

THE UNIVERSITY OF CHICAGO

SELF-OPACITY, HUMAN AGENCY, AND ETHICS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE HUMANITIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHILOSOPHY

BY

FRANCEY RUSSELL

CHICAGO, ILLINOIS

AUGUST 2017

For my family.

Table of Contents

<i>Acknowledgments</i>	iv
Introduction: Getting Opacity Into View	1
Part I.	
1. Deliberative Endorsement, Agency, and Self-Knowledge	17
2. “Murkiness is the Rule”	45
Conclusion to Part I	70
Part II.	
3. The Conflictual Unity of Rational Animals	72
4. “The All But Unending Vulnerability of Human Action”	106
Conclusion to Part II	140
Part III.	
5. The Difficulty and Value of Self-Opacity	141
Conclusion: “Something You Can’t Get At”	178
Bibliography	187

Acknowledgments

Throughout writing this dissertation, I have felt so lucky to have such a supportive, creative, and challenging committee. It has been wonderful to work with my advisor, Candace Vogler. When I began formulating my ideas, I read her essay “Sex and Talk,” and it has been a kind of touchstone for me, reminding me to try to go further in thinking of what philosophy can say and how it can say it. I want to thank Jonathan Lear, Daniel Brudney, and Kyla Ebels-Duggan, all of whose philosophical work and conversation has been so incredibly productive and inspiring. I’m very grateful to have been able to engage with each of you as I moved through this project.

I owe huge thanks to everyone in the Philosophy Department at the University of Chicago—the faculty, the students, and the administrators.

My experience as a graduate student was more challenging, more exciting, and more supportive through participating in student activist groups. These groups transformed my sense of what it means to be an educator, a philosopher, and a member of the academy at this moment in history. At the University of Chicago, I want to thank the undergraduate and graduate Women in Philosophy groups, the Graduate Student Union, the Gender and Academia Working Group, the Diversity and Philosophy Working Group, and also People in Support of Women in Philosophy at the New School.

There are so many friends and philosophers and philosopher-friends that I want to thank from the bottom of my heart: Amichai Amit, Pascal Brixel, Agnes Callard, Matt Congdon, Emily Dupree, Jaime Edwards, Brendan Fernandes, Janna van Grunsven, Gregg Horowitz, Grayson Hunt, Katie

Kelley, Thomas Khurana, Claire Kirwin, Netanel Lipshitz, Agnes Malinowska, Dan Morgan, Karen Ng, Rory O’Connell, David Rodowick, and Alex Wolfson. Thank you to my precious friends Vanessa Brown, Andrew Sutherland, Andrea Wong, and Will Widmer and Jeanne Firth. I want to thank Tom Krell for so many conversations about philosophy and self-opacity. I asked him once if he thought self-opacity was always a kind of failure, and he said that it didn’t have to be, that it could be stirring. Throughout writing I have tried to hold onto this insight that our self-opacity can stir us, that it is a key to our sense of life as stirring—strange and important.

To my mom, dad, Frances, and Allie: your support has been so enduring and has come to me in so many forms. You are such searching, curious people, so thoughtful in your ways with the world—thank you for everything, for all the love and conversation and mystery.

Why should the mind be less dense and empty and mazed and pocked and clotted
—and why less a whole—than the world is?
Stanley Cavell

There is perhaps in the end no peace between those who think that morality
is complex and various, and those who think it is simple and unitary,
or between those who think that other people are usually hard to understand
and those who think they are usually easy to understand.
All one can do is lay one's cards on the table.
Iris Murdoch

Introduction: Getting Opacity into View

opacity: from Latin *opacus* 'darkened.'

1. adjective: not able to be seen through, not transparent
2. hard to understand, unfathomable
3. obscurity of meaning

§1. A Scene, in the Dark

Ingmar Bergman's *Scenes from a Marriage* (1973) opens with a couple preparing to give an interview about who they are and about their relationship. After agreeing that they are "not particularly nervous," the interviewer asks them how they would describe themselves in a few words. Johann provides a list of his attributes and considered opinions, a list that begins half jokingly ("I'm an excellent lover") but eventually becomes a more ordinary self-report about his beliefs and commitments. When the question turns to Marianne, she falters: "what can I say? I doubt if I have the same natural appreciation of my own excellence as Johann. But to tell the truth, I'm glad I live the life I do. It's a good life, if you know what I mean. Well, what else can I say... Oh dear this is difficult!" The interview moves forward, with Johann and Marianne confirming that after ten years

they are happy, they get along with hardly any friction, *and* they are “well aware” that the very congeniality of their relationship could itself be a problem, though actually it isn’t.

It is striking to begin a movie about two characters and their long and intimate relationship with an interview. What is our response, as viewers, to this scene of two characters avowing their beliefs and desires? On the one hand, we probably think that if we want to know what someone believes or desires, that person is the right person to ask. On the other hand, it is significant “that we do not accord absolute authority to the explicit avowals of others when it comes to determining what attitudes are appropriately attributable to them” (Smith 2004, 341). I think we regard Marianne and Johann as in the familiar, ordinary position of both knowing and not knowing themselves, all at once, which means that we hear their explicit self-avowals, not with cynical suspicion, but with an awareness that there is probably more going on than they are ready to say. The subtlety, compassion, and humor of *Scenes from a Marriage* derives from the film’s sensitivity to the ways in which human beings occupy and navigate this grey area.

Importantly, the film does not suggest that we should be skeptical about Johann or Marianne. Later in the film, after they have decided to separate, Marianne wonders how things would have gone if they had “always told the truth and not kept anything secret from each other,” to which Johann replies: “did we even know that we kept things secret?” Bergman thus suggests that the characters have not been lying to each other about who they are, and he implies that greater explicit truthfulness about themselves was not what was needed. But if they were not deceiving each other, then maybe these characters were *self*-deceived? Yet though Johann is a little too self-certain, a little too committed to his avowed attitudes, a little too confident about his choices, there is no suggestion that he is actively deceiving himself about anything in particular. Rather than imply that he is self-deceived, the film instead suggests that his very confidence about who he is and his

reasons for acting is importantly connected to a certain lack of compassion, humility, responsiveness to others, and a lack of humor about himself. That is, the film suggests that Johann's confidence in himself and his ability to know himself constitutes a kind of ethical failure.

Marianne is far more tentative. In the first scene she does not evidence such comfort in making confident claims about who she is. Later in the film, after the separation but when she is still seeking to connect with Johann, she asks if he will listen to her read aloud from her diary (a kind of self-interviewing device). In her diary, she has written about the experience of looking at old photographs of herself: "I seemed to be aware of something that had been lying in readiness for a long time but beyond my grasp. To my surprise I have to admit that *I don't know who I am*. I haven't the vaguest idea." Marianne does not seem distressed or overwrought about this realization, she even smiles as she speaks. It is almost as if she feels some relief in knowing that she does not know. Moreover, in wanting to read this passage to Johann, Marianne seems to think that telling Johann this—that she doesn't know who she is—could be a way of telling him something intimate and real about herself. And that this telling is a way of *being* intimate and real with him. But when she looks up, she and we see that Johann has fallen asleep on the couch.

The film moves through various scenes of conversation over many years. Again and again we are invited to take Marianne and Johann's claims about themselves, about what they believe and what they want, at once seriously and not seriously. We are not asked to doubt them or view them with a cynical eye, as though these characters were simply deluded. On the other hand, we are also not positioned to simply accept what they say at face value. Even in their most emotionally charged, most emphatic interactions, where it might seem like they are really "being honest," we know that what they avow doesn't capture all of what they believe or feel. The film asks us to take up what I would call a remarkably humane position towards them: we understand that these characters are

working to figure out who they are and how to live, with themselves and with each other, all the while in partial darkness.

The way the film ends is as telling and suggestive as it's beginning. In a section or "scene" of the film titled "In the Middle of the Night in a Dark House," Marianne wakes suddenly from a dream, initially frightened, eventually comforted, though no less confused. Marianne wonders why she had such a strange dream: "what do you think causes it?" she asks. Johann replies: "in your extremely well-ordered world, there is something you can't get at." So this film moves from an opening scene of alert self-assertion on daytime television, to dreams and disorienting darkness, to something you can't get at. And most crucially: Bergman presents this development as a kind of growth or progress. That is, the increase in darkness and tentativeness about the self, and the easing of the need to be clear and more certain, is presented as somehow humanly good and important. While they were always in the dark about themselves, these two are now a bit more welcoming of this fact: Johann is a little less threatened by it, and has also lost some of his resolute self-assurance; Marianne is no more self-certain, but she has gained a bit of confidence.¹ Both have developed a sense of humor about themselves and about their relationship.

Bergman's vision is realistic and compassionate, deeply moral and not at all moralizing. It moves with what Cora Diamond calls "a sense of the mystery of individual life" (2010, 56). His film suggests that an increase in the capacity to allow for such mystery, in ourselves and others, and an easing of the demand for decisive self-certainty constitute forms of moral progress. That we feel so kindred toward Bergman's film and these characters suggests that we appreciate that we ourselves

¹ Candace Vogler's "Sex and Talk" is extremely illuminating for thinking about the kind of ethical, intimate progress made in *Scenes from a Marriage*. Vogler is interested in exploring forms of sex and forms of talk in which the participants precisely do not communicate their well-defined selves but are rather given an opportunity to be less than full, rational, self-knowing selves. Vogler calls these kinds of connections "depersonalizing intimacies." Importantly, the end of *Scenes*, Johann has gotten a bit better at talking with and listening to Marianne, and Marianne has grown more at ease and confident in her sexuality. I discuss *Scenes* and Vogler's essay a bit more in the conclusion.

are very much like these two humans: working things out in the dark, “all unknown to ourselves” (Cavell 1999, 325).

*

There are (at least) two ways in which it hard to square our response to this film and its characters with much of how we are invited to think about persons within the context of contemporary moral philosophy.

First, according to a dominant philosophical picture of agency, we human beings are for the most part lucid about ourselves, able at any moment to step back and reflect on our attitudes and subject them to rational scrutiny in order to decisively act. On this picture, the pressing practical question is not whether we have understood ourselves, but whether we will be sufficiently resolved to carry out our rationally deliberated decisions. So, there is, as it were, a descriptive mismatch between the kinds of beings that moral philosophy suggests we actually are, and the kinds of beings presented in films like *Scenes*. (This is not to say that philosophy’s picture of the human is not present in the film: to the contrary, we might interpret Johann as working very hard to realize such a form of agency).

Second, this philosophical picture maintains that treating another human being as less than authoritative or lucid about herself constitutes a kind of disrespect: “we not only defer to others’ claims about themselves, we *owe* them that deference so far as we treat them as rational, responsible agents” (McGeer 2007, 82). But in the final scene of the film, Marianne and Johann take each other’s words with large grains of salt, poking fun at each other and at the preposterousness of the idea of taking themselves or each other wholly seriously. And this poking fun is presented as loving, as a way of loving and being good to each other.

Finally, according to this dominant picture of agency, while most of us have not achieved such reflective self-possession yet, we should *want* to realize such an ideal of deliberative agency, to become “perfectly self-aware and so perfectly detached and free” (Murdoch 1999, 340). Perhaps as a matter of fact we are unknown to ourselves, but we should strive to overcome that state in order to achieve the greatest possible reflective and deliberative clarity and *thereby* greater freedom. But this makes it difficult to understand our appreciation of how *Scenes from a Marriage* ends, and the idea that a kind of ethical, human progress has been made.

