietzsche’s discussion of the free spirit’s studied ambivalence regarding her own “For and Against.” Her ambivalence is thoroughgoing: it applies even to itself. Her attitude regarding the value and role of commitments is a nuanced one. She recognizes the stupidity and injustice of any determinate stand, but soberly perceives it in the light of its necessity for life, mindful of the value and importance of commitments and the dangerousness of her practice of radical questioning:\textsuperscript{311}

One has to test oneself to see that one is destined for independence and command...One should not dodge one’s tests, though they may be the most

\textsuperscript{308} Nietzsche, BGE Part Two, “The Free Spirit:” 41  
\textsuperscript{309} \textit{Ibid.}  
\textsuperscript{310} \textit{Ibid.}  
\textsuperscript{311} We return to this critical qualification of the free spirit’s radical practice in our discussion below of Nietzsche’s endorsement of “brief habits.”
dangerous game one could play and are tests that are taken in the end before no
witness or judge but ourselves.\textsuperscript{312}

So the free spirit is not uncommitted, but her style of commitment is distinguished
by the skepticism and ambivalence which it requires her to maintain. She is committed to
irreverent but serious play. This commitment is itself provocative and it manifests
seriousness because there is much at stake in being this way. The free spirit lives in light
of the possibility of her identity’s disruption and demolition and thereby of finding
herself disoriented. She tampers with the conditions of her “life,” indeed courting her
own death in Butler’s sense of losing one’s place in language. She strives to outgrow her
horizons by releasing her identity, by losing herself: \textsuperscript{313}

\textit{Losing oneself.} – Once one has found oneself one must understand how from
time to time to \textit{lose} oneself – and then how to find oneself again: supposing, that
is, that one is a thinker. For to the thinker it is disadvantageous to be tied to one
person all the time. \textsuperscript{314}

One must be able to lose oneself occasionally if one wants to learn something
from things different from oneself. \textsuperscript{315}

The free spirit’s habitual self-loss helps explain why Nietzsche always underscores the
profound loneliness to which he is made subject by his calling. By reaching beyond
existing resources for thought and action, he isolates herself from others and, having “lost
his way,” “becomes lonely, and is torn piecemeal by some minotaur of conscience.” He
therefore “cannot go back any longer,” “nor can he go back to the pity of men.” \textsuperscript{316} But
the free spirit’s habitual self-loss makes his predicament more austere yet: he is lonely
even in relation to himself. \textsuperscript{317}

\textsuperscript{312} Nietzsche, BGE Part Two, “The Free Spirit:” 41
\textsuperscript{313} Recall that Korsgaard takes SI to be an account of self-command or self-mastery, too.
\textsuperscript{314} Nietzsche, HATH II: Part Two, “The Wanderer and His Shadow:” 306
\textsuperscript{315} \textit{Ibid.}, Sec. 305
\textsuperscript{316} Nietzsche, BGE Part Two, “The Free Spirit:” 29
\textsuperscript{317} \textit{Ibid.}, Sec. 29
To be invested in serious play is therefore to be deeply committed, but not especially in the sense of maintaining a consistent pattern of conduct in light of a fixed ideal. For serious play is aimed in part at disrupting such patterns in order to expose and explore their internal tensions and ambiguous outer boundaries. Such playful disruption is a necessary prelude to the task Nietzsche regarded as the highest calling, suitable only for those capable of enduring its rigors: constituting genuinely fresh possibilities for thought and action that have been opened up by the Death of God. In *The Gay Science*, Nietzsche refers to the destructive creativity characteristic of the free spirit in her guise as “a discoverer and conqueror of the ideal” as “*the great health:*”

Another ideal runs ahead of us, a strange, tempting, dangerous ideal to which we should not wish to persuade anybody because we do not readily concede *the right to it* to anyone: the ideal of a spirit who plays naively—that is, not deliberately but from overflowing power and abundance—with all that was hitherto called holy, good, untouchable, divine; for whom those supreme things that the people naturally accept as their value standards, signify danger, decay, debasement…the ideal of a human, superhuman well-being and benevolence that will often appear *inhuman*—for example, when it confronts all earthly seriousness so far, all solemnity in gesture, word, tone, eye, morality, and task so far, as if it were their most incarnate and involuntary parody—and in spite of all this, it is perhaps only with him that *great seriousness* really begins, that the real question mark is posed for the first time, that the destiny of the soul changes, the hand moves forward, the tragedy *begins.*  

One might worry that this way of framing the task of serious play entails a dualism akin to the one we have criticized under the heading of “horizontality.” For our discussion of power in Chapter IV helped us to see that, in an important sense, concrete practices where questions of meaning and value are at stake are not appropriately seen as bivalent, with antecedent normative horizons figured over against discrete, ‘foreground’ activity that can only be rendered intelligible in light of those prior patterns. We found that it is not helpful to frame normative practice *tout court* in terms of ontologically

---

318 Nietzsche, GS Book Five: 382
stable dialectical relations of priority and posteriority between background patterns and foreground activity. The reason is that the figure of “normative practice tout court” is philosophically mythological. To frame questions of agency and identity in this fashion reinforces what Butler calls “the sovereign conceit” regarding agency—most saliently, the agency exercised by the theorist herself as a linguistic subject. And as we saw, doing so ignores the role of power in subjection and agency, whose excessive and productive operation transgresses efforts through speech to exhaustively circumscribe the normative domain of action and identity.

With this reminder in place, we can see that Nietzsche’s concern with the free spirit is at least in part a concern to articulate the style of activity most appropriate for an agent acting in light of her acknowledgement of the finitude of her own horizons, where ‘finitude’ denotes the productive, dynamic and excessive character of agency under AR. If the ideals embodied in our identities and expressed through our conduct are always subject to “resignification” in Butler’s sense, then possibilities for reinvention abound. In a sense, we are always reinventing our ideals. For if Nietzsche and Butler are right, and horizons of meaning and value are unstable and ambiguous, partially conditioning but outstripping the agency we possess in relation to them, then our efforts to live out the roles through which our values take shape will be as unstable, as ambiguously shaped, as those values themselves. What this means for our discussion of Nietzsche’s free spirit is that if AR is defensible, then we should want not only an account of the sort of agency possible in light of that, which we sketched in Chapter IV as part of our exposition of AR, but also a normative ideal consistent with its political psychology of power. And as we have already indicated, the free spirit is precisely that.
Recall that Judith Butler’s account of social power and linguistic agency in *Excitable Speech* was a valuable resource for our exposition and defense of AR in Chapter IV. Since her assault there on (linguistically) sovereign power is ultimately designed to motivate a broadly agonistic “politics of the performative,” her account can help shed light on the ethical possibilities represented by Nietzsche’s free spirit. We saw that Butler argues that to be a speaking subject is to be subject to speech, vulnerable to the vagaries and excesses of agency cultivated through ritual practices of definition, naming and calling. Insofar as the free spirit is a speaking subject, then, she lives in light of her “linguistic vulnerability” so construed, exploiting what Butler calls “the promising ambivalence of the norm.”

Butler is particularly interested in the political value of contestations that draw attention to the perspectival, or determinate social and historical, contingency of attempts such as Korsgaard’s to articulate “standards of universality” ostensibly governing moral or political practice as such. Examples include critical race theoretic attempts to expose the way in which constructions of rational agency encode a white-privileging ideology, as well as feminist contestations of universal rationality as it is represented in modern liberal theories of justice. By Butler’s lights, such assaults on the convictions of modern liberal politics—some of them internal to the liberal tradition itself, as in the case of Susan Moller Okin’s seminal *Justice, Gender, and the*
Family—“constitute valuable contestations crucial to the continuing elaboration of the universal itself.”

Framing things this way helps us to see that Butler’s performative politics and Nietzsche’s free spirited ethical ideal are agonistic in the sense relevant here. Putting aside the question of whether Butler wholly embraces Nietzschean perspectivism and regards any serious appeal to universal standards as misguided, it is enough for our purposes that she follows Nietzsche in seeing “the utterance”—or key, disputed meanings in political practice—as “a scene of conflict” where legalistic attempts “to prosecute the utterance in order, finally, to ‘fix’ its meaning” are confused and counter-productive. Butler’s politics echoes this chapter’s epigram in which Nietzsche cautions against the temptation “to divest existence of its rich ambiguity,” as the ambiguity Butler celebrates is recurrently refreshed by the “potential incommensurability between intention and utterance (not saying what one means), utterance and action (not doing what one says), and intention and action (not doing what one meant).” For Butler, “such disjunctures produce the possibility for a politically consequential renegotiation of language that exploits the undetermined character of these relations…” Indeed, she insists that the “lack of finality” introduced by the fact that “no final adjudication of positions can emerge…is precisely the interpretive dilemma to be valued, for it suspends the need for final judgment in favor of an affirmation of a certain linguistic vulnerability to reappropriation.”

In short, Butler celebrates the finitude of speaking subjects as tenuously empowered inhabitants of radically open spaces of possibility.