In this dissertation my aim is to provide a philosophical account of agency and moral psychology that aligns us closer to Marianne and Johann. I think the best way to conceive of the kind of flickering self-understanding that we see evidenced in these characters is to see them as *self-opaque*. Thus my aim is to bring the concept of self-opacity (to be elaborated below) to the center of our thinking about agency and ethics. While it is no doubt true that, in simple cases, we often know what we are doing and why, I will argue that in ethically complex circumstances where our values and self-conceptions are at stake, we are often not clear why we do what we do, we do not understand our motivations, and our aims are opaque. In much contemporary moral philosophy, self-opacity is treated cursorily because it is usually conceived negatively, as a contingent failure of agency and as an obstacle to ethics. I will argue instead not only that self-opacity is our ineradicable condition, but more controversially, that our self-opacity can contribute positively to our ethical lives: our self-strangeness can be a source of ethical insight and flourishing.

§2. What is Self-Opacity?

Self-opacity is a strange idea. It is not something philosophy can get clearly into view with a neat definition; to borrow a term from Rahel Jaeggi (in describing the concept of alienation), self-opacity

is a concept with “fuzzy edges” (2014, 3). Very broadly speaking, self-opacity involves the failure to practically understand oneself, where that can include one’s actions, one’s reasons, one’s attitudes, or one’s motives, *and* where this is not a merely contingent failure: either there is resistance to knowing something (because, for example, it challenges one’s self-conception), or it concerns features of the self that cannot be straightforwardly recognized from the perspective of agency (for example, a sense of my own finitude).

Above I cited three definitions of “opacity:” 1) not able to be seen through, not transparent; 2) hard to understand, unfathomable; 3) obscurity of meaning. The first definition suggests that opacity is an epistemic defect: here I am self-opaque when there is something of myself that I lack access to or don’t know about, as, for example, when I don’t realize the real motive for my actions (I think I am being deferential, but am leveraging control). The second and third definitions move us away from issues of epistemic access to broader issues of self-comprehension: in these senses, I am self-opaque when I cannot understand or fathom something about myself, for example, some attitude or desire or action of mine, or when the meaning of some attitude or experience is “obscure” (this is connected to Diamond’s “sense of mystery”). In these cases, I may enjoy a kind of epistemic access or knowledge—I see what I am or what I’m doing, maybe I recognize that my generosity is not quite so sincere—but I lack the ordinary kind of self-understanding that I enjoy when moving unproblematically through the world, when my attitudes and actions simply make sense.

My aim is to show that self-opacity is a constitutive dimension of our practical lives, rather than a contingent failure. I will analyze self-opacity as tied to two mutually informing aspects of human life: our animality (Chapter Three) and our dependency on others (Chapter Four), which is to say, our basic non-self-sufficiency. As such dependent, rational animals, we live in a space of

meaning and value, where this space is sufficiently complex, dynamic, and indeterminate that there can be serious uncertainty and disagreement about the meaning and value of any stretch of human activity.

Crucial to the idea of self-opacity is the fact that there is more to the self than it could, or should, deliberately endorse or avow or identify with or commit to; there is more to the self than what rationality can redeem. That there will always be more to the self than it could or should deliberately endorse or avow is not just because unproblematic, affirmative, rational self-understanding is an ideal out of reach. It is also because, as I will try to show, sometimes such unproblematic self-understanding is not the ideal to strive for; sometimes the very insistence on such self-understanding can get in the way of ethically significant forms of opaque relations with self and others. Here I am following Iris Murdoch's recommendation that "what we require is a renewed sense of the difficulty and complexity of the moral life and the opacity of persons" (1997, 293).

§3. Motivation for the Project

The motivation for this project initially arose from an almost visceral reaction to two aspects of a dominant philosophical picture of agency, which in Chapter One I will refer to as the Deliberative Endorsement theory of agency (DE). On this picture, human agents or persons are conceived, as John Doris puts it, as "autonomous individuals engaging in reflective deliberation funded by accurate self-awareness" (2015, 164). According to this picture, which focuses on the position and perspective of an agent at the moment of deciding what to do, human agency is exemplified by our capacity to step back from any of our desires, emotions, beliefs, and commitments, subject them to deliberative, reflective scrutiny, and decide what to do. This picture seemed not only too managerial,

like a consumer model of agency, according to which I choose my attitudes and actions as I would choose a product, but too lucid.

The second relevant aspect of this theory is the normative implication of the picture. If reflective deliberation and decision exemplify human agency, then this is also the kind of agency that we are meant to strive for. What “we” should want, on this view, is to be able to reflect on and rationally endorse any and all of our beliefs, desires, and attitudes, to stand behind them wholeheartedly, and make resolute decisions on their basis; “we” do not want any part of who we are to remain in the dark, and we never want to be passive (Korsgaard 1996, 168). Our aim on this picture is to “encompass ourselves completely” (Nagel 1989, 118), to be wholly “self-satisfied,” as Frankfurt puts it, where self-satisfaction is defined as the “absence of restlessness or resistance” (1992, 12). We are guided by an ideal end-point of self-awareness and self-control where we would be, as Murdoch sardonically puts it, “perfectly self-aware and so perfectly detached and free.”²

The Deliberative Endorsement picture of agency and the values it promotes have come to be almost taken for granted in contemporary analytic moral philosophy. It can seem that agency *simply is* actively deliberating and choosing; and that otherwise one is *passive*, i.e., *not an agent*. The values and ideals associated with this picture can seem almost incontestable: who wouldn’t want to be wholly self-satisfied? who wouldn’t want to be perfectly self-aware and so perfectly free? And even if these are ideals towards which we can only strive and never reach, aren’t these the right goals to pursue?

² Of course there are important differences in the normative views of each philosopher, and I will discuss this more in Chapter One. Frankfurt would reject the ideal of free and rational detachment, since he maintains that who we are depends in part on our having certain “volitional necessities,” attitudes that define who we are and that we did not voluntaristically choose. However, Frankfurt also maintains that even though we do not choose or create these contours of our characters, we *want* them to be as they are. So even though we do not in fact choose these aspects of ourselves, we *would* choose them.

The guiding thought for this dissertation is: not necessarily, not always. It would be one thing to complain that the Deliberative Endorsement picture is too demanding. In response to this, advocates of the view might simply reply that ideals *are* demanding, and that this is no demerit. However, I think that in emphasizing these narrow aspects of agency (self-awareness, deliberation, choice, action), such a picture obscures significant aspects of our practical lives and certain ethically significant values and experiences. So my initial, visceral intuition was not that the ideal is too hard to reach, but that there might be reasons *not* to aim to encompass ourselves completely and be wholly self-satisfied, that restlessness might be not only ineradicable but fecund, and that sometimes failing to understand oneself, remaining opaque, can be ethically important. These then were my initial reactions to a dominant and compelling philosophical picture. Let me say something briefly about the relation between content and methodology.

§4. Philosophical Methodology

There are two broad approaches to theorizing about agency and moral psychology. On the one hand, it can seem that we usually enjoy unproblematic practical self-knowledge regarding our reasons for action, that we are not radically in error about ourselves: if you ask me *why* I am acting as I am, I can usually tell you, without observation and without any problem. This way of thinking about agency tends to go with a certain philosophical method: insofar as the first-person, deliberative perspective at the moment of choice exemplifies human agency, *how* we picture and theorize agency must do justice to this point of view. Because agency is essentially reflexive, has a perspective on both its world and itself, we *cannot* in principle understand agency “from the outside,” as if it were just a kind of empirical thing. So the importance of the first-person perspective in deliberation and action must be reflected in any philosophical theory of agency: the connection

between the topic (agency) and the method is internal. So agency must be analyzed “from the inside:” any account of the agent or self must express how the self is *for itself* and be explained in the terms in which it understands itself (see Gardner 2004, 2009).

On the other hand, we know from experience (including an increasing body of empirical evidence³) that we can be wrong about ourselves, especially when it comes to our evaluatively significant actions and attitudes, and that we are not just conscious centers of will but are rather “complex beings... with various dimensions to which [we] lack access” (Sher 2009, 21). This has suggested to some philosophers that agency cannot be analyzed “internally” precisely in virtue of the fact that the agent’s perspective is so often erroneous and opaque. Since agents act for reasons they themselves do not know of or understand, their perspective on what they are doing should not be privileged in our philosophical analysis. Given that the human agent is typically opaque, with an erroneous perspective on both its world and itself, we *cannot* understand agency “from the inside;” it just *is* a kind of empirical thing. This means that agency must be analyzed theoretically or “from the outside:” an account of the agent or self must explain what the self is *in fact*.

What I want to do take insight from each approach, and provide an *internal* analysis of human agency, which respects our self-consciousness and reflexivity, but that nonetheless emphasizes our self-opacity. But this means that we cannot analyze self-opacity as just a mere fact about our species—“we are mammals, bipedal, self-opaque...” Rather, we need to analyze self-opacity as part of the practical dimension of our lives, hence something we have to deal with from the practical perspective. In terms of content, I want to maintain that human reality is significantly constituted both by subjectivity *and* by a lack of awareness regarding the deeper meanings that shape and structure that subjectivity (Lear 1999, 4). In terms of method, I want to maintain that agency

³ For philosophical accounts of agency and self that rely on this evidence see: Arpaly, Carruthers, Doris, Flanagan.

must be theorized internally so as to reflect the self's essential reflexivity and yet in such a way that does not idealize or overly privilege the agent's own self-understanding.

§5. On the Difference Between Self-Opacity and Self-Deception

I want to say something now about the difference between self-opacity and self-deception. While many different analyses of self-deception are available, on any conception self-deception involves acquiring and maintaining a false belief despite available evidence to the contrary, and where the person is motivated to hold the false belief and discount the evidence. Some philosophers emphasize that self-deception involves an *intention* to deceive (see Davidson 1986); others, in an effort to avoid the paradoxes that arise with the idea of intentionally deceiving oneself, maintain a more deflationary view and hold that self-deception is just a species of motivated, emotionally biased belief (see Mele 2001).

Without entering the debate⁴ about how to conceive of self-deception, the relevant points for our purposes is that self-deception is specific and selective, limited in scope, motivated, precarious and unstable, and conceivably correctable. So, using an example from Carla Bagnoli (2012): if a mother is self-deceived about her daughter's eating disorder then a) her self-deception will *specifically* target beliefs and evidence pertaining to the eating disorder, b) her self-deception will *only* target those beliefs and bits of evidence, i.e., will not distort the rest of reality, c) she will not *want* to know or acknowledge the truth, i.e., be motivated not to know, d) there is always a risk of truth or reality "breaking in" and unmasking the self-deception (the evidence mounts, other people

⁴ For what it is worth, I am sympathetic to Bagnoli (2012). She argues that we need to understand self-deception, not as a case of theoretical irrationality (the holding of contradictory beliefs about oneself or the world) but as a qualified form of practical irrationality: the agent is irrational because she does not take available evidence to count as a practical reason (when she should), but the agent is doing something rational insofar as she is trying to maintain the stability of herself. On this view, what one needs to overcome self-deception is not some new information about the state of the world, but a change in one's practical attitude toward the evidence one already has.

start commenting, etc.), and e) it is clear what it would mean for the mother *not* to be self-deceived: she would need to recognize—in either an epistemic or practical sense, depending on your account—that her daughter has an eating disorder.

Self-opacity, by contrast, is pervasive and it is not conceivably correctable because, as I put it above, this condition is grounded in our animality and our dependency, that is, our non-self-sufficiency. Moreover, we live in a space of meaning and value where this space is sufficiently complex, dynamic, and indeterminate that there can be disagreements and uncertainty about any stretch of human activity. As Marcia Baron writes, while it is clear what it would be not to be self-deceived, “it is not entirely clear what it would mean to have figured out what one is about” (1988, 441), what it would be to understand oneself in a complete and untroubled way.

As I will argue in what follows, part of what it is to be a rational animal is to be formed through deep bodily attachment to norms and values in ways that cannot be made reflectively, rationally available (Chapter Three). We are also constitutively dependent on others for realizing our agency in the world, in ways that cannot be made transparent from the first-person deliberative perspective (Chapter Four). That our rational, reflective agency is both supported and outstripped by our animality and our sociality explains why we cannot expect to be transparent to ourselves. And, to anticipate, that our rational agency is outstripped and sometimes even undone by these dimensions of our lives need not be taken as a cause to lament our shortcomings or wish things were otherwise. Rather, I think that our self-opacity, our being only ever incompletely self-encompassed might be essential to a good human life, and that this is made perspicuous by certain experiences (Chapter Five).

§6. Two Initial Objections

Before turning to the arguments, let me raise and respond to two objections that could arise in response to the very idea of focusing on self-opacity as a central concept for thinking about agency and moral psychology.

The first objection raises the specter of self-skepticism: “yes, we can fail to know ourselves in all kinds of ways. But insisting on our self-opacity is like to moving from the fact that we *can* fail to know, to the conclusion that we can *never* really know. But just because we *can* fail to know or understand ourselves doesn’t mean we should think of ourselves as fundamentally opaque. This is a form of self-skepticism, as though anything I might seem to know about myself were somehow doubtful.”

In response to the charge that I am advocating a form of self-skepticism, my response is: yes and no. I am advocating for a form of self-skepticism in one sense. There are plenty of reasons to think that our self-conscious avowed conceptions of self and world tend to be skewed, whether that is due to our propensity to self-conceit, our vulnerability to biases and ideologies, our dependence on forms of cognition that do not rise to the level of consciousness, our inability to know or understand the various meanings and significances of our attitudes and actions. Granting all of this means that we should embrace what Lorraine Code calls “a commonsense, practical skepticism of everyday life” (2006, 224), a kind of readiness to critique along and recognition of one’s own fallibility (I discuss this further in Chapter One). I take this to be one of the generalizable lessons of feminist philosophy and critical race theory, with which any moral or political philosophy needs to contend.

That this form of skepticism is *practical* means that it constitutes a way of being actively engaged with oneself and others; it is a way of acknowledging the complexity of our inner and outer lives, where certainty precisely cannot be hoped for. But that I am advocating for this as a form of

active moral-political engagement means that, in another sense, I am *not* advocating for self-skepticism. Cavell describes skepticism as expressing a “wish to deny one’s humanity” (1999, 109). Here, skepticism involves disengaging and taking up a spectatorial view on one’s own attitudes and actions, and stepping out of one’s worldly practical involvements. So if I am advocating a kind of self-skepticism, I mean it not as a way of disengaging but as a way of engaging in appropriately critical, curious, and creative self and other relations.

A second objection asks whether or not my real target isn’t, after all, with self-*knowledge*. With all this emphasis on critical self-relations, doesn’t this mean that the point is to know oneself better? Again, the answer is: yes and no. In wanting to emphasize not only that we are self-opaque but that there can be value in relating to ourselves as self-opaque, I am not advocating anything like complacent self-ignorance or indifference to self-knowledge. Conceiving of ourselves as self-opaque can be a way of stepping into ethically significant, meaningful self-relations, where this involves the willingness to experience oneself and one’s relation to the world as hard to understand, as not wholly straightforward or obvious, as beyond one’s comprehension. Acknowledging our self-opacity is a kind of self-knowledge, but it is different from knowing something about oneself or knowing oneself through avowal or endorsement or scouring one’s soul for ever more information. Rather, acknowledging our self-opacity means knowing that one cannot ever “encompass oneself completely,” and trying to live well with this.

In all this, I am in deep agreement with Sebastian Gardner’s assessment of the project of self-knowledge as advocated by Richard Moran, to be discussed at length in Chapter One. For Moran, I *know* my own attitudes when I can rationally endorse or avow them, when the question of what I *do* believe or desire is transparent to the question of what I have *reason* to believe or desire. Gardner writes, and I quote him at length:

Moran does not see any inherent limitations to the project of self-avowal. Now one might ask whether this is the whole story. One might think, first, that the demands of personal relationships, or of psychological flourishing, can provide not just psychological barriers to Transparency but reasons for setting limits to its scope. Second, there are philosophical and religious views from which again it would follow that Transparency is in certain domains an abortive or misguided endeavor. Should either prove to be so, then there are elements in and of the self with respect to which avowability in Moran's sense is not proper—objects of self-knowledge that I should not, of course, disavow in the sense of counting them into the not-self, but that equally I should not strive to avow in the sense of 'committing' myself to them. These objects I should regard not as 'up to me' but as transcending the scope of what is properly subject to my deliberation. 2004, 266-267.

My aim here is to precisely to provide a conception of human agency with dimensions that transcend the scope of deliberation and with respect to which avowability is not proper, and to explore the personal relationships and forms of flourishing that provide not mere limits to the project of self-avowal but reasons for setting limits to its scope.

*

This dissertation explores how our non-self-sufficiency—our dependency on our animal nature and our social world—is *registered* from the self-conscious practical perspective of agency. My argument is that it is registered as a kind of challenge to the authority and lucidity of that perspective, a challenge that makes us hard to understand, obscure, sometimes unfathomable to ourselves. This is a challenge that we can be tempted to defend against or avoid, both as philosophers—by insisting on the authority and sufficiency of the deliberative perspective—and as actual agents—through strategies of defensiveness and deflection. But I think it is essential if we are to appreciate the complexity of human agency as worldly and real, and as valuable thereby.

Part I

*

Chapter One: Deliberative Endorsement, Agency, and Self-Knowledge

How much should the first-person perspective matter for agency? There are two levels to this question: in practice, does an agent's own self-understanding enjoy any special privilege? And in philosophy, should the perspective of an agent¹ constrain how we theorize philosophically about agency?

My aim in this dissertation is to find the right way to think about self-opacity with respect to human agency and ethics, *in light of* the importance of the first-person perspective. My sense that self-opacity is relevant to how we understand human agency arose in response to a prominent philosophical picture of agency that regards the connection between agency and self-knowledge as very close. I will call this the Deliberative Endorsement conception of agency (DE). DE begins from the thought that there is a difference between *my* beliefs and desires and the actions that issue from them, and beliefs and desires I simply happen to have. So the question of agency, for DE, is following: how is it that any of these attitudes and actions are properly *mine*, attributable to and expressive of *me*?

As we shall see, according to DE, an attitude or action is properly mine when I can deliberately endorse it. Moreover, this relationship of deliberative endorsement is meant to characterize *both* agency and self-knowledge. This is because, in both cases, what we are trying to understand is a special kind of *first-person authority*, the very authority through which I am the "author" of some attitude or action (it is *mine* or me) rather than a passive bystander to inner events.

¹ Because the authors I discuss here use "agent," "person," and "self" interchangeably, I will as well.

In this chapter, I will present DE and begin to indicate some of the problems it faces and what I think is missing (though the whole of the dissertation will be concerned with that task). Insofar as they take *rational deliberation* culminating in a *decision* to exemplify agency as such, I argue that DE opts for an unduly narrow, exclusively deliberative conception of agency and what it means to occupy a “practical” stance. DE maintains that determining *what to do* (or believe or desire) is *the* question for understanding human life in its practical dimension.

As we will see, maintaining this kind of “deliberative purism” (McGeer 2007) forecloses the possibility of recognizing crucial exercises of agency or “ways of being practical” that are not narrowly deliberative. Moreover, I will argue that the deliberative ideal of wholehearted self-endorsement runs special risks of motivating “defensive” forms of rationality and of generating illusory, defensive versions of the ideal (and, perhaps surprisingly, I turn to Kant for help with this argument). Thus, I claim that DE is insufficiently critical in exalting rational deliberation and decisiveness. And yet, it will also turn out that on a narrowly deliberative conception of the practical, there is no real room for self-critique, for according to DE, this form of questioning self-relation can only be understood as *theoretical*. I conclude that the first-person practical perspective *is* essential for understanding agency, which means that self-opacity must be analyzed as a practical problem encountered within that perspective, and yet deliberative purism cannot be the way to go.

§1. Motivations and Methods

The most fundamental aim of DE is to provide an account of how it is that a person can be not merely a vessel of or bystander to efficacious inner forces or mental facts, but an agent, either with respect to what she does or with respect to her own mental life. In presenting this account, I will focus on Christine Korsgaard, Harry Frankfurt, and Richard Moran. But given that the former two

are focused on agency, while Moran is concerned with self-knowledge, why does it make sense to think of all three as engaged in a common project?

The key idea connecting all three thinkers is that agency *and* self-knowledge are both exercises of “first-person authority.” When is someone an agent? These thinkers will argue, broadly speaking, that agency involves standing in a certain kind of active self-relation and deliberately endorsing one’s attitudes and actions: agency is manifested when the agent herself deliberately determines what to do. And when does someone enjoy self-knowledge? Here too, self-knowledge involves standing in a certain kind of active self-relation and endorsing one’s attitudes: self-knowledge involves the agent herself deliberately determining what she believes and desires. So the same answer is given in response to questions about both agency and self-knowledge, because both agency and self-knowledge are exercises of first-person authority.

Before turning to the task of spelling out this account in more detail, I want to frame this discussion in terms of philosophical method. Because these thinkers maintain that agency and self-knowledge involve a special kind of authoritative self-relation, they give special normative privilege to the first person practical perspective. And more specifically, they model their picture of agency on the perspective of an agent self-consciously deliberating about what to do or desire or believe: this is their exemplary “scene” of action.

I will eventually argue that DE’s conception of this standpoint is overly narrow, but what I want to call attention to now is that it is essential to DE that any philosophical understanding of what human agency essentially *is*, must be articulated in terms of what agency is actually like *from* that very point of view. The idea is that agency (and selfhood and personhood) is the kind of things that must be understood, not from an observer’s perspective, as if it were a kind of given object, but “from the inside.” As Sebastian Gardner writes of this method: “the philosophical picture of the

self is intended to match the self's own picture of itself" (2004, 252). This means that we can conceive the connection between topic and methodology as internal. It is important to point out that while DE seeks to provide internal criteria for agency and self-knowledge, these criteria are supposed not to be merely subjective; that is, they are not meant to depend on individual preferences or aims (Jaeggi 2014, 41). As we will see, however, it is possible to understand and critique DE's normative picture of agency is that it may be offering a particular, substantive, and optional ideal under the guise of a self-evident truth about the basic structure of agency.²

As I said, DE privileges the first person or practical perspective, where this perspective is modeled on the point of view of an agent self-consciously deliberating about what to do or desire or believe, where this scene is meant to make vivid the difference between being an agent and being a bystander of one's attitudes or actions. As Korsgaard writes:

I must be able to *see myself* as something that is distinct from any of my particular, first-order, impulses and motives, as the reflective standpoint in any case requires. Minimally, then, I am not the mere location of a causally effective desire but rather am the *agent* who acts *on* the desire. 1996, 227.