---

323 Butler 1997 (1), p. 89
324 Ibid., p. 92
Of course Nietzsche himself is no stranger to the practice of revaluation. And insofar as Nietzsche’s free spirit represents an ethical possibility for us, she is no herculean entity but an “all-too-human” subject gratefully exhilarated before the prospects of overcoming vouchsafed by her own frailty and finitude. To be exhilarated in the midst of conflict, before one’s adversaries and so before one’s own vulnerability to death, rather than bitterly recoiling from adversarial engagement in the spirit of ressentiment, is the essence and intended ethical-aesthetic appeal of “living dangerously.” That the free spirit is a possibility available to finite agents is something Nietzsche performatively demonstrates by carrying out the sort of transgressive resignifications and transformative renegotiations of language in his own work—in the very articulation of the ideal he thereby prefiguratively embodies—in which Butler, concerned with prospects for countering the injurious efficacy of hate speech, locates “the basis of an ironic hopefulness that the conventional relation between word and would might become tenuous and even broken over time.”

We have already encountered Nietzsche in this mode, but there is another passage from Book Five of The Gay Science where he plays at ‘revaluing’ precisely the vocabulary of value and agency over which advocates of Sovereign Idealism may suppose they have privileged jurisdiction. As usual, Nietzsche is out to remind us that there is no such thing:

Believers and their need to believe.— How much one needs a faith in order to flourish, how much that is ‘firm’ and that one does not wish to be shaken because one clings to it, that is a measure of the degree of one’s strength (or, to put the point more clearly, of one’s weakness)…Metaphysics is still needed by some; but so is that impetuous demand for certainty that today discharges itself among

---

325 Butler 1997 (1), p. 100. Cf. Lear’s conception of “radical hope.” In Radical Hope, Lear comes close to asserting sovereign agency. He does so by representing the prospects for resignifying cultural norms that are in crisis in terms of a social ontology that sees crisis or breakdown as a deviation from the paradigm, as we saw in Chapter IV.
large numbers of people in scientific-positivistic form. The demand that one
wants by all means that something should be firm...—this, too, is still the demand
for a support, a prop, in short, that instinct of weakness which, to be sure, does
not create religious, metaphysical systems, and convictions of all kinds but—
conserves them. 326

Here Nietzsche reminds us that he is not opposed to creation—the construction of new
standards, fresh convictions, new ways of going on. Instead he wants to identify the
dangers to the prospects for attempts at creation posed by standards already in place, at
least to the extent that the standards in question are expressed by one’s own convictions
and one handles them too gently. 327 Moreover:

Faith is always coveted most and needed most urgently where will is lacking; for
will, as the affect of command, is the decisive sign of sovereignty and strength.
In other words, the less one knows how to command, the more urgently one
covets someone who commands, who commands severely—a god, prince, class,
physician, father confessor, dogma, or party conscience...For fanaticism is the
only ‘strength of the will’ that even the weak and insecure can be brought to
attain, being a sort of hypnotism of the whole system of the senses and the
intellect for the benefit of an excessive nourishment (hypertrophy) of a single
point of view and feeling that henceforth becomes dominant—which the
Christian calls his faith. Once a human being reaches the fundamental conviction
that he must be commanded, he becomes ‘a believer.’ Conversely, one could
conceive of such a pleasure and power of self-determination, such a freedom of
the will that the spirit would take leave of all faith and every wish for certainty,
being practiced in maintaining himself on insubstantial ropes and possibilities
and dancing even near abysses. Such a spirit would be the free spirit par
excellence. 328

326 Nietzsche, GS Book Five: 347
327 See, for instance, BGE: 171: “In a man devoted to knowledge, pity seems almost ridiculous, like
delicate hands on a cyclops;” also GS: 325, an often misunderstood passage entitled “What belongs to
greatness.— Who will attain anything great if he does not find in himself the strength and the will to inflict
great suffering? Being able to suffer is the least thing: weak women and even slaves often achieve
virtuosity in that. But not to perish of internal distress and uncertainty when one inflicts great suffering and
hears the cry of this suffering—that is great, that belongs to greatness.” As usual, encouraging crude
brutality or bullying is the farthest thing from Nietzsche’s mind. His own literary and philosophical
career—marked as it was by undeserved obscurity, precarious health only made worse by the almost
desperate production of a prodigious volume of work, professional ridicule and self-alienation from his few
friends and family relations—furnishes an example of the greatness available to those willing to distress
themselves and others for the sake of creation. One must have the courage of one’s convictions, yes: but as
we have seen, the free spirit’s guiding conviction requires her to assault ideals and values held no less
dearly by herself than by her contemporaries.
328 Nietzsche, Book Five: 347
This passage is worth quoting at length because it contains a provocative resignification of so many of the core concepts employed by SI. Nietzsche has appropriated the vocabulary of self and action at home in the German idealist tradition to which his psychological naturalism and moral skepticism so strongly oppose him, turning their rhetorical force against his intellectual adversaries. The free spirit can now be seen in terms of her dignified “sovereignty” and masterful capacity for “self-determination.” She is crowned by the affect of command: she has the wherewithal, “the will,” to endure the rigors of life lived in the absence of comforting certitude. The reliance upon settled standards for thought and action is shown to be a symptom of weakness, a “disease of the will” afflicting those with feeble or declining creative energies and so incapable of finding their own way forward. To frame the point even more provocatively, the intellectual or ethical resort to standards seen as already established, “constitutive standards,” for instance, is nothing to admire. It is the recourse most suitable for ‘believers,’ not creators; for sick spirits, not free ones.

What can we expect from Sovereign Idealists by way of rejoinder to the Nietzschean ethos of agonistic self-cultivation we are offering in place of its moral legislation? For admittedly, though her theory of agency elides the formative role of power in the ways we identified in Chapter IV, Korsgaard acknowledges that the roles we occupy, and in whose terms we articulate our commitments, are not settled affairs but rather social artifacts always under construction:

I think it is important, at least in some cases, to think of a form of identity…as a role with a point. For there is room for argument about whether a particular way of acting is the best way or the only way to go about being, say, a teacher or a citizen—think of an argument about civil disobedience, for example—and it is with reference to the role or point of that form of identity that thought and

---

329 Nietzsche, GS Book Five: 347
argument about different interpretations of that form of identity can go on. There is room for creativity here, as well as argument: one might find a new way of being a friend.\footnote{Korsgaard 2009, p. 21}

But precisely by virtue of eliding power, SI has no appropriate resources for accounting for why this is or what we ought to make of its ethical significance. As we saw in Chapter IV, Korsgaard’s account of discourse, namely that norms are public resources shared through speech, is harnessed to a communicative ideal of reason-exchange beset with the same explanatory deficiency for which it might otherwise supply a remedy—it has no resources for accounting for the excess, out-of-controlness or ‘excitability’ of speech in Butler’s sense. AR, on the other hand, is inter alia an account of why precisely this feature of agency explains the necessity of reinvention for the dynamic life of a psychic or social body. And this in turn explains why Nietzsche’s free spirit is a suitable exposition of the possibility of agency represented by the figure of the “odd stray” within John Haugeland’s Heideggerian social ontology. This means that the free spirit represents not only a valuable ethical resource but a needed analytical resource, too—at least insofar as we want to explain not only the possible ethical value of nonconformity but also its integral role in maintaining the dynamism of social practices in which it may occur. Indeed, AR encourages us to go farther than this and assert that the “odd stray” represents a possibility which is everywhere manifested; all agents are nonconformists. The free spirit consummates this status by living dangerously in light of it, by cultivating

The ability to contradict.— Everybody knows nowadays that the ability to accept criticism and contradiction is a sign of high culture. Some people actually realize that higher human beings desire and provoke contradiction in order to receive some hint about their own injustices of which they are as yet unaware. But the ability to contradict, the attainment of a good conscience when one feels hostile
to what is accustomed, traditional, and hallowed—that is still more excellent and constitutes what is really great, new, and amazing in our culture; this is the step of steps of the liberated spirit: Who knows that?331

As we have already indicated, “the liberated spirit” cultivates her “ability to contradict” in part by aspiring to creation. This is why her irreverence is an indispensable ethical resource. After all, a leitmotif of Nietzsche’s mature thought is that contradiction and destruction form a critical part of creative processes: smashing old “tables of value” and fashioning new ones are two dimensions of the same ethical practice. Nietzschean agonism is aspirationally constructive.

In the 1886 Preface to Human, All Too Human, Nietzsche describes a process of emancipation free spirits may undergo when passing from a fettered state into their “great liberation:”

It is at the same time a sickness that can destroy the man who has it, this first outbreak of strength and will to self-determination, to evaluating on one’s own account, this will to free will: and how much sickness is expressed in the wild experiments and singularities through which the liberated prisoner now seeks to demonstrate his mastery over things! He prowls cruelly around with an unslaked lasciviousness…With a wicked laugh he turns round whatever he finds veiled and through some sense of shame or other spared and pampered: he puts to the test what these things look like when they are reversed.332

This passage typifies an aspect of Nietzsche’s philosophical activity during the time at which he wrote the preface to the first volume of Human, All Too Human. It was drafted in 1886, shortly before the publication of Beyond Good and Evil, a book rife with attempted revaluations of such familiar concepts as freedom, virtue, knowledge and the pursuit of truth, and morality itself. In this passage Nietzsche demonstrates the style of activity he at the same time describes—the “imperious” reversal of accustomed valences,

331 Nietzsche, GS Book Four: 297
332 Nietzsche, HATH I: Preface: 3
of “everything [one] had hitherto loved!”—by carrying out his own provocative reversals in the course of providing the description.

The Kantian rhetoric Korsgaard uses to motivate her formal ideal of unified agency—the rhetoric of “free will” understood as self-determination, reflective criticism by one’s own unfettered lights—Nietzsche subjects to customary revaluation and transforms into an antipodal ideal that is meant to have its own, quite different allure. In *Self-Constitution*, Korsgaard argues that her Kantian moral ideal of self-constitution through psychic self-unification is essentially the same as the conception of justice as self-mastery defended by Socrates in Plato’s *Republic*.\(^3\)\(^3\)\(^3\) It is fitting, then, that Nietzsche also revaluates the Platonic image of the liberated prisoner, here not shown as bound for communion with transcendent reality, as in the *Republic*, but as robustly engaged in an earthbound struggle over the meaning and value of the pieties associated with one’s accustomed “duties”—the “value of our values,” as he puts it elsewhere.\(^3\)\(^3\)\(^4\)

In keeping with Nietzsche’s commitment to psychological naturalism and his fondness for agonal metaphors, he portrays his exemplar’s unsparing critical activity as animalic and predatory. She “prowls” rather than ascends.\(^3\)\(^3\)\(^5\) Accordingly, Nietzsche’s unfettered former captive is characteristically moved not by respect for the felt authority of a formal constraint of reason but by a “wicked” wish to probe and expose the artifice

---

\(^{333}\) Korsgaard 2009, Ch. 7 *et passim.*

\(^{334}\) Nietzsche, GM: Preface.

\(^{335}\) Of course Nietzsche often employs a vocabulary of ascent and descent that calls to mind Plato’s allegory of the cave. Notably, the story of Zarathustra begins with his “going down” from his mountain fastness into mixed company, and the “true seeker after knowledge” portrayed in BGE is compelled by the requirements of his vocation to “go down” in a similar fashion. Though we cannot adequately explore the comparison here, I think there are pronounced dissimilarities deriving from Nietzsche’s rigorously naturalistic conception of the “all-too-human” character of the objects of his studies, namely the ideals or “tables of value” which orient human practices; to say nothing of other salient differences.
by which the homely, “all-too-human” pedigree of such lofty conceptions is veiled.\(^{336}\) Nietzsche’s own subversive genealogical practice of inquiry thus models the agonistic integrity of the free spirit. Skeptical irreverence before “habitual evaluations and valued habits,” not Kantian reverence for universal law, is her motivational drive-wheel.\(^{337}\) The posture most befitting her ideal is one of defiant, experimental zeal, not solemn, dutiful obedience. Indeed, to return to this chapter’s epigram, playful irreverence for accustomed modes of thought and action is the negative, critical side of her joyful, prefigurative “reverence for all that lies beyond [her] horizon.”\(^{338}\) And to return to Nietzsche’s discussion in *The Gay Science* of “the great health,” hers is “the ideal of a spirit who plays naively—that is, not deliberately but from overflowing power and abundance.”

The preface of *Human, All Too Human* quoted above provides a rich psychological characterization of Nietzsche’s free spirit. Her irreverence and thirst for conquest is leavened by the discipline and circumspection required of athletes, whose periodic respites from the rigors of training and competition invite comparisons with the free spirit’s gratefully embraced, intermittent “convalescence.” Such detours ultimately enable her to attain to that

---

\(^{336}\) The rationale for Nietzsche’s genealogical method of inquiry can be understood in this light. Indeed, the expressly genealogical mode was most prominent in Nietzsche’s published writings (1886-1887) around the time of the writing of the first preface for *Human, All Too Human* (Spring 1886).

\(^{337}\) Nietzsche, HATH I: Preface: 1

\(^{338}\) Nietzsche, GS Book Five: 373, ‘‘Science’ as a prejudice.—’’ In this passage Nietzsche takes aim at positivism in the sciences, arguing that “an essentially mechanical world would be an essentially meaningless world.” Korsgaard has often leveled a similar objection against what she calls The Scientific World View (2009, p. 41). So it is ironic that, as we argued in Chapter IV, Korsgaard’s own transcendental account of agency is designed to fundamentally purge our self-conception as agents of any ambiguity and so vitiates the ethical case for a posture of ambivalence in the manner Nietzsche ascribes to scientific positivism, though for different reasons. Korsgaard thus rejects one exhaustive, univocal view of human action in favor of another.
tremendous overflowing certainty and health which may not dispense even with wickedness, as a means and fish-hook of knowledge, to that mature freedom of spirit which is equally self-mastery and discipline of the heart and permits access to many and contradictory modes of thought – to that inner spaciousness and indulgence of superabundance which excludes the danger that the spirit may even on its own road perhaps lose itself and become infatuated and remain seated intoxicated in some corner or other, to that superfluity of formative, curative, moulding and restorative forces which is precisely the sign of great health, that superfluity which grants to the free spirit the dangerous privilege of living experimentally and of being allowed to offer itself to adventure: the master’s privilege of the free spirit!

It is in light of his elevation of irreverence and wide-ranging experimentation to an ethical ideal that Nietzsche’s often misunderstood praise of “cruelty” and “hardness,” as well as his often overlooked celebration of the human capacity for joyful activity, are to be understood.

The discipline made possible by the former is an indispensable resource for embracing one’s life in the joyful, yea-saying spirit of gratitude in which Nietzsche thinks humanity must place its highest hopes. Nietzsche’s elitist rhetoric manifests his recognition of the great difficulty of affirming one’s entire life from first to last; to say the least, not everyone will be capable of doing this. Nevertheless, this is the challenge proposed by the demon of The Gay Science through whose nocturnal visitation to one in one’s “loneliest loneliness” Nietzsche first announced his doctrine of the eternal recurrence:

‘This life as you now live it and have lived it, you will have to live once more and innumerable times more; and there will be nothing new in it, but every pain and every joy and every thought and sigh and everything unutterably small or great in your life will have to return to you, all in the same succession and sequence—even this spider and this moonlight between the trees, and even this moment and I myself. The eternal hourglass of existence is turned upside down again and again, and you with it, speck of dust!’

339 Nietzsche, HATH I: Preface, Sec. 4
340 Following Nietzsche’s typical usage, ‘humanity’ here denotes late modern Western European civilization in the aftermath of the Death of God, that is, during the dissolution of traditional, especially religious and moral, sources of authority.
341 Nietzsche, GS Book Four: 341, “The Greatest Weight"
Nietzsche’s ethos of self-cultivation is therefore not a universalistic one. There is nothing to guarantee that every human agent will have favorable, let alone interpersonally comparable, prospects for flourishing in the relevant sense. One way of putting this is that some will be unable to bear up under the strain of encountering the manifold of differences through which one amasses a greater wealth of understanding:

Happiness and virtue are no arguments. But people like to forget—even sober spirits—that making unhappy and evil are no counter-arguments. Something might be true while being harmful and dangerous in the highest degree. Indeed, it might be a basic characteristic of existence that those who would know it completely would perish, in which case the strength of a spirit should be measured according to how much of the ‘truth’ one could still barely endure…Perhaps hardness and cunning furnish more favorable conditions for the origin of the strong, independent spirit and philosopher than that gentle, fine, conciliatory good-naturedness and art of taking things lightly which people prize, and prize rightly, in a scholar.\(^{342}\)

Here hardness and cunning are aligned with strength and independence of spirit. Since the former are also contrasted with the retiring diffidence said to be typical of the blandly agreeable, scholarly type skilled in the “art of taking things lightly,” by implication we have another alignment of the free spirit with the art of taking things seriously, now represented as requiring a disciplined practice of exposing oneself to an open-ended range of potentially “harmful and dangerous” perspectives.

Nietzsche’s free spirit therefore represents a dangerous but ethically promising possibility, that of embracing one’s life from first to last. The challenge posed by the demon of \textit{The Gay Science} is to say yes to one’s suffering without saying no to the life, to one’s finite, this-worldly, “all-too-human” life with all its peril and frailty, of which that suffering is an integral part. Indeed, Nietzsche’s call for joyful living is inseparable from his denunciation of those who would rid life of suffering if they could. For, by

\(^{342}\) Nietzsche, BGE Part Two: 39
Nietzsche’s lights, suffering is integral to one’s prospects for finding “one’s own way,” for confronting one’s finitude on one’s own terms. Nietzsche’s exhortation “to give style to one’s character”\textsuperscript{343} is therefore to be understood in terms of “the personal necessity of distress.”\textsuperscript{344}

5: worries

What we have been saying on Nietzsche’s behalf may grate on the ears. Why call this sort of activity, one defined by skeptical irreverence for one’s own accustomed perspectives, ‘integrity’? Wouldn’t conducting oneself in this fashion be irresponsible—maybe even paradigmatically so? And won’t it be corrosive? In short, why should a practice defined in these terms to be sufficiently attractive to designate it with a strongly approbative term like ‘integrity’?