So, there is a difference *from the first person perspective* between merely having some motive or desire and properly authorizing that desire, either by acting on it or by identifying with or endorsing it. The latter kinds of active intervention, as we'll see, are what make any attitude or action properly *mine*.

Notice already three features of the account: first, there is a very strong distinction drawn between what is properly *me* and what is *not me*, where this tracks an equally strong distinction between activity and passivity. On this picture, any attitude that I *qua* agent do not deliberately endorse is just thereby "not me," does not reflect my agency but is rather a brute force to which I

² See for instance Cheshire Calhoun's "What Good Is Commitment?" (2009) where she argues that the idea that agency essentially involves shaping one's life around "commitments" is better understood as an optional *style* for managing one's diachronic existence. It is a "style of life," where this style is well-suited for certain dispositions and situations, but not "agency as such."

am passively subject. Second, thinking about agency from the first person perspective at a moment of deliberation can make seem as though the agent is, or sees herself, distinct from her various attitudes and must join up with them *by way of* some expression of agency; so none of “my” attitudes are *me* until I choose or *endorse* them as mine. Third, analyzing the practical point view through emphasis on choice and endorsement means that, on this picture, “the practical point of view [is] *defined* by our interest in the deliberative question of what we ought to do” (Wallace 2006, 153). Deliberative questions of what to do are thus taken to exhaust the kinds of questions we can pose from the practical point of view.

§2. Self-Conscious, Deliberative Endorsement, and Agency

DE grounds its conception of agency in the fact that our mind is self-conscious in the sense that it is essentially reflective. Advocates of DE are careful to note that self-consciousness need not imply any kind of self-transparency or introspective infallibility. Instead, self-consciousness names the basic structure of the mind such that any conscious relation to the world at the same time involves a reflexive self-relation, not in that there is an additional instance of knowing, but in that “being conscious is identical with being self-conscious. Consciousness *is* self-consciousness” (Frankfurt 2005, 162).

For DE, this kind of self-consciousness is essentially a consciousness of the *grounds* of our various attitudes and actions. For Korsgaard, self-consciousness constitutes the capacity to “step back and reflect” where this “capacity to turn our attention on to our own mental activities is also a capacity to distance ourselves from them, and to call them into question” (2013, 79). So, for instance, if I have some desire or am disposed to form some belief, that my consciousness of this is *self-conscious* means that I must decide whether there is reason to act on this desire or maintain this

belief. As self-conscious agents it is not enough that we simply have desires and beliefs, we have to “take control” of them by giving them grounds.

Now it is not immediately clear why the capacity to step back from and attend to one’s own mental states should have anything to do with agency. It might at first seem that this kind of reflective distance constitutes a kind of passivity, as though the person were a spectator of her own attitudes rather than their source. So what does being self-reflective have to do with the active, authoritative self-relation we are trying to understand? If DE defines agency in part by contrasting the relevant kind of self-relation with a passive or spectatorial form of self-relation, then we need to see taking reflective distance could contribute to the former rather than the latter. Notice also that this contrast is far less intuitive when we are thinking about self-*knowledge*. For it might seem that knowing oneself or one’s attitudes means standing in a kind of introspective epistemic self-relation, where this would seem to involve a kind of passivity: in order to know what I think or want I need to look and see. But, as we will see, this is precisely what Moran wants to reject. For Moran, self-knowledge is not knowledge of an object that just happens to be the same as the self who knows it. Rather as we shall see, knowing what one believes or wants involves *determining* what one believes or wants, for reasons. Such first-person authority is not a matter of being an expert on or especially proximate witness to what goes on in my mind; rather I am such an authority because it is *my* mind. As he writes, “the fact that I not only have special access to someone’s mental life, but [the fact] that it is mine, expressive of my relation to the world, subject to my evaluation, corrections, doubts, and tensions” (2001, 37). So the question now is: how should we understand this special form of authoritative self-relation by which I make something *me* or *mine*?

Frankfurt suggests that in addition to the formal kind of self-consciousness outlined above, there is “another sort of reflexivity or self-consciousness...[namely] that we *care* about what we are”

(2005, 163). The capacity for higher order or reflective attitudes is then not an epistemic ability to take first-order attitudes as objects for observation, but more fundamentally the capacity to be invested in and concerned with the state of one's attitudes. How we are *matters* to us, where this caring is essentially connected with a sense of agency, that how things go is not something about which I wait and see but something I actively determine, *because* I care.

Advocates of DE hold that the relevant non-spectatorial forms of concerned, responsible self-relation are *endorsement* and *identification*. These are the ways in which we make some attitude or action our own. Recall again that on this model, agency *is* deliberative (Wallace 2006, 158). Because of this, endorsement and identification are properly practical only when they stand as the culminations of a deliberative process: I recognize that I desire something, I deliberate about whether I should desire it or whether I should do something about it, and then I endorse or identify with the decision I make. Endorsement and identification have three main functions for agency: they provide intra-subjective organization, volitional resolve, and normative authority.

§3. Endorsement and Identification: Organization, Volition, Authority

(1) Endorsement and identification are modes of relating to one's own attitudes and actions that function to *organize* and unify the self, so that certain contents count as *me* and others are *not me*. Advocates of DE begin with the assumption that persons are complex and that agential unity is not given but must be actively achieved; it must be *my* achievement. Korsgaard, for instance, holds that our most basic task as agents or persons is “the ongoing struggle for integrity, the struggle for psychic unity, the struggle to be, in the face of psychic complexity, a single unified agent” (2009, 7). While someone may have, as a matter of contingent fact, various desires, beliefs, commitments, etc., their hanging together in a coherent whole—as a coherent self—is a function of the agent's activity,

her endorsements and identifications. If you endorse some attitude or identify with some role, “then it’s you” (Korsgaard 2009, 23). As Frankfurt, writes, “it is these acts of ordering and rejection—integration and separation—that create a self out of the raw materials of inner life” (2005, 170). Of course, we may not ever be wholly free of inner conflict, but “with respect to any such conflict, the [agent] himself [must] be fully resolved...he has to know where he himself stands” (Frankfurt 1999, 100). Absent this kind of resolute self-ordering, we are nothing more than, as Korsgaard puts it, “a mere heap,” and this threatens to “destroy one as a person.” So, exercises of endorsement and identification organize what would otherwise be a “mere heap” into a proper self.

(2) In identifying with or endorsing some attitude, the agent makes a “wholehearted decision” (Frankfurt 2005, 172) and thereby knows where he stands. This not only settles the organizational question of what is *him* and what is an external or alien “*force* with which he does not identify” (ibid., 165), it also imbues this organization with *volitional resolve* (ibid.). The connection between organization and resolve is on display in his passage from Korsgaard:

I decide to get some work done...I will certainly want to stop...Timidity, idleness, and depression will exert their claims in turn, will attempt to *control or overrule* my will...Am I to let these *forces determine* my movements? At each moment I must say to them: “*I am not you; my will is my work.* 2009, 69-my emphases.

In this passage, Korsgaard identifies with one of her attitudes (her decision to work) and must, as it were, hold down the fort against imposing forces that attempt to control and overrule her. In rejecting or dis-identifying with these forces, she not only places them in the organizational category of *not me*, but exerts volitional force against them. In another passage, Korsgaard makes even clearer how the organizational and volitional dimensions of agency work together:

We must *repress* [these alien forces] in order to be one, to be unified, to be whole. We must repress them in order to maintain our personal or practical identity. And we must do that in order to maintain our agency itself. 2009, 26-my emphasis.

In emphasizing repression, Korsgaard makes explicit that psychic, agential organization is a *dynamic* organization, where order must not only be achieved but also continually maintained against the threat of disintegration. Frankfurt also occasionally speaks in favor of repression, as when he writes that we may need to “introduce barriers of repression and inhibition” between motivations that “are dangerously antithetical” to our intentions and conceptions of ourselves (Frankfurt 2006, 10).³ Recall now that DE analyzes agency from the position of deliberator. This means that endorsement and identification are modeled on kinds of decisions, in both the organizational and the volitional sense. As Angela Smith writes, “this model...takes the standpoint of practical deliberation as the fundamental standpoint of agency, and locates the activity of ‘the person’ in the activity of conscious choice or decision” (2004, 332).

So endorsement and identification organize the person into what is properly *her* and what is merely present in her in fact, where such organization is maintained through volitional resolve. The question that now arises naturally is: by what *right* are these acts of ordering and rejection undertaken? The exemplary scene of agency for DE is, as we’ve seen, a moment of conflict in need of deliberative resolution: while I have decided to work, other “alien forces” threaten to overrule my will. But how does one know which attitude to endorse as *one’s own*? To ask this is to ask about the *principles* governing patterns of identification and endorsement with reference to which any given decision can be justifiably made.

(3) While identification and endorsement target first-order attitudes, they target such attitudes *through* broader reflective patterns of care and commitment. Korsgaard calls these “practical identities” and for Frankfurt this constitutes “character.” These provide the ways in

³ Though at other times Frankfurt explicitly rejects repressive conceptions of unity: “psychic unity obviously cannot be achieved by dividing oneself” (1999, 106).

which an agent *conceives* of herself, such that for any relevant deliberative choice, part of what is being decided is how *this* choice fits into her broader identity. Such an identity provides a pattern in both a synchronic and a diachronic sense: it provides “some degree of persistence” (Frankfurt 2005, 84) and “carries me into the future” (Korsgaard 1989, 113). In these ways, who I take myself to be rules in and rules out candidates for endorsement and identification.

Now, from all that has been said, practical identities could be as arbitrary as the decisions they are meant to authorize. So how do these patterns provide the sought-for authority? We can differentiate types of answers to this question with reference to the role of reason. For Frankfurt, one’s character is a function of what one loves or cares about, where this constitutes a configuration of the will, providing not just organization and resolution but authority (1999, 138). Yet while such identities and cares *create* reasons, they need not themselves be rationally grounded. As he puts it, “we can identify with various psychic elements, and we can be wholehearted in various thoughts and attitudes, without having any reasons for doing so. On the other hand it is in virtue of these states of our wills that certain things will count for us as reasons” (Frankfurt 2002, 218). For Korsgaard and Moran by contrast, endorsements and identifications are rationally grounded, and for Korsgaard, they are made in light of a practical identity which is itself elected for reasons. As Korsgaard puts it, “your identity is never just a given...how you construct it is really up to you” (2011, 79). While Korsgaard and Moran set different bars for what counts as rational, the point is that for these thinkers, the identities that filter narrower decisions must themselves survive rational scrutiny. If such an identity does not survive such scrutiny, it must be rejected, and if it cannot be rejected, then it is something with respect to which one is fundamentally passive.