To begin with, we can identify a programmatic reason for reserving the term ‘integrity’ for what is sought when one strives to overcome one’s own horizons through transformative encounters with difference. Recall that we want to develop a conception of integrity based on our account of subject formation and agency, Agonistic Realism. Recall too that AR represents human agency as paradigmatically productive of normatively salient differences, excessive in its relation to discrete exercises of volition, and dynamic in that the range of differences at stake in human practices is indefinite. In light of this, it should come as no surprise that we should see integrity in terms of contestation. For this is as natural for us as it is for those who think the self is paradigmatically whole, Sovereign Idealists, to see integrity in terms of its unification. In

\textsuperscript{343} Nietzsche, GS Book Four: 290
\textsuperscript{344} Ibid., Sec. 338. See also ibid., Sec. 318, 325, and Nietzsche, BGE: 42.
each case, integrity is seen as something sought through struggle amidst the complexity of our psychic lives, and what is sought is a concerted, skillful mode of the activity that defines the condition of the self that acts.

On both accounts of agency, then, ‘integrity’ denotes a lived consummation of the conditions for the formation and normative empowerment of subjects. One salient difference is that in the case of SI, the conditions are to be construed in metaphysically robust terms, whereas AR eschews any resort to “final horizons” construed teleologically or otherwise. This was part of our account of the methodological implications of the ‘dynamism’ of agency under AR. We see now that there is a further implication for—or, rather, about—those who resort to final horizons in the exercise of their own powers of abstraction: they lack integrity in the sense AR makes possible; they lack agonistic integrity. To translate this into robustly agonistic terms, they run afoul of Nietzsche’s heroic criterion of craving a worthy adversary. For resorting to ‘final’ horizons of analysis is like trying to play a trump card. One thereby strives for discursive closure, and in doing so one effectively tries to outpace any possible adversary, to settle once and for all whatever is at issue. On an agonistic conception of integrity, doing so manifests a lack of respect for the adversarial relations in whose midst one’s own finite horizons and normative powers are precariously situated.

What about the suspicion that Nietzsche’s free-spiritedness is really a recipe for being a deeply irresponsible person? Is the radically skeptical practice he extols really compatible with caring deeply for anyone or anything?

To begin with, we need not take Nietzsche’s description of the free spirit to be an exhaustive account of the psychology and conduct of any actual or possible person. As
always, he is out to model and provoke independent engagement with the topics he addresses, to diagnose problems of thought and action and prefigure ways forward by offering fresh perspectives, not to universally legislate or give a full accounting of what anyone can or should be. His refusal to resort to such “final horizons” of value or interpretation manifests his sense of the importance of the art of nuance and prescriptive restraint. It is his pedagogical exercise of agonistic integrity: he wants neither followers nor adherents but peers and rivals. It is part of what is suggested by the phrase “beyond good and evil”—taking such matters seriously requires abandoning the resort to bivalent interpretations of psychic phenomena by, for instance, seeing them as instantiations of some univocal category like “rationality,” “normative authority,” “autonomy,” and so on.

The free spirit’s seriousness is closely tied to the necessity of her prescriptive and pedagogical restraint:

> Whoever is a teacher through and through takes all things seriously only in relation to his students—even himself.345

To take a complicated thing seriously, and to take others seriously in relation both to it and to one’s own efforts to understand it, requires the discipline of not terminating the engagement with all-embracing solutions. Owing to the excess of agency discussed in Chapter III, any perspective is sure to encounter potentially transformative resistance from other perspectives. Given this, the Nietzschean hypothesis is that cultivating the ineluctable play of differences by embracing agonism, rather than trying in vain to suppress it through discursive closure, is healthier and will prove to be a more fruitful

---

345 Nietzsche, BGE Part Four: 63
strategy. Crucially, like the doctrine of the will to power itself, this is to be regarded as a hypothesis rather than a transcendental argument.\textsuperscript{346}

We may thus justifiably regard Nietzsche’s portrait of the free spirit as a stylized rendering of a possibility for life in the late modern social conditions that Nietzsche is concerned with. So construed, the free spirit represents a dimension of possible ethical practice Nietzsche regards as attractive for the reasons we have been considering. Thus taken as a resource for ethical thought and action rather than an exhaustive blueprint for a worthwhile life, much of the worry that Nietzsche may be encouraging us to abandon our families or adopt a life of crime as “experiments in living” may lose its force.\textsuperscript{347}

But we need not leave it at that. Nietzsche is keenly aware of the importance of maintaining commitments in the straightforward sense discussed by Williams and Korsgaard. But he identifies a style of commitment more congenial to the practice of experimentation we have been discussing:

Brief habits.— I love brief habits and consider them an inestimable means for getting to know many things and states, down to the bottom of their sweetness and bitternesses…I always believe that here is something that will give me lasting satisfaction—brief habits, too, have this faith of passion, this faith in eternity…Enduring habits I hate. I feel as if a tyrant had come near me and as if the air I breathe had thickened when events take such a turn that it appears that they will inevitably give rise to enduring habits; for example, owing to an official position, constant association with the same people, a permanent domicile, or unique good health. Yes, at the very bottom of my soul I feel grateful to all my misery and bouts of sickness and everything about me that is imperfect, because this sort of thing leaves me with a hundred backdoors through which I can escape from enduring habits.\textsuperscript{348}

\textsuperscript{346} Nietzsche, BGE Part Four: 36
\textsuperscript{347} Besides, Nietzsche consistently repudiates overtly ‘immoral’ behavior such as the ruthless self-assertion of the bully as flatly incompatible with his ideal of strong, independent spirits. After all, that bullying (for instance) is a paradigmatic manifestation of weakness and insecurity is no arcane psychological doctrine; it is learned in the sandbox.
\textsuperscript{348} Nietzsche, GS Book Four: 295
As Walter Kaufmann observes in his translation of *The Gay Science*, Nietzsche hastens to qualify this elaboration of living dangerously. He concludes the “brief habits” section by recognizing the special importance of stable patterns for the sustainability and long-term effectiveness of a practice otherwise directed at radical questioning. This crucial qualification recalls our initial discussion of the free spirit’s characteristic style of serious play. To play with one’s own ideals or those of one’s culture—to “philosophize with a hammer” in Nietzsche’s autobiographical formulation—is to live dangerously in a number of respects. Nietzsche’s free spirit well knows that attachment to any particular perspective involves a “quantum of stupidity” and injustice because it comes at the expense of alternative possibilities for thought and action. But she is equally aware that taking a stand, having one’s “For and Against,” is necessary for life. In short, one cannot suspend judgment about everything or all at once; such a feat would erode the basis for gathering one’s energies and channeling them in a measured way:

> Most intolerable, to be sure, and the terrible par excellence would be for me a life entirely devoid of habits, a life that would demand perpetual improvisation. That would be my exile and my Siberia.³⁴⁹

There is another connection between the free spirit’s prospects for serious play and her need for periodic withdrawals into spiritual convalescence; the connection is forged by the heroic trait of craving a worthy adversary. The idea that retreating from radical questioning helps one gather one’s energies for renewed excursions is Nietzsche’s revalued endorsement of the capacity “to forget” which he ascribes to the “noble” or “master” type of *Genealogy* I, as it is the luxury of forgetting which frees the noble type who can afford it from the corrupting influence of *ressentiment*:

³⁴⁹ Nietzsche, GS Book Four: 295
To be incapable of taking one’s enemies, one’s accidents, even one’s misdeeds seriously for very long—that is the sign of the strong, full natures in whom there is an excess of the power to form, to mold, to recuperate and to forget...Such a man shakes off with a single shrug many vermin that eat deep into others; here alone genuine ‘love of one’s enemies’ is possible—supposing it to be possible at all on earth. How much reverence has a noble man for his enemies!—and such reverence is a bridge to love.— For he desires his enemy for himself, as his mark of distinction: he can endure no other enemy than one in whom there is nothing to despise and very much to honor.\footnote{\textit{Nietzsche, GM I:10}}

Of course, a leitmotif of Nietzsche’s mature work is that late modern people are all inheritors of the threefold legacy of “the slave revolt in morality” whose conditions of origin and value for life the \textit{Genealogy} diagnoses: \textit{ressentiment}, bad conscience and ascetic values. Nietzsche’s own solitary and scholarly life was deeply ascetic, and the free spirited type he exalts—embodied \textit{inter alia} by the figure of “the philosopher of the future”—exhibits characteristics which equally recall the nobles and the slaves of \textit{Genealogy I}. As we have seen, his free spirit is noble, strong and independent, but not after the crudely violent fashion of the primitive early conquerors, the “blonde beasts,” of \textit{Genealogy I}. The strength and courage of his free-spirited, radical questioner are those required for perilous excursions into deep problems, for an exercise of the capacities for “remembering” and for sustained, focused activity first made possible by “the slave revolt in morality” and “the internalization of man.” In this way, the free spirit is more akin to the slavish ascetic type.