So one is an agent insofar as one endorses or identifies with various attitudes and actions, based on reasons that flow from one’s practical identity. But how are endorsement and

identification relevant for self-knowledge? While these forms of self-relation might seem plausible expressions of *agency* (though perhaps narrowly deliberative or volitional; is it really “up to me” to “construct” who I am?), it seems harder to understand how endorsement and identification could be expressions of self-*knowledge*. Mustn't self-knowledge involve cognition of what is there, who I am in fact?

§4. Self-Knowledge as Deliberative Endorsement

Moran's central proposal is to understand self-knowledge as a manifestation of first-personal authority, the same authority involved in agency. He thinks that in wanting to understand how someone *knows* his own mind, what we really want to know is what it means for someone to have special *authority* regarding his own mind. Why is *he* the person to ask if we want to know what he believes or desires? Moran thinks that there are two kinds of authority involved in self-knowledge, epistemic and practical. Epistemic authority means that a person knows his own mind *immediately* and does not have to rely on evidence or inference, as he would with someone else. Practical authority involves *authorial* authority, and this “concerns the authority of the person to make up his mind, change his mind, endorse some attitude or disavow it. This is a form of authority tied to the presuppositions of rational agency” (2001, 92). It is this latter form of authority that is properly first personal, in the sense that it is the unique authority expressed insofar as one *actively is* the mind to be known rather than a well-placed spectator.

For Moran, then, self-knowledge is not an instance of theoretical, object-oriented knowledge applied to the self but is a special and distinctly first-personal *form* of knowledge gained not by observation but by deliberation and endorsement. To see this, consider again what it means to think of one's attitudes as one's *own*. It seems that beliefs and desires don't just sit in my mind like stones,

for me to know or not. Rather, they are *my* beliefs and desires, which means I am responsible for having them, in the sense that it is my own rational activity that establishes and maintains them. For Moran, knowing what my beliefs and desires are involves, not gazing at some inner realm, but, in the paradigmatic case, deliberately deciding what I have reason to believe and desire.

Insofar as self-knowledge is achieved, not by reflecting upon attitudes that are fixed and given, but by deliberating about the reasons that speak for those attitudes, Moran thinks that questions about what I *do* believe are “transparent” to questions about what I am *to* believe (2001, 63). That is, theoretical questions are transparent to deliberative questions. Transparency is both a condition for proper self-knowledge and first-person authority *and* names the method by which questions of self-knowledge are considered and answered. For Moran, while theoretical inquiries involve observation and they terminate in a description, deliberative inquiries involve the consideration of reasons and they terminate in an avowal, endorsement, identification, commitment, or decision. If reasons recommend rejecting or disavowing some attitude, this is either sufficient to be rid of it *or*, if the desire persists despite the rational verdict, then, for Moran, I am just thereby alienated from it: I do not identify with it, it is *not me*. And because in this case I am unable to actively “make up” my mind, I am passive with respect to it and know it theoretically as if it were a kind of object.

Now, it may seem controversial that on Moran’s picture questions of self-knowledge are taken to be transparent to questions of first-person rational authority, and questions of theoretical self-knowledge are transparent to deliberative questions. It may look like we are losing exactly the *knowledge* component of self-knowledge. If I know my mind by making it up, doesn’t this suggest that I can fix it according to whatever I think reason demands? And yet isn’t there something to which my knowledge is meant to be responsive and that may in fact diverge from whatever I

conclude is most reasonable? For example, if I am trying to understand what I really desire, isn't this a quite different question than trying determine what I am *to* desire, given the reasons? Is there not conceptual space to prize apart what I desire from what is desirable, and does it not seem that knowing the former (regardless of the latter) is what counts as proper self-knowledge? (see Lawlor 2009).

While Moran is explicit that his view does not commit him to voluntarism (the idea that I can simply stipulate my attitudes at will), he is willing to bite the rationalist bullet. In cases where I have some belief or desire Y in spite of deliberating about and accepting that the reasons count against it, then while I might know *of* that attitude, I do not enjoy real self-knowledge *because* I do not exercise first-person authority. In this case, my attitude sits there like a stone, indifferent to my reflections and deliberations. I am simply “given over” it and insofar as I see the reasons counting against it, I cannot endorse it. In this case I only know my attitude theoretically or third personally, where “this information may as well be about some other person, or about voices in [my] head” (2001, 93). So this is knowledge of “an interior occurrence to which [I am] passively subject” (2001, 58), but it is not proper self-knowledge.

With respect to any mental state known in this way, the attitude is one I am essentially passive with respect to; or put even more strongly, it is “inflicted” upon me (2001, 117). For Moran, this kind of non-deliberative and thereby non-authoritative knowledge of one's attitudes is derivative, typically pathological, a kind “infirmity,” as he puts it. As Victoria McGeer notes, for Moran a person's rational activity is “compromised to the extent that his psychological states and processes do not ‘spontaneously’ conform—i.e. through deliberation—to the dictates of reason” (2007, 91). Once compromised in this way, the relationship is thereby theoretical, and my attitude is like a stone.

As should be clear, there is a strong normative, even moral connotation to Moran's picture. The problem with occupying a non-deliberative stance with regards to one's own mind is that it fails to exhibit proper rational authority, and for Moran this thereby constitutes an evasion of responsibility (2001, 187-194). Taking a theoretical stance involves treating one's own mental life as simply given (like a stone), as though there were nothing one could do about it. Moran goes as far as claiming that any non-deliberative or observational knowledge necessarily constitutes a form of alienation, and even a form of self-opacity (2001, 57). If, for Moran, proper self-knowledge involves resolution and decisiveness, this means that any failure of self-knowledge is in fact an expression of *indecisiveness* (ibid.). So if I cannot know my mind *by* avowing it, I do not really know myself, where this is explained by my being indecisive. I want to flag here that I think Moran is right to conceive of self-opacity as a *practical* problem, which is to say, a problem proper to the practical dimension of our lives and so something that concerns how we live, though I will want to reject his more narrowly deliberativist approach to the practical. This means that I do not think that self-opacity needs to be analyzed as a form of indecisiveness calling for greater resolve.

We find a very similar idea in Frankfurt, who describes ambivalence as a form of "ignorance or uncertainty" (1999, 100). For both thinkers, volitional failures and epistemic failures go together, and for both thinkers, such failures call for a decisive resolution. Where Moran opposes self-opacity to self-endorsement, Frankfurt opposes ambivalence to wholeheartedness or "self-satisfaction," which he defines as an "absence of restlessness or resistance" (ibid., 103). When one is self-satisfied, one knows what one wants and where one stands.

For Frankfurt, to be self-satisfied, to have "no ambition for improvement" (ibid., 104). This involves being wholehearted in one's feelings, intentions, and thoughts, and this, Frankfurt asserts, is what "all of us" want (ibid., 106). So for Moran, Frankfurt, and Korsgaard, being able to endorse

and avow all of one's attitudes is the ideal to aim for, not only as agents, but as the ideal for a good human life. So there is a great deal of normative baggage packed into this deliberativist picture, where what is "good" or "wanted" or conducive to "psychic well-being" is presented as though it were not only obvious, but universally applicable, for all persons at all times.

*

We have before us a compelling account of agential authority, where this authority is expressed in an agent's determining of her own attitudes and actions. Part of what is meant to make the argument compelling is its apparent familiarity and its promise. We know from experience what it is like to face a moment of decision like the ones described by DE,⁴ and it can be encouraging to imagine that our agency is paradigmatically expressed in these forceful exertions of individual or rational resolve. So it can seem as though agency just *is*, at its best, *like this*, and that being wholeheartedly resolved is what "all of us want."

It is important to notice, with the ideals of wholeheartedness and self-satisfaction, that DE is trying to provide not only a philosophical picture of what agency is, but also a formal conception of human flourishing meant to be both psychologically accurate and intuitively appealing: who among us wouldn't want to be wholly self-satisfied, wholly free of restlessness?

The rest of the dissertation will be an extended engagement and critique with advocates of DE. There is much that is worth keeping: I think that DE is right that we need to understand agency "internally," from the practical perspective, and that, from this perspective, self-opacity constitutes not a mere empirical fact about us but a practical problem that we have to learn how to live with (and as we shall see, this need not mean: simply resigning ourselves to an unfortunate part

⁴ I should qualify this idea that DE is "familiar." While I think we know what it is like to step back, deliberate, and decide, this is hardly how we act most of the time, and yet we would still seem to qualify as agents. So DE's phenomenology is familiar and yet so demanding as to seem absurd. As John Kekes puts it simply: "I act from time to time...yet I have never in my life gone through the process Korsgaard claims is necessary for action" (2011, 338).

of our lot as finite). I want now to argue that in focusing on deliberation and resolve, DE is insufficiently attentive to the specific risks that come with championing deliberative agency and self-knowledge. For creatures like us, there are reasons to be cautious about adopting DE's ideals. I will then show that DE's commitment to a deliberatively purist framework makes it impossible to recognize certain expressions of agency or practical elements of human life *as* practical, simply because they are not narrowly concerned with deliberation and decision-making. I will show that we need to make room for *critical* forms of self-relations that are not deliberative but no less practical or agential. This is precisely the mode of agency and engagement needed in order to curb the risks of the deliberative ideal.

§5. The Risks of Deliberative Purism

We have seen that DE's guiding ideal is what Frankfurt calls "wholeheartedness" or "self-satisfaction," which he defines as an absence of restlessness or resistance. For Korsgaard, this ideal is spelled out in terms of agential unity (one has "pulled oneself together"), and for Moran this means enjoying deliberative self-knowledge: knowing one's mental states, not observationally, but in the mode of avowal. This ideal can initially seem so attractive and intuitive that it is difficult to fathom why or how one would question it. Would giving up the project of wholeheartedness mean that we should *want* to be restless or see some kind of value in restlessness? Does allowing for self-opacity mean letting parts of ourselves remain in the dark? If we check the imperative to "encompass ourselves completely," as Nagel puts it, does this mean accepting that there are parts of me that are not *up to me*? In fact, this is exactly what I think. For the remainder of this chapter, I will argue that we need to better understand the risks to which we are especially vulnerable once we adopt the deliberative ideal, and that narrowly deliberative exercises of agency will not only be insufficient to respond to these risks, but can even encourage them.

The risk, to put it concisely, is that the promise of wholeheartedness can motivate illusory, defensive, ersatz versions. So it is not quite that the ideal is too demanding, too out of reach, but that it is almost too attractive. Here one might wonder why it is worthwhile to dwell on this possibility: isn't worrying about illusion just like worrying about the mere chance of error when considering the aim of securing knowledge? It can seem a bit panicky, a bit skeptical, to insist on this. But note that there is a difference between the mere possibility of error and what we could call a *motivated* tendency.

As finite, imperfect beings, it is of course possible that we could make an erroneous theoretical judgment, but this need not detain our attention: it is a possibility that cannot be ruled out and simply follows from our being finite, and it is not something we are positively attracted to. However, a *motivated* tendency is something towards which we are positively prone. As such, motivated forms of error or illusion are not mere possibilities but are something we have special reason to watch out for. In asserting that we humans are at risk of creating defensive, ersatz versions of avowal or wholeheartedness, the point is not that it this is just a possibility, but that it is the kind of error that positively attracts us because of the promise of self-satisfaction that it holds. An advocate of a demanding agential ideal may try to defend it by asserting that even out-of-reach ideals are acceptable if we bear in mind that they are *merely* regulative, something towards which we must continually strive. I am trying to bring out that pursuing such regulative ideals may require a special critical, attentiveness *given* the propensity to defensively imagine that we have "already made it." Kant, for example, saw that human reason has a special difficulty bearing in mind the merely regulative status of its ideals, that it has an internal, natural tendency to illusion, and that precisely in

light of this, reason needs *critique*.⁵ The point for now is that however we conceive of an ideal of agency, it must be formed in light of an appreciation of such positive proneness to illusion.