But Nietzsche is emphatic that the free spirit is not slavish but masterful. To see what this means, we need only recall that the impotent, self-lacerating rage of \textit{ressentiment}, fueled by the dream of revenge upon the source of one’s constraints, is the signal affect of the slave type of \textit{Genealogy I}. Whereas the primitive noble type is shallow and capable of no great spiritual achievements like the feats of creative
revaluation to which their slaves are driven, it is free of *ressentiment* and other corrosively reactive affects associated with it such as the bad conscience. What crucially distinguishes the possibility the free spirit represents from the slave type of *Genealogy I* is the former’s triumph over the disempowering urge to take revenge upon life, her overcoming and self-overcoming of *ressentiment*:

*That man be delivered from revenge,* that is for me the bridge to the highest hope.\(^{351}\)

She accomplishes this, as we have seen, by embracing life from first to last, affirming her suffering as integral to her prospects for self-overcoming in which she finds her joy and highest calling:

*Who suffers much makes the devil envious, and is thrown out into heaven.*\(^ {352}\)

A free spirit in Nietzsche’s sense is someone possessing the psychic depth and creative potential inherent in the slave type—she points the way beyond reigning values, towards new ideals—without the longing for an existence unconditioned by finitude and contingency.

Nietzsche insists that we all equally masters and slaves; that the conflict between the ethical possibilities represented by these psychic types is as yet unresolved. The message is that this conflict has a future, and that its future depends upon our capacity to live it out in a bold exploratory spirit of agonistic respect for the rivalries in whose midst we are contingently formed, sustained, threatened and potentially ennobled:

> The two *opposing* values ‘good and bad,’ ‘good and evil’ have been engaged in a fearful struggle on earth for thousands of years; and though the latter value has certainly been on top for a long time, there are still places where the struggle is as yet undecided. One might even say that it has risen ever higher and thus become more and more profound and spiritual: so that today there is perhaps no more

---

\(^{351}\) Nietzsche, Z II:7, "On the Tarantulas"

decisive mark of a ‘higher nature,’ a more spiritual nature, than that of being divided in this sense and a genuine battleground of these opposed values.\textsuperscript{353}

Let us conclude by returning once more to this chapter’s epigram and to the passage from \textit{The Gay Science} in which Nietzsche exalts “the ideal of a spirit who plays naively…with all that was hitherto called holy, good, untouchable, divine” and “those supreme things that the people naturally accept as their value standards.”\textsuperscript{354} Such an irreverent innovator has no wish “to divest existence of its rich ambiguity but instead exploits what Butler calls “the promising ambivalence of the norm.”\textsuperscript{355} We might still wonder what, if anything, integrity understood agonistically could concretely look like in contemporary life. In the following section, we will offer some concrete examples of what Nietzsche’s free spirited ideal might look like when put into practice.

6: Illustrations of Agonistic Integrity: literature as self-overcoming

\textit{Charles Bukowski & gendered ambivalence}

“\textit{The Bluebird}”\textsuperscript{356}

there's a bluebird in my heart that
wants to get out
but I'm too tough for him,
I say, stay in there, I'm not going
to let anybody see
you.

there's a bluebird in my heart that
wants to get out
but I pour whiskey on him and inhale
cigarette smoke
and the whores and the bartenders
and the grocery clerks

\textsuperscript{353} Nietzsche, GM I: 16
\textsuperscript{354} Nietzsche, \textit{The Gay Science}, Sec. 382
\textsuperscript{355} Butler 1997 (1), p. 91
\textsuperscript{356} Bukowski 1992, p. 120
never know that
he's
in there.

there's a bluebird in my heart that
wants to get out
but I'm too tough for him,
I say,
stay down, do you want to mess
me up?
you want to screw up the
works?
you want to blow my book sales in
Europe?

there's a bluebird in my heart that
wants to get out
but I'm too clever, I only let him out
at night sometimes
when everybody's asleep.
I say, I know that you're there,
so don't be
sad.

then I put him back,
but he's singing a little
in there, I haven't quite let him
die
and we sleep together like
that
with our
secret pact
and it's nice enough to
make a man
weep, but I don't
weep, do
you?

The poem culminates in a brisk, subtle rhetorical performance that affords a way
to unpack the complexity disguised by Bukowski’s homely language. Bukowski’s final
word on the subject (“…but I don’t weep, do you?”) is ambiguous. On the one hand, we
are offered a dismissive macho gesture, a tough manly pose that would obviate the
foregoing revelation but misfires, calling attention to its own transparent inadequacy as a compensatory affectation. For of course the poet weeps. His poem is a sustained weeping; a lyrical, comic self-parody. Indeed, this much is attested by another available reading of the closing interrogative, namely as a “tell,” a doubtfully self-conscious expression of insecurity about the self-assured pose to which the expression simultaneously, but unconvincingly, aspires. As a comic self-parody, “The Bluebird” is an artistic exercise of good humor, humility and courage. But its deceptively subtle, ironically laconic handling of the figure of stoical masculinity goes farther and manifests agonistic integrity when placed in the larger context of Bukowski’s life and literary output.

In his work, Bukowski often acts and lives out that macho pose. There is an odd way in which his literary alter egos (paradigmatically, “Henry Chinaski”) do so in earnest, all the while calling attention to its status as a farcical compensatory mechanism. Bukowski neither pig-headedly (and piggishly) embraces traditionally masculine postures as unproblematic nor discards them wholesale in a futile effort to bootstrap himself ex nihilo into a more viable way of inhabiting his masculinity. Through his writing and in this poem especially, he enacts a deeply felt ambivalence about his own gendered horizons, exhibiting their troubling normative complexity for the reader, who is left to decide for herself what to think of the stylized experimentation carried out by the writer and his characters (cf. “Some of my readers,” about the violent feminist backlash to Bukowski’s work in Germany, which he unresentfully documents there).

357 Thanks to James Olsen for helping me recognize this additional layer of complexity in the poem’s closing line.
To return to “The Bluebird,” Bukowski never fully abandons the macho affectation ostensibly threatened by the titular creature. The dismissive rhetorical posture we identified is after all the final word, albeit an elliptical one (“…but I don’t weep, do you?”). This is as important as the fact, which we have already observed, that by the poem’s end that final word rings hollow and absurd to the observant reader. The poem is not an exercise in self-loathing but an experiment in playful self-deprecation in which we are encouraged to share the poet’s good-natured amusement as well as his mixed feelings. The possibility of laughter is as important to the poem’s delicate mood as the possibility of overcoming the shame of weeping. In the context of the present discussion, we naturally recall Nietzsche’s recurrent exaltation of laughter—especially laughter before an adversary, in the midst of loss, in the face of one’s finitude—as a profound indicator of existential circumspection and psychical good health.

The figure of the bluebird elegantly represents all of this for us. It embodies a set of contested affects which, among other things, take themselves as their own object. Its endangered predicament tells us something about Bukowski’s complicated attitude

---

358 For instance: “The Olympian Vice.—In despite of that philosopher who, being a real Englishman, tried to bring laughter into ill repute among all thinking men—‘laughing is a bad infirmity of human nature, which every thinking mind will strive to overcome’ (Hobbes)—I should actually risk an order of rank among philosophers depending on the rank of their laughter—all the way up to those capable of golden laughter. And supposing that gods, too, philosophize, which has been suggested to me by many an inference—I should not doubt that they also know how to laugh the while in a superhuman and new way—and at the expense of all serious things. Gods enjoy mockery: it seems they cannot suppress laughter even during holy rites.” (BGE Part Nine, “What Is Noble:” 294); “Caricature is the beginning of art. That something signifies, delights. That whatever signifies, should mock and be laughed at, delights still more. Laughing at something is the first sign of a higher psychic life (as in the fine arts).” (Nachlass, Musarion edition, IX, 413); “Not by wrath does one kill but by laughter. Come, let us kill the spirit of gravity!” (Z First Part, “On Reading and Writing:” 7); “What has so far been the greatest sin here on earth? Was it not the word of him who said, ‘Woe unto those who laugh here’?...He did not love enough: else he would also have loved us who laugh. But he hated and mocked us: howling and gnashing of teeth he promised us...Laughter I have pronounced holy; you higher men, learn to laugh!” (Z Fourth and Last Part, “On the Higher Man:” 13; also 16 and 20); “Perhaps this is where we shall still discover the realm of our invention, that realm in which we, too, can still be original, say, as parodists of world history and God’s buffoons—perhaps, even if nothing else today has any future, our laughter may yet have a future.” (BGE Part Seven: 223).
towards the demise of the way of being a man into which he was socialized by, among others, his emotionally stunted, abusive, alcoholic father. The poem reveals that, within Bukowski’s divided relation to himself, the unspecified set of affects represented by the bluebird are imperiled by the very alcoholism and dumb inarticulacy which Bukowski inherited from his father’s World War I generation (“there’s a bluebird in my heart that wants to get out but I pour whiskey on him and inhale cigarette smoke”).