In just this spirit, David Velleman (2006) argues that Frankfurt offers us not only an unrealizable ideal but an essentially *defensive* ideal, a kind of motivated fantasy. Velleman argues that we may be inclined to accept Frankfurt's ideal because the promise of achieving such a condition—an absence of restlessness, with no need to change—would satisfy a deep human *wish* to be free of conflict (which is almost to say: to be free of the human condition). The charge here is that as a fantasy, the ideal of wholeheartedness can compromise agency because it is so apt to motivate defensiveness, repression, and self-deception.⁶ So the problem is not that it is an ideal impossible to realize but that the uncritical pursuit of it renders us especially vulnerable to certain kinds of pathologies. Velleman writes:

wholeheartedness is an object of wishes that do not necessarily represent a healthy trend in our thought. Our attraction to the idea of being wholehearted is one manifestation of the fears that move us to defend ourselves against our own emotions. Hence our affinity for Frankfurt's ideal may not indicate that he's right about the constitution of the self; it may indicate no more than our own defensiveness. 2006, 346.

⁵ I will discuss Kant in a moment, but we can look to the first *Critique* for an example of the mere and motivated possibilities: while the possibility of making erroneous judgment is given no real special treatment at all, the entire Dialectic (“the logic of illusion”) as well as much of the Appendix are concerned with a positive propensity “in the very nature of human reason” to mistake subjective conditions for objective conditions, where this propensity is both internal to reason and unavoidable. The point, for our purposes, is that there is a difference between the possibility of error and the propensity for illusion, and this difference accounts for need for philosophical attention in the latter case.

⁶ The idea that DE promotes a defensive ideal can be understood with reference to Freud: according to Freud's early formulations of the ego's activity, the ego is concerned primarily with defense. That is repressed which is judged incompatible with the ego: for example, in 1910 Freud writes: “the incompatibility of [a] wish...with the patient's ego was the motive for the repression; the subject's ethical and other standards were the repressing forces...which [were] thus revealed as one of the devices serving to protect the mental personality” (2008, 24). As Hans Loewald observes, “Freud's first conception of the ego was that it represents the repressive, defensive agency within the psychic apparatus” (1951, 10). Insofar as DE emphasizes the organizational work of identification and endorsement, which separates what is properly “me” from what is an “outlaw” or “not me,” it seems that DE's conception of self mirrors Freud's early, defensive conception of the ego.

That is to say, the ideal of wholeheartedness voices a defensive and wishful fantasy for self-satisfaction rather than a merely internal standard for agency.⁷ As Velleman puts it, “my worry is that Frankfurt’s conception of the self appeals to us only because its implicit ideal represents us as we wish we could be” (2006, 341). Again, the point is not that a conception of agency should not provide ideals that may be difficult to achieve, but that a) any such conception should not be tailored to reflect and satisfy our fantasies, and b) such regulative ideals bring special risks with them for creatures like us, and so there a special need to be aware of the propensity to illusion (where there is not such need in the case of a mere possibility of error).

In response to Moran’s ideal of self-knowledge as deliberative endorsement, Jonathan Lear and Victoria McGeer each offer structurally analogous criticisms. Their charge is that as a “deliberative purist,” Moran is insufficiently attentive to our human propensity to engage in “warped” exercises of rationality, ways in which we can use reason to mask and facilitate tendencies towards illusion and self-deception. Focusing on deliberation and resolve overlooks the possibility that one’s deliberations, resolutions, and practical identities might be essentially defensive, distorted and distorting. So, for instance, McGeer describes a character whose deliberative conclusions are spontaneously psychologically effective—that is, he realizes Moran’s deliberative ideal—and yet these deliberative exercises turn out to be defensive rationalizations: “[this character] is particularly adept at putting his reason to work in the service of a feature of his psychology to which he is determinedly blind—viz., his immense need of something important and predominating” (2007, 97-8). So while he effectively deliberates and resolves, it is in the service of promoting and maintaining

⁷ Frankfurt occasionally speaks in favor of repression, as when he writes that we may need to “introduce barriers of repression and inhibition” between motivations that “are dangerously antithetical” to our intentions and conceptions of ourselves (Frankfurt 2006, 10; see also Poltera 2010, 297). At other times, however, he explicitly rejects repressive conceptions of unity: “psychic unity obviously cannot be achieved by dividing oneself” (1999, 106). Korsgaard also is not clear about how she conceives of repression, since later in *Self-Constitution* she contrasts her favored picture of agential unity with a form of unity that depends on repression (2013, 166).

a fantasy of which he could never become aware—let alone begin to dismantle—by deliberative means alone.

The philosophical celebration of, near obsession with, the merits of self-conscious deliberative reflection is usually attributed to Kant's lasting influence. Both Lear and McGeer connect the shortcomings of Moran's account to his allegiance to the Kantian tradition, according to which agential freedom finds its most exemplary expression in self-conscious reflection and deliberative autonomy (Lear 2004, 452; McGeer 2007, 93), where this tradition not only privileges this capacity but is insufficiently critical of it or aware of its destructive potential. Yet while the "Kantian" tradition may be overly deliberative and evidence uncritical faith in the powers of human reason, we do well to recall that Kant himself was acutely aware of the illusions to which human reason is naturally and unavoidably prone and to the risks of defensive reflection. For Kant "self-conceit" constitutes the illusion to which human reason is positively prone in the practical realm.

In the third chapter of the second *Critique*, Kant introduces the idea of self-conceit (*Eigendiinkeel*): a perversion of self-love, a positive tendency of human reason, and a form of illusion (*Wahn*). The point of Kant's discussion of self-conceit is to specify precisely what the moral law must "strike down" in order to operate as a motive for action. For our purposes, what I want to bring out is that self-conceit is neither a mere error nor some kind of "alien" desire that interrupts rational reflection from without or undermines volitional resolve. Rather it is a way of engaging the method of rational deliberation and reflection for the purposes of self-defense.

For Kant, self-conceit is a form of self-certainty and self-satisfaction that involves the (faulty) judgment that one's personal interests qualify as universal, legislative, justificatory reasons, both for oneself and for others. Self-conceit involves asserting a particular claim regarding one's own worth *as if* it were an objective rational judgment, where crucially, the propensity to make such

faulty claims resides in human reason itself (5:58). In this way, self-conceit maintains that its judgments are not just right and good (a claim to validity), but uniquely so (a claim to unconditioned legislative authority) (5:73). So self-conceit not only prioritizes its own interests and overestimates its capacity for rational judgment, it employs the latter capacity to advance the former under rational cover. Regardless of what we think of Kant's moral philosophy, it is significant that, *contra* the "Kantian" tradition, in his moral psychology Kant is more concerned with our tendency to defensive rationalization and the propensity to illusion than the possibility of weakness or lack of deliberative commitment, and he cautions against our attraction to the seductive experience of self-certainty.

Kant is very clear that self-conceit is not simply an erroneous judgment but an "illusion" or "delusion." The German word is *Wahn*, which Kant describes in the *Critique of Judgment* as "dreaming according to principles, raving with reason" (1987, 275), and which is related to both madness (*Wahnsinn*) and mania (*Wahmwitz*). Whereas an error is as a mistake that can be isolated and corrected, *Wahn* is a way of going wrong in one's overall orientation, such that the whole procedure of rational reflection itself is distorted. We can make further sense of self-conceit's participating in the structure of *Wahn* when we notice that whereas error arises from something like ignorance or innocence, self-conceit, in its legislative aspirations, operates in light of "some antecedent, if only implicit, consciousness of the moral law" (Engstrom 2010, 111). So one must be capable of proper reflective judgment and see its merits, in order to be capable of self-conceit; one must recognize the claims of reason in order to set up defensive ersatz forms. Because of this, self-conceit constitutes an illusion to which a deliberative purist is especially prone.

It is also significant that in other writings Kant cautions against a kind of arrogance that comes from enjoying a charmed and unchallenging life, where this involves taking moral credit for one's good conduct where, "strictly speaking," one is simply lucky enough to have escaped real

temptations to vice (6:460). Self-conceit here involves mistaking one's good fortune—what we might now call *privilege*—as evidence of one's excellent moral disposition. Here self-conceit involves taking credit for what is actually a lucky accident or a contingent social arrangement. Kant conceives of this as a form of dishonesty “by which we throw dust in our own eyes” (6:38). So again, given the attractiveness of the ideal of rational self-satisfaction, we humans are especially prone to distort not only our own psychologies, but the role that luck and social privilege has played in facilitating that very sense of satisfaction and certainty.

Given all this, we can see that when Korsgaard describes “the person who acts from self-love” as “not actively willing at all, but simply *allowing* herself to be *controlled* by the passive part of her nature” (1996, 168), whatever one thinks of this picture of moral psychology, Korsgaard is at least wrong about Kant.⁸ My suggestion is that, despite its Kantian heritage, DE does not adequately recognize what Kant saw clearly: that principles and ideals are apt to inspire fantasies, that human reason itself can *rave* (1987, 275). Within the structure of illusion, deliberative rational reflection functions as just another move within that framework and hence cannot be employed to critique or puncture that illusion. On Kant's subtle conception, self-conceit is not a way of being “controlled” by something “outside” reason, but is a defensive propensity of human reason itself. As Stephen Darwall writes, in self-conceit “thoughts are represented to oneself as the result of rational judgment, as supported by reasons, when in fact that representation's very purpose is to defend oneself against the honest critical thought” (1988, 416).⁹ So self-conceit presents *as* reason while refusing what Kant saw to be reason's condition *sine qua non*: critique, where self-critique

⁸ Pippin makes a similar critique of the way Korsgaard describes failures of rationality. He argues that when she talks about alien forces, this “all sounds like a disease or exogenous interference, not *my* failure or weakness” (2008, 83).

⁹ Darwall's essay is specifically about self-deception, not self-conceit. But he is writing about self-deception as a special problem for Kantian moral philosophy, so it is fair to assume that his claims could apply to Kant's conception of self-conceit as well.

involves precisely the *courage* (6:405) to disrupt the kind of deliberative self-satisfaction that DE regards as our ultimate aspiration.

There are thus special risks that come with adopting a deliberately purist conception of agency, where the risk involve going through the motions of deliberative self-reflection as a way of evading real self-knowledge. I also showed that while DE's failure to appreciate this risk is often traced to its Kantian commitments, Kant himself was wholly sensitive to our propensity to defensive reflection and rationalization. Kant helps us see that there is an internal relationship holding between ideals of reason and the propensity to illusion, hence the special need for critique.¹⁰

I want to close this chapter by looking at the conceptual framework that places deliberation in such a privileged position and that makes it difficult to appreciate the need for critique, an importantly different form of practical self-relation.