For Bukowski, the new possibilities of going on as a man represented by the bluebird are as yet too inchoate for him to inhabit, so he imprints them with an image of something heart-breakingly lovely (“it’s nice enough to make a man weep”), sonorous but inarticulate (“he’s singing a little in there”), and extremely vulnerable (“I haven’t quite let him die”). Those who know Bukowski’s work will know that he typically wrote late into the night each night, surrounded by a growing pile of empty bottles of beer and liquor. The poem leaves little room for doubt, then, that the bluebird inside the poet is, among other things, a carefully chosen image of a possibility which he is inclined to embrace: “I'm too clever, I only let him out at night sometimes when everybody's asleep. I say, I know that you're there, so don't be sad.” The poem calls the normative status of this nascent possibility into question while celebrating its fragile beauty, puzzling over its peculiar resilience, and leaving its fate as uncertain as that of the gentle, endangered creature which the narrator anxiously conceals inside himself in the presence of others.

Or so he ironically tells us. For of course the poem itself exhibits for all to see that which the bluebird represents. It is an indirect self-disclosure through a transparent artifice of concealment. By making a show of what he cannot show us, Bukowski makes mentionable what men like him are characteristically ashamed to name—their confusion
about what to make of the evident failures of their own way of being a man. That these familiar failures—other-directed physical and psychological violence, self-neglect, emotional inarticulacy—were seen as such by Bukowski is manifest throughout his oeuvre.

These considerations bear on the charge of misogyny leveled by some of Bukowski’s critics. The accusation is popular but I think it rests on a shallow, dismissive reading. Though it would go too far to call him a feminist, much of his work lays out many familiar grievances with patriarchal constructions of masculinity in robust, candidly experiential terms. His life and work show him to have been an early and lifelong victim of the violence inherent in patriarchal ways of being masculine, as his father was an emotionally stunted drunk who abused him and left him with many of the traits in question. Bukowski’s work shows a vivid awareness of the ugliness of these traits, and the self-deprecating element in his writing caricatures it by representing his alter egos as clownish and broken.

These facts are relevant because they manifest a great deal of integrity if integrity is read agonistically as a playful, confrontational style of critical engagement with one’s own horizons of identity and value. Through his literary output, Bukowski uses autobiographical self-parody to lyrically muster the courage for an attack on those of his convictions with which he struggled to come to grips. Much of his work manifests and advances that struggle; like Nietzsche’s own writings, his poetry and prose is a record of
the inconclusive battles waged by an individual striving to make sense of what he was and could be in relation to those people and projects of great concern to him.\textsuperscript{359}

7: illustrations: bodily confrontation

\textit{MLK}

Consider the following passage from Martin Luther King, Jr.’s “Letter from a Birmingham Jail:”

Birmingham is probably the most thoroughly segregated city in the United States. Its ugly record of brutality is widely known...As in so many past experiences, our hopes had been blasted, and the shadow of deep disappointment settled upon us. We had no alternative except to prepare for direct action, whereby we would \textit{present our very bodies as a means of laying our case before the conscience of the local and the national community}...Nonviolent direct action seeks to create such a crisis and foster such a tension that a community which has constantly refused to negotiate is forced to confront the issue. It seeks so to dramatize the issue that it can no longer be ignored...I must confess that I am not afraid of the word ‘tension.’ I have earnestly opposed violent tension, but \textit{there is a type of constructive, nonviolent tension which is necessary for growth.}\textsuperscript{360}

The case of nonviolent direct action so understood supplies a context for recognizing structural features of social life overlooked or diminished in Sovereign Idealist views of agency and politics. King’s description of the strategic situation confronting his movement highlights the role of an agonistic interpersonal stance in the pursuit of democratic aims. Direct actions may challenge and pressure the sanctioned or accepted boundaries of discourse where the need to do so is keenly felt by agents who find themselves on the losing end of the power formations it expresses and helps to reinforce.

\textsuperscript{359}Interestingly, for both men, the literary and gendered horizons they inhabited furnished much of grist for their creative projects of self-diagnosis and overcoming. “Nietzsche’s women” and “Bukowski’s women” are equally immortalized in their works.

\textsuperscript{360}Cahn 2002, p. 1201
The case of nonviolent direct action is thus illustrative of a broader range of phenomena which are constitutive of politics by underwriting many of its possibilities for intersubjective engagement. A suitably well-executed, enacted acknowledgement of these phenomena by political actors can supply them with resources for constructively extending their political practice. King’s letter expressly testifies to the value of constituting of new spaces of what is interpersonally possible through a tactically discerning, bodily adversarial politics. So the case of nonviolent direct action helps dramatize the role of lived, bodily engagement in the domain of power.

In what way is the ‘bodily’ dimension of politics more pervasive than it might appear at first? Why suppose that nonviolent direct action of the sort carried by civil rights protestors in the 1960s South is emblematic of political possibilities (and perils) more broadly? The simplest answer is that politics is constituted by the operation of power. This is power in the Nietzschean sense required by AR: a polymorphous field of effects and relations for the production and exercise of capacities for normatively salient action and response. The “field” consists in networks of agencies whose determinate constitution—their peculiar capacities for action and reaction—are intrinsically at stake in the relations obtaining among them. On such a view, the terms ‘power,’ ‘capacity,’ ‘agency’ and their cognates cannot be defined independently of one another or given determinate content other than by reference to accounts—partly descriptive, partly evaluative—of historically concrete constellations of interpersonal difference. Whatever priorities or activities are encompassed by a particular politics—conversation, argument, identity, rhetorical performance, resource distribution and management, collective action—are conditioned by and realized through indefinitely proliferating deployments
and redployments of power so understood. What this means in relation to the example furnished by King’s letter is that political transactions are shot through with the sort of potentially constructive “tension” to which he refers.

In his letter, King points to the tension inherent in the fabric of his racially segregated society: “Actually, we who engage in nonviolent direct action are not the creators of tension. We merely bring to the surface the hidden tension that is already alive. We bring it out into the open, where it can be seen and dealt with.” This insight can be broadened, too. Instability and vulnerability are constitutive features of the norms composing any social practice or set thereof. Any concretely realized structure of norms opens up a range of possibilities for action and forecloses others. Moreover, since the norms themselves are relational, porous affairs, the language in which they inhere is itself inherently “excitable,” the range of acts they permit and foreclose indefinite and hence susceptible of being tampered with by those they constitute as agents. This dynamic underwrites the freedom of concretely engaged subjects to perform in a normatively intelligible fashion, but always within certain constraints. Any formation of socially instituted power therefore enables more or less determinate patterns of intelligible action on the part of participants embedded in the power relations in question; their freedom is real but bounded in this way by the space of normatively available alternatives for engagement.

However, it is essential to see that the scope of freedom so understood is necessarily indefinite. Were it not, social and political life would be a static affair, its boundaries once and for all time fixed. In this case, it would not the dynamic, bounded though open-ended [this bivalence of the medium of agency suggests a correlative
attitude of ambivalence on the part of agents themselves; this is in short the central claim of my project] manifold of action familiar to concrete political actors who may always intelligibly endeavor to exploit the fluidity of its constraints to intervene in a regime of norms for some specific purpose or other. The porousness of the boundaries of any social practice is therefore what makes transformative interventions into its character possible. One way in which this can be seen is in terms of the importance of timing in the success of any such attempted intervention.

To sum up the relevant implications of the foregoing sketch of power in its relation to agency. There can be no fully realized agent; no inherently stable, context-transcending, fixed position of action or assessment from which words or deeds can issue forth with irresistible force or unambiguous import; no standpoint of evaluation beyond the destabilizing/potentially contextually delegitimizing influence of robust contestation. No normative force can exist that is not concretely operative on, and inherent in the always embodied relations between, historical persons.

1990s Riotgrrl/Queercore youth movements

The ‘Riot grrrl’ and ‘queercore’ youth cultural movements of the 90s supply further contemporary illustrations of the potential value of agonistic integrity for politics. Here is Juliana Luecking, representing sexuality as a contested field of potentially prefigurative practice:361

Esperanza was a young girl who dreamed of other little girls kissing on her. Esperanza knew she didn’t wanna be anybody’s husband, she didn’t wanna be anybody’s wife. She didn’t wanna be one of those peaceable-like lesbian pacifist people she sees on the national public television talk shows. She knew she didn’t

361 From the She’s Good People compilation of Luecking’s spoken-word performances.
wanna be no lezzie prostitute addict case, hardship on the street. What she wanted to be was a fiercely independent homosexual woman. But she didn’t have a picture of that in her head.\textsuperscript{362}

Another illustrative case is the “Free to Fight” multi-media project released in 1995 by Candy Ass Records, Jody Bleyle’s Queercore independent record label. The project consisted of a seventy-five page booklet and double-album compilation featuring all-female bands within the Riot grrrl and Queercore movements. The project, which addressed issues of violence against women and provided personal testimonies and self-defense resources, was prompted in part by Bleyle’s experiences with violence at underground shows and her frustration at the failure of her efforts to seek adequate recourse through the legal system as a self-identified dyke. Bleyle’s then-band Team Dresch toured to support Free to Fight in 1995. They were accompanied by martial arts instructor Alice Stagg, who staged self-defense demonstrations for female audience members before the band’s set.

In Lucy Thane’s 1997 documentary film, \textit{She’s Real, Worse than Queer}, Stagg observes that by encouraging women to perform an aggressive, confrontational and heavily politicized style of music, the Queercore and Riot grrrl scenes of the mid-1990s changed the local space of possibilities for women:

So they see the demo, and then…first of all, they’re more aware of their circumstances there, where they are, at that show, which a lot of times, those places aren’t safe, they’re not made safe for women, and so already maybe the safety increases…But also just the awareness of like power and fighting on [sic] a really physical way. And then you watch the music and…maybe it would be easier to see like when a woman is like screaming over a microphone, like she’s using her whole body and to be able to do that, that’s not something that we learn at the age of three, like we learn to be quiet, you know. So all of a sudden you…can see the social barriers that are being broken down. And…a woman playing a guitar really hard is like this physical act, you know you can turn around and take a guitar and hit somebody with it. You know, you can pick up a microphone and hit somebody with it…And so, all of a sudden these things, this

\textsuperscript{362} Thane 1997, Pt. 1; my transcription and emphasis.
music and these instruments become tools and weapons…and our own voice. And I think it’s like the self-defense was like this metaphor or something for like what is already happening. And so for me it’s just kind of pointing out a little bit to women who are already doing really intense things…how powerful what they’re doing can be. (Alice Stagg)

Here Stagg makes a number of salient observations about the performative dynamics of activity and empowerment within the practice she is describing. I take her to be identifying two related ways in which the “girl culture” scenes of the 1990s changed the space of what was possible for women at the time. First, these scenes made room for a new image of what a woman could be by fostering an immediate, visceral impression of female empowerment in audiences. Second, the production of this representation helped to constitute the strength so depicted through the explosive physicality typical of performance within the subgenre. By engaging in a form of activity not yet fully available to them or intelligible as theirs, women in these scenes helped prefigure the possibility of being something which that very involvement constituted—the “fiercely independent homosexual [women]” a suitable image of which “she didn’t have a picture…in her head,” to revisit Juliana Luecking’s words.

So what does all this have to do with the conceptual and phenomenological issues within moral and political psychology we’ve been concerned with? The practices at issue in the Queercore and Riot grrrl scenes of the 1990s have a strongly agonistic and prefigurative character. They therefore help illustrate the possibility and value of a style of public performance and interaction that does not readily conform to the consensus-oriented, deliberative templates insisted upon by writers in the ethical and political tradition with which we contrasted the view of agonistic integrity defended here. They illustrate that an immediate performance can provide the occasion for a futural becoming-
other, an overcoming of presently-felt limitations through the enactment of a confrontation with them.

We need to be careful about the status of the ‘futural’ or prefigurative dimension of the practice described by Stagg and which I argued is an agonistic one. The future-oriented dimension of the activities nurtured by such a practice hardly need to be conceived along voluntaristic lines or as made possible through the exercise of prospective deliberation or reflective control over one’s conduct in the strong sense insisted upon by Korsgaard. Indeed I think they paradigmatically lack these features. What’s more, the possibility and value of the futural dimension of agonistic practice relevant here does not require the present capacity for deliberative control through prospective reflection. Korsgaard tries to qualify her theory of action by restricting its requirements in this way, but I think the attempt fails the phenomenological test on which we have seen she must rest her case:

It wasn’t like, well we wanted to get into this scene or cultivate this sound or we have this big plan and this is what we want to do. It wasn’t like that at all. It was just like ‘Ha ha! Let’s have fun and be stupid and just enjoy that.’

But despite its possibly unreflective character, the sort of agonistic practice in question is not blind. The activity of its participants paradigmatically embodies a form of practical intelligence. It consists less in the capacity for formal abstraction and the formulation of general policies or maxims from an impartial or impersonal standpoint, and more in the ability to live out or make trial with unusual possibilities prematurely, often without a clear or definite sense of precisely what one is doing or becoming, or of what it can or will entail:

---

363 Thane 1997, 18:20 of Pt. 2 of She’s Real; member of Tribe 8
Everybody kind of has touched on the idea of, you know, ‘ok well there’s nothing out there that I can relate to, so I’m just going to create it, I’m going to be it. I’m going to pretend that the revolution has already happened and I’m just existing as I am, who I am.’

Moreover, though, as with Riot grrrl and Queercore, it can be substantially oppositional, the activity is not merely reactive. As we have already seen, it is constructive, productive of new, albeit inchoate or nascent/incipient possibilities of identity and action. In Stagg’s words:

It’s not worth it to be able to say ‘no’ if you can’t go out and get what you want...And I think that’s what a queer identity is about, too. You know, it’s not just about all this political bullshit, it’s about how we have sex.

8: is my view of integrity “horizontal”?

The following worry might arise about what I have said so far about integrity: “Isn’t ‘noise’ in the relevant sense (normative polyvalence, or whatever) just another ‘horizon’”? Am I not saying something like—“One cannot establish any universal, permanent or constitutive features of normative practice because of its universal, permanent and constitutive susceptibility to challenge in various ways”—and isn’t this contradictory?

The response is to bite the bullet with certain qualifications. There is a way in which I am indeed saying something about agency and normativity as such. However, what I am saying is indirect and indeterminate in a way that significantly distinguishes my view from the sort of view of normative practice I criticize, which I call “horizontal.” And it is the basis for my criticism of what I take to be their characteristic ambitions and rationale.

---

364 Thane 1997, Pt. 2, around 22:30
365 Ibid., 18:50, Pt. 2
Recall what makes a view of normative practice horizontal in my sense. Horizontal views of normativity and agency are general accounts of the determinate constraints governing normative practice as such. To motivate my case against horizontal views I focus on some prominent theories of normative practice which I take to be especially illustrative by virtue of their clarity and impact. Korsgaard’s recent work on the nature of action and Habermas’s discourse ethics, respectively moral and political forms of constructivism about value and agency, are cases in point.

Here in a nutshell is the reason why horizontal views are dubious. It is that each such account is potentially properly subject to contestation over any of its determinate features, where these range from the conceptual embellishments that divide adherents down to its supposed basic starting-points. This in turn is because the standpoint furnished by a theory designed to capture the range of relevant phenomena requires articulation at indefinitely many stages throughout its construction (or putative ‘reconstruction’) and defense. Such points of articulation are equally potential objects of appropriate challenge. And this in turn is in tension with the point of such views, which, as we said, aim to hold all the phenomena in a unified conceptual vision by specifying determinate conditions that govern and constitute them as the phenomena they are.\(^{366}\)

Now there is nothing obviously wrong with some feature of something being in tension with other of its features. On the contrary, tension can be of value, and the view of integrity I defend here broadly requires that one face up to and even cultivate tensions in one’s outlook or conduct. In the relevant Nietzschean sense, tensions (and their subjective manifestations as suffering or experienced hardships of various kinds)

\(^{366}\) They entail that unruly, transgressive conduct is not essentially intelligible: sure, but the point I wish to make is that nothing is essentially intelligible, so we can’t generalize here and say that “norms” are what furnish intelligibility \textit{tout court}.\[258\]
condition dynamic life and grown and so can be fruitful if properly cultivated or engaged with. However, I think the tension in horizontal views of agency which I have just described is not a fruitful one but rather one that saddles such views with a bad conscience, also in a Nietzschean sense. They effectively (methodologically and rhetorically?) gainsay the inexhaustible character of the phenomena they would capture by treating it as though it can be captured exhaustively, and this tends to hamper rather than enhance their functioning. The tension manifests itself in a lack of flexibility and openness, and these are symptoms of a lack of integrity as I understand the term.