§5. Beyond Deliberation

Recall the rationale for DE's "internal" methodology: because an agent is not a mere thing in the world but essentially reflexive and self-conscious, this is something that any philosophical account of agency must capture. That is, any philosophical picture of agency must reflect what it is like to *be* an agent. I think we should accept this internalist constraint on how philosophically to proceed and yet object to DE's next move. For DE makes a particular, substantive choice about which agential "stance" to analyze as representative of agency in general: the moment of conscious deliberation culminating in a decision. While some exercises of agency are no doubt like this, the moment of deliberation and decision is just one, rather narrow, manifestation of our agency, and so just one, rather narrow, conception of what constitutes the "practical perspective" (see Smith 2004, 336). In

¹⁰ See my 2012 "Unity and Synthesis in the Ego Ideal: Reading Freud's Concept through Kant's Philosophy" for a further discussion of the relationship between ideals and illusion.

the last section we saw that a narrow deliberative conception of agency, and the ideal it generates, tends not to recognize the particular risks that come with pursuing such an ideal. In this section I want to show that this narrow conception cannot give an appropriate place to what Kant saw as a possible remedy to the ills of deliberative agency, namely critique.

To begin, recall that if we accept DE's reduction of the broadly practical with the narrowly deliberative, this results in a hard distinction between *me* and *not me*, and between *activity* and *passivity*, where the agent is properly identified only with what she can actively endorse and everything else is framed as a kind of alien or outlaw (not me) or brute force (not agency). While I will later argue that we should give up such a hard line, my point now to indicate that it is DE's particular, deliberative model of agency that primes us to see whatever is not deliberatively endorsed as *not me*, to see our own desires and less-than-rational attitudes as "inflicted on me" (Moran 2001, 117). But it is only once this picture captivates us that our own attitudes are experienced as aliens. Notice also that such a picture positions us to be, if not wholly paranoid about being overwhelmed by alien influences, then at least strikingly defensive, in just the ways that Velleman, Lear, and McGeer cautioned.

I want now to look at an example of someone who experiences herself as not wholly comprehensible, in order to see where and how the narrowly deliberative account of agency misfires. The example is from Rahel Jaeggi (2014, 99-130): she asks us to imagine someone who identifies as a feminist, and yet who has desires and behavioral dispositions that she cannot square with that self-conception. Jaeggi describes someone who demurs and giggles in the company of men, who desires that her date would, say, open the door and pay for her dinners, and yet cannot make sense of why she is like this: these attitudes and actions do not fit her practical identity, does not endorse them, and sees all the reasons counting against them.

As DE would regard this scenario, this person's actions and attitudes find no deliberative support: her considerations of the reasons have settled the question of what she should do or desire, and yet the habits and desires persist. According to a "deliberatively purist view" like Moran's, that this person's attitudes do not spontaneously conform to the dictates of reason is enough to qualify her as in an essentially compromised state (McGeer 2007, 90). And according to such a view, any non-deliberative mode of self-relation will be, just thereby, not properly agential at all but rather a kind of observation-plus-control, as though she were either observing herself as a spectator or wrangling brute forces. So, according to DE, not only are we passive—"inflicted"—with regards to any attitudes that are not the deliverances of or spontaneously responsive to rational deliberation, but our relation to these aspects of ourselves is *just thereby* theoretical or observational, i.e., not practical or properly first-personal: if I am not "making up my mind," I am thereby a kind of a spectator.

But now notice that on a "deliberatively purist view," the stance of self-*critique* would have to be analyzed as *theoretical* rather than practical. In her discussion of the imperfect feminist, Jaeggi observes that the question the feminist needs to ask herself is not deliberative but critical: it is not "what do I want?" or "what should I do?" but "*what am I actually doing in what I already do?*" (2014, 128). So what are the relevant differences between these types of questions, and how would DE understand them?

The deliberative question takes the available field of values and reasons for granted in order to weigh those reasons and come to a rational decision; from this position, the inability to commit constitutes a kind of agential vice or infirmity, the only cure for which would be more deliberation plus greater resolve. Which seems unhelpful here, since this person finds herself in a situation in which deliberation and resolve are no longer what she needs. By contrast, the critical question—

what am I actually doing in what I already do?—begins by assuming that the available deliberative field cannot now be taken for granted, as the starting point for rational deliberation. The critical question is primarily concerned with whether the familiar and intuitive ways of making sense have somehow blinded one to possible alternatives (whether we have “thrown dust in our own eyes”). From the critical position, there is no passive inability to commit but rather a suspension of the deliberative imperative to commit, an abstention from decisive self-assertiveness engaged for the sake of critique. The point is not (yet) to make up one’s mind but to understand it better.¹¹

Notice that if we take the practical to be equivalent to and coextensive with the deliberative—if practical questions just *are* deliberative and directed towards decision-making and action—then critical questions (what am I doing? how can I better understand my desires?) will have to be analyzed as theoretical, third-personal, and observational, and where occupying such a stance thereby constitutes an evasion of responsibility. If one cannot avow one’s attitude and endorsing how it discloses the world, one is thereby relating to it theoretically, as a psychological fact with which one is simply stuck (“nothing I can do about it”).

However, the fact that the critical stance is not directed towards making a deliberative decision does not reason to characterize the stance as either theoretical *or* as an evasion of responsibility. It seems to be neither, and moreover, it seems that it is *only* fidelity to deliberative purism that inclines us to see critical questions as theoretical rather than properly practical, or as a way of shrugging off rather than taking up responsibility. There is no pre-philosophical reason to

¹¹ In his critique of Korsgaard, Kekes describes the case of a deeply ambivalent person, and tries to show that Korsgaard’s prescription of autonomy (complete with universalizability of maxims) is not the right response to such a case. But Kekes says: “Korsgaard is right in this: [he] should deliberate further” (2011, 348). Kekes describes deliberating as “a matter of testing the realism of his beliefs, emotions, and motives. It is to entertain seriously the possibility that they reflect illusions that sustain a false picture of himself.” And he says that deliberation can proceed in the direction of “the past, [which] involves considering what has led to our present ambivalence” (2011, 34). I agree with Kekes’ description of he thinks is needed in this kind of case, but I disagree that this is just “further deliberation.” Insofar as it is not directed at resolve and decision, I think it is helpful to distinguish deliberation from this process here. I would call it self-critique.

think that our practical engagement with the world must be essentially deliberative, or that being decisive is always responsible, or that abstaining from commitment or action is always evasive. Not only that, there are also social and political reasons to temper the inclination to decisiveness and engage in self-critique. If you take ideology or privilege seriously, for instance, a policy of deliberative purism and the pursuit of self-satisfaction begins to seem naïve.¹²

Critical questions begin with the assumption that I am in some way opaque, that my attitudes are not transparent to their objects, that my spontaneous responses to the world are in some way distorted. Relating to oneself as opaque for the sake of critique is not *just as such* spectatorial or observational or irresponsible; rather this is a basic practical posture for less-than-ideally-rational creatures like us (in a less than ideal world). Given our attractions to the promise of self-satisfaction and our propensities to defensive reflection and self-conceit—to name just two forms of self-obfuscation—arresting the rush to decisiveness and engaging in non-deliberative self-critical reflection seems not only not theoretical, but a quite essential manifestation of our agency. This would be a form of agency with a positive role for restlessness to play (I will return to this idea of critical, productive restlessness in Chapter Three).

*

In the next chapter I engage with Nomy Arpaly's rejection of DE and her positive alternative for understanding opaque agency. While she is not an outright skeptic about self-knowledge, she maintains that it is so exceptional—and that murkiness and opacity are so ubiquitous—that we would do better to abandon our philosophical commitment to the first-person perspective and first-person authority. Yet while much of Arpaly's views and philosophical approach seems absolutely

¹² Against moral philosophy's pressure to self-assertiveness and its tacit assumption "that I can verify my own judgment" (2001, 156), Adorno insists that "we need to hold fast to moral norms, to self-criticism, to the question of right and wrong, and at the same time to a sense of the fallibility of the authority that has the confidence to undertake such self-criticism" (ibid., 169).

right—principally, her commitment not to insert false philosophical clarity into the complexities and opacity of human experience—I will show that she goes too far in rejecting the relevance of the first-person perspective, and that we have reason to retain a sense of its importance, even when opaque.

Chapter Two: “Murkiness is the Rule”

As we saw, the central feature of DE is the special importance it grants to the “first-person perspective,” the agent’s reflective, deliberative endorsement of her own attitudes and actions. Because this approach can seem so idealized and unrealistic, moral philosophers turned their attention to complicated, self-opaque agents, and have challenged the basic commitments and presuppositions of DE.¹ Articulating an alternative to DE raises an important question about the role of the first-person perspective and the agent’s own self-understanding. If we want to understand self-opaque agency, how important is the agent’s own practical self-understanding?

In *Unprincipled Virtue* (2003), Nomy Arpaly argues that there is no need to emphasize the agent’s self-conception in an account of responsible, rational, moral agency. Arpaly observes that while most contemporary moral philosophers would agree that the human mind is not transparent to itself and prone to self-opacity, these philosophers tend to construct their theories of agency and personhood as though this observation could be made only to be set aside. As she notes:

Cases in which the mind of the agent seems quite transparent to her, cases in which she either acts for reasons that she understands and endorses or is carried to action by some atavistic force, cases in which what she regards as important to her is important to her, and cases in which, even when she is irrational, she knows it—these are all cases familiar from the [philosophical] literature. And yet the unconscious² does not usually get its due. 2003, 29.

¹ See for example: John Doris *Talking to Ourselves: Reflection, Ignorance, and Agency* (2015), George Sher *Who Knew? Responsibility Without Awareness* (2009); Angela Smith, especially “Conflicting Attitudes, Moral Agency, and Conceptions of the Self” (2004).

² Arpaly is not committed to a specifically Freudian conception of the unconscious. All she wants and needs is what she calls “garden-variety” forms of unconsciousness and unconscious motivation. By this she simply means forms of motivation that are not (presently) available to self-conscious reflection and human beings’ not infrequent inability to distinguish reason from unreason from within the deliberative perspective (or, our ability to meaningfully ask ourselves or others: “am I being crazy??”).

Arpaly argues that by focusing on cases that take an agent's self-knowledge for granted, these philosophers proceed as if insights about the mind's opacity could be accommodated without challenging the normative privilege accorded to the first-person perspective. So while challenges to the first-person perspective have not been simply ignored, they also have not been allowed to really transform our conceptions of agency and moral psychology. Against this trend, Arpaly focuses on cases which demonstrate "the sort of murkiness and complexity which is the *rule*, not the exception, in life" (2003, 30).

Because of the ubiquity of self-opacity, Arpaly argues, *contra* DE, that "to the extent that moral psychology emphasizes the first-person perspective, it risks misunderstanding [opaque] cases" (2003, 18). A person can be wrong about her reasons, or wrong about what counts as a good reason, or she can be bad at articulating her reasons or at thinking abstractly, or not particularly reflective, or a bit defensive, or in any other way opaque, and yet, according to Arpaly, such a person can still act rationally and responsibly. Arpaly concludes that if this is so, then the first-person perspective is not normatively significant and what matters for agency is not whatever the agent thinks or any form of self-conscious self-relation but instead her *real* reason for acting, which could only be discerned with a "God's eye view of [her] mental states" (2007, 431).