Nietzsche recognized that philosophical activity carries with it precisely the threat to integrity I am trying to identify:

_Being philosophically minded._ – We usually endeavor to acquire a _single_ deportment of feeling, a _single_ attitude of mind towards all the events and situations of life – that above all is what is called being philosophically minded. But for the enrichment of knowledge it may be of more value not to reduce oneself to uniformity in this way, but to listen instead to the gentle voice of each of life’s different situations; these will suggest the attitude of mind appropriate to them. Through thus ceasing to treat oneself as a _single_ rigid and unchanging individuum one takes an intelligent interest in the life and being of many others.\(^{367}\)

Recall that I think integrity requires taking responsibility for the limits one finds oneself confronted with, but doing so unresentfully, that is, without thereby striving to overcome limits as such (however the transcendence of limitation is specifically framed as something possible and desirable—Korsgaard’s Kantian strategy is one we are by now familiar with). In good Nietzschean form, doing so is a way of saying “yes” to suffering without saying “no” to life.\(^{368}\) After all, if there really is a worthwhile sense of integrity as somehow owning up to one’s own finitude as an agent, as one who can in some way

\(^{367}\) Nietzsche, HATH Part 9, “Man Alone With Himself;” 618

\(^{368}\) Nietzsche embraced the Aeschylean dictum, “pathei mathos,” which held that human beings learn through hardship and suffering: “Who suffers much makes the devil envious, and is thrown out into heaven.” (Gesammelte Werke (14:44))
own up or take responsibility at all, then even theories, general views of agency, *as* exercises of some person’s or persons’ agency, are properly subject to the constraint of integrity so understood. Accordingly, one way of stating my complaint with horizontal views would be say that they exhibit a lack of integrity in this sense.

Now notice that so far I have said nothing about what agency or normativity is like or requires always and everywhere—at least I have said nothing determinate about what a pattern of activity must be like to count as properly subject to normative assessment (whether the relevant sort is moral, epistemic, or whatever). What I have done (and what agonistic critics of much modern political thought have done in the same vein) is make two related assertions: first, I affirmed the in-principle contestability of determinate, informative statements of what normative practice is like as such; second, I assert that efforts at contestation are subject to the same open-ended or ambiguous ‘constraint’ (that of being “contestable”) and so cannot be exhaustively captured either.

There must be an avowed ellipsis folded even into such an account as I give here, but, importantly, it cannot be set out as a further assertion alongside those just enumerated. It may be referred to, as I have just referred to it, but also must be taken up performatively in one’s ongoing dealings with the phenomena at issue as an engaged participant.

9: conclusion: underground overmen?

We began our discussion of integrity in the previous chapter with a preliminary diagnosis of what is wrong with Dostoevsky’s Underground Man. We saw that his psychic life is manic and recklessly episodic. He identifies his own psychopathology in terms of a failure of personal integrity, a failure he characterizes in terms of his inability
to properly engage with the multitude of divergent perspectives that well up from within and that confront him from without. He either refuses to take seriously the changing perspectives to which he is subject, scathingly jeering at himself and his audience, or over-commits to what he takes them to entail, apparently unable to establish practically effective critical distance from his own compulsive behavior.

The Underground Man embodies a dramatically undisciplined Nietzschean psychology. Like any human soul, his is a multiplicity of conflicting drives. But though his activity is not wholly lacking in the “constraint of style” which Nietzsche praises in *GS* Sec. 290 (perhaps no one’s is), his style is bad. This is not necessarily because he has failed at the task that Korsgaard represents as constitutive for any agent: “the task of uniting [our many particular practical identities] into a coherent whole.” After all, under AR, this is not even an intelligible aim.

Moreover, and for essentially the same reason, the badness of the Underground Man’s style isn’t a function of its failure to satisfy a formal requirement of consistency in willing such as the one expressed by the Categorical Imperative of Kantian morality. For AR undermines the Moral Sovereignty Thesis in whose terms advocates of SI defend that requirement. Our discussion of the free spirit helps us to see that the agonistic ethic we are offering contains no general formula for settling the question of the goodness or badness of the Underground Man’s style.

Rather, we can see now that the Underground Man’s style is bad because he lacks agonistic integrity. He is confrontational, but recklessly and fruitlessly so. He lives at odds with the standards of his time—standards of proper dress, romantic courtship, professional success, political influence, and so on—but he does so bitterly and

---

*Korsgaard 2009, p. 21*
impotently. He makes—believes that he can make—no fruitful effort to constitute a viable alternative to the deficient way of life he so vehemently resents. Instead he fatally embraces his wretchedness and ironically makes a self-contemptuous “virtue” out of it rather than attempting to overcome and live beyond it. In short, the Underground Man is stuck underground. His subterranean mode of existence recalls the canny but one-dimensionally reactive slaves of the First Essay of Nietzsche’s *Genealogy* who also gnash their teeth and impotently rage against their galling fetters. But even Nietzsche’s slave type is creative, however poisonous and nihilistic their values have turned out to be. Indeed, their revaluation of the master’s values, the victory of their “good vs. evil” over the old, aristocratic “good vs. bad,” was a milestone in human development that, in many respects, Nietzsche admired in its own right and took as the model for his own attempt to prefigure a new, life-affirming “table of values.” In light of this, it is more accurate to say that the Underground Man reflects one aspect of the slave type—its purely negative, reactive side.

The free spirit, by contrast, is both a creator and a yea-sayer. By embracing conflict, which is the condition of life, she affirms life and embodies her own justification. But her capacity for serious play requires her to be multi-dimensional, endowed with the capacity for criticism and negation so robustly cultivated by the slave type of *Genealogy* I. And like the slaves, the free spirit disciplines and employs her capacity for negation constructively. Recall that Nietzsche concludes *Genealogy* I by exhorting us to “be the battleground” between slave and master values, to cultivate the tension between the reactive and the constructive resources made available by the slave revolt in morality and by “the internalization of man” later addressed in Essay Two. The
free spirit does precisely this, harnessing the creative, “form-giving” energy unleashed by the slavish, reactive affects bred into her through socialization—the bad conscience, in a word—and drawing upon these affects as resources for prefiguring new ways forward.

What sets her apart from the slave type, and ultimately above it in the ethical “order of rank,” is the heroically agonal, and therefore life-affirming, spirit characteristic of the master type. And what sets her apart from and above the master type as it is portrayed in Genealogy I is the complexity and sophistication of the psychic resources she brings to her creative tasks. After all, since psychic depth and complexity are consistently associated with slavish, repressed conditions of life, the free spirit’s capacity for serious play has its roots in that type too. The slave in her is capable of seriousness—of commitment to focused, protracted undertakings; commitment being understood in terms of “the right to make promises” to oneself, to hold oneself to account. The master within the free spirit endows her with the capacities requisite for play—the thirst for adventure and conquest as well as the ability “to forget,” to lose or release oneself into the fray, to assault or abandon one’s own convictions in a spirit of experimentation.

So through her playful but serious, destructive and creative activity, the free spirit embodies and preserves a fruitful tension between the masterful and slavish resources through which the activity defining her is sustained. Fueled by this internal conflict, she plays amidst the wreckage of ideals in crisis, risking herself to expose and exploit their vulnerability to “revaluation.” This committed, adaptive but constructive resiliency is what her integrity consists in. The Underground Man doesn’t even come close to achieving this. He is a shadow of Nietzsche’s slavish caricature, an eternal prisoner of the underground, a specter of pure negativity. This is ultimately why he is right to
observe that he lacks integrity. And even this measure of clarity about his own condition is an enabling symptom of his impotence: the diagnosis comes in the form of an accusation he directs at himself from an imaginary critic whose standpoint his self-imposed isolation from life amidst others requires him to adopt on his own behalf. As such, it is more bitter irony, availing nothing.\textsuperscript{370}

According to Nietzsche’s diagnosis, we are all potentially Underground agents. Human conduct is constrained and enabled by contestation among bearers of different identities. Power is the social and historical condition of the proliferation of meaningful differences. So we are led to ask: what style of engagement with diverse perspectives, with other agents, avoids the dual pitfalls of paralytic overcommitment and self-ironizing triviality to which the Underground Man is subject? What style of engagement can underwrite a seriousness that is not stultifying but playful, critical distance sufficient for taking up new perspectives in good conscience—with the experimental daring to genuinely make trial with them—but which preserves our capacity to do so resiliently, that is, in such a way that we remain engaged with differences that both threaten to confound our efforts and help sustain the context on which we depend in making our efforts? We are wondering about the dynamics of overcoming, the existential possibility of ‘serious play’—Nietzsche’s abiding ethical concern, as well as ours, to the extent that we are asking after the character of a kind of integrity that’s especially valuable when power constrains but conditions our capacity to act.

I have argued that discursively responsive yet agonistic integrity is an attractive candidate for the style of engagement that we have identified. Notably, the Underground

\textsuperscript{370} This detail suggests that, in order to secure integrity of the sort we have argued the Underground Man lacks, the requisite agonistic engagements must have a robustly second-personal dimension.
Man does not engage with his own perspectives or those of others in a way that is satisfactory, in a way that places each party at risk so that the course of interaction remains fruitfully undecided. Undergoing the transformation from Underground Man to Overman requires cultivating social and psychological conditions conductive of the style of interaction requisite for agonistic integrity. Identifying those conditions more precisely would help us to better identify the prospects for agonistic integrity in modern democratic conditions. But that will have to wait for another time.


Wilde, Oscar. Lady Windermere’s Fan, Act I, ???, 1892.