Arpaly thinks that philosophers' emphasis "the first-person perspective" causes them to misunderstand the kind of complex, opaque cases which she thinks makes up the bulk of ordinary life. How exactly does she understand "the first-person perspective"? While she does not provide a definition, it is clear that what she is really concerned to reject is the idea that reflective endorsement, conscious commitment, deliberation, or lucid self-awareness are criterial for agency. I agree with this, and yet I will try to show that de-emphasizing these demanding intellectualist criteria may not require that we jettison the "first-person perspective."

Part of the problem is that Arpaly conceives of the murky first-person perspective as a kind of obstacle that masks an agent's real reasons for acting, those reasons which we could see from a God's eye view. However, I think that if we look at many of her characters, a different picture emerges: it is not that her characters' reasons are settled and hidden from (their and our) view. Rather her characters' reasons and motivations are themselves indeterminate,³ unsettled, slightly chaotic *and*—and this is crucial—this indeterminacy is *registered from the agent's own perspective*.

Bringing out this kind of motivational or moral psychological indeterminacy provides a way of keeping the agent's first-person perspective in appropriate focus, without granting it absolute authority *or* conceiving of that perspective in terms of self-conscious deliberative endorsement. This means that we are able to mark a difference between “emphasizing” the first-person perspective in our understanding of human agency, and being naively credulous about first-person reports, or granting self-avowals canonical status, or giving too much weight to conscious deliberation. Arpaly argues that her account can do better justice to our ordinary intuitions about self-opaque agents; that is, her account is preferable because realistic and non-revisionary. One of the merits of my approach is that we can acknowledge self-opacity as “the rule” without having to de-emphasize the agent's perspective, which—given how much that perspective matters both for agents themselves and for those with whom they interact—seems too revisionary to be satisfying.

In the first section I present the main features of Arpaly's account of self-opaque agency. In the second section I examine three of her main case studies and offer different interpretations of them, arguing that my interpretations do better justice to the characters and to our intuitions about them. In section three I clarify the philosophical upshot of these re-interpretations. In the final

³ Reading and discussing Avner Baz's paper “Motivational Indeterminacy” (forthcoming 2016) was extremely helpful during the writing of this paper. I agree with much of what he says about indeterminacy, but since his view is indebted to Merleau-Ponty's and Wittgenstein's respective conceptions of our perceptual relationship with the world, engaging with Baz would take me too far from my aim here, which is to critically engage with Arpaly.

section I recall Moran’s conception of the relationship between opacity and indeterminacy—or what he calls “indecisiveness”—in order to differentiate his account from my own.

§1. Arpaly’s Motivations and Methodology

Arpaly’s positive aim in *Unprincipled Virtue* is to provide an account of rational, responsible, moral agency in light of the murkiness and complexity which is the *rule* in human life (2003, 30). For Arpaly, accounts of agency that emphasize the agent’s self-conscious deliberative perspective cannot do justice to such cases. As she puts it: “to the extent that moral psychology emphasizes the first-person perspective, it risks misunderstanding [opaque, complex] cases” (2003, 18).⁴ While Arpaly’s primary aim is not to completely disprove deliberative endorsement accounts, she wants to expose the limitations inherent in any account that gives special normative standing to an agent’s first-personal perspective where she understands that perspective primarily in terms of conscious deliberation and reflective endorsement. For Arpaly, these accounts are simply unrealistic: people are often wrong about their reasons for acting, people often act rationally and responsibly without engaging in any process of reflective deliberation, *and*—as we saw in Chapter One—reflective deliberation can function as a defensive mechanism used to avoid rationality and self-knowledge.

But it is important to note that Arpaly has accepted DE’s particular conception of the first person practical perspective as narrowly deliberative, such that in rejecting the importance latter she thinks she needs to reject the former. I showed in Chapter One that we have reason to reject DE’s narrow conception of the first person practical perspective. This means it might be possible to give

⁴ Arpaly explicitly frames her account as a critique of the role of the first-person perspective. But she also critiques the idea that an agent’s “best judgment” is normatively special, and she claims that an agent’s sense of her own rationality is just “one belief amongst others.” So Arpaly moves fairly fluidly between these three ideas. I will occasionally use these terms but more often will refer to an agent’s self-conception, since I think this concept captures the essence of Arpaly’s target. In addition, I think the first-person/third-person, practical/theoretical, participant/observer are still in need of clarification and do not always map neatly on to one another.

up their emphasis of reflective, self-conscious deliberation without de-emphasizing the first person perspective generally.

Arpaly's strategy is to argue by example, presenting compelling cases of self-opaque agents who nonetheless seem to be rational and responsible for what they do. While she discusses dozens of examples, her principal case study is the famous scene in the *Adventures of Huckleberry Finn*, where Huck has the opportunity to turn Jim in but does not. Huck consciously holds racist beliefs and, according to Arpaly, his "best judgment" is to turn Jim in to the authorities. Yet Huck does not do this when the opportunity presents and he conceives of himself as weak-willed and bad, unable to do what he believes he should. Arpaly thinks that we intuitively see Huck as having done the right thing for the right reasons, and hence as praiseworthy; the only thing that gets in the way of our recognizing that Huck acted rightly and rationally is that Huck himself does not see things this way.

Arpaly writes:

while Huckleberry does not conceptualize his realization, it is this awareness of Jim's humanity that causes him to become emotionally incapable of turning Jim in...[he is] acting for a moral motive—*without knowing* that it is a moral motive. 2003, 10.

Arpaly wants us to see that Huck is a familiar figure from ordinary moral life, as is our admiring moral response to him. She thinks that we recognize, pre-philosophically, that agents like Huck often act without knowing either *what* their reasons are or that their reasons are *good* reasons. On her theory, an agent can do the right thing *for* the right moral reasons, and this does not depend on what the agent consciously *believes* to be either the right reasons or her reasons (2003, 73). Whatever people think they are up to, whatever people think count as good or bad reasons, is beside the point. Arpaly wants us to see that, like so many of us, Huck did the right thing for the right reasons—he just didn't know it.

In light of both her commitment to messy cases and her rejection of any commitment to the idea of first-person authority, Arpaly pursues a wholly different philosophical strategy than DE. There is, as Arpaly often puts it, “*a fact of the matter*” (see especially 2006, 16; 2003 *passim*) as to a person’s reasons or the state of her will, and yet the agent herself need not be aware of these facts. An agent’s understanding of her reasons, her sense of what counts as a good reason, her conscious deliberations: none of this bears on the facts of her moral psychology, and the latter are what is relevant for responsible, moral agency. Notice how different this is from DE: whereas, for example, Moran rejects any picture on which an agent’s attitudes are simply present and indifferent to the agent’s own deliberations, this is exactly the picture that Arpaly embraces: the agent’s reasons are there as a matter of fact, whatever the agent herself thinks or knows about this. Because she does not see the first-person perspective as having special authority, Arpaly also sees no reason for it to provide any special methodological constraint on our philosophical theorizing. Thus she provides what she calls a “*third-person theory*” of rational agency (2003, 33). If the agent’s perspective is typically murky and opaque, why should it enjoy any practical or philosophical privilege?

Now, one might think that if we accept that self-opacity is the rule in human life and that clear-eyed deliberation terminating in a decision is rare, then we should abandon the ideas that agents are responsible for their actions and/or that persons are morally evaluable. One example of this can be found in Brian Leiter’s critique of morality and agent-responsibility, articulated through his naturalistic reading of Nietzsche (2015). He maintains that once we give up the ideas of self-transparent agency and the epistemic accessibility of motives, it no longer makes sense to hold people responsible for what they do or subject them to moral evaluation. For Leiter, the opacity of

motives means that agent responsibility and moral appraisal are unintelligible ideas and incoherent practices that must be rejected or revised.⁵

For Arpaly, however, the opacity of motives does not require the abandonment of responsibility or our ordinary practices of moral evaluation. What really distinguishes her project is that she wants to reject the idea that reflective deliberation or self-knowledge are important for responsible moral agency and yet retain our pre-theoretical ideas about responsibility, rationality, and moral worth. In fact Arpaly thinks that *de-emphasizing* an agent's avowed self-understanding is the only way of doing justice to our moral intuitions, the only way of making "all the fine-grained moral distinctions that need to be made" (2003, 169).

Returning to Huck, the idea is that if the target of our assessment were an agent's conscious, endorsed, deliberated attitudes, then we would have to focus on Huck's avowed racist attitudes and his expressions of regret at not turning Jim in; this means that we would have to judge him both blameworthy (for holding immoral attitudes) and irrational (for failing to act on his best judgment). But Arpaly thinks that this is not how we regard Huck. Instead, she thinks we are inclined to see Huck as acting for a good reason and hence as praiseworthy (2003, 10). This means that an agent's *actual* reasons should be the object of our assessment, not however he conceives of himself or his express moral commitments or whatever he consciously avows. As Candace Vogler puts it, Arpaly

⁵ See Leiter's 2015 *Nietzsche on Morality*. For example, on Nietzsche's critique of the idea of self-transparency, Leiter writes: "we do not have epistemic access to what the causally effective motives really are. Consequently we cannot access actions in terms of their motives, since we lack knowledge of the relevant differentiating features. Thus Nietzsche's critique of the descriptive presuppositions of [Morality in the Pejorative Sense] is simply that if motives are to be the basis for the moral evaluation of actions, then such evaluation is impossible, since the motives for action are largely unknown" (2015, 83). Leiter, according to his reading of Nietzsche, also holds the incompatibilist view that since human beings do not have free wills (or even more strongly, do not have wills *qua* internal locus of agency at all), they cannot be responsible or held responsible for their actions. In *Merit, Meaning, and Human Bondage* (2006), Arpaly defends the compatibilist view that a deterministic world does not preclude responsibility and reason responsiveness.

wants to show that “the virtuous agent’s sense of what she is up to and why may be inarticulate and slightly chaotic. What she is up to, and why she merits our praise, need not be” (2006, 373).

This is the idea that I will challenge, that Arpaly’s characters are best or rightly described as having inarticulate and chaotic self-conceptions, yet with determinate reasons and motives on which they act, which would be clear from a God’s eye view. I will propose that the agent’s chaotic self-conception is often importantly related to “chaos” at the level of her reasons and motives themselves. In this way, the agent’s self-opacity, her inability to understand what she does, appropriately *reflects* the murkiness and unclarity at the level of her motives themselves. As we will see, this allows us to emphasize the first-person perspective even in the case of opaque agents, and thus retain a minimally unified agential perspective, even if quite conflicted and murky.

§2. Opacity and Indeterminacy

As we have seen, Arpaly’s account of agency involves the idea that one can act *for* a reason *because* it is a good reason and yet not know either what one’s reason is or that it is good. In response, critics have argued that Arpaly never explains what is involved in acting *for* a reason, if this can be fully independent of what the agent herself thinks she is doing. Robert Pippin, to take one example, counters that the “for” in “acting *for* reasons” “must mean *something* motivationally ‘present’ to consciousness or we will have a simple case of accidental rationality” (2007, 297).⁶

I will challenge Arpaly’s picture from another angle: while Pippin contends against Arpaly

⁶ Arpaly’s account of acting for reasons without agent awareness or agent endorsement has been criticized by a number of philosophers for failing to provide a positive alternative to what Mason calls the “justification internalism constraint on rationality,” i.e., the idea that an agent must be able justify her belief or action from her own point of view, in order to count as rational. Put negatively, when my belief is true or my action is good, and yet I see no reason to justify my belief or action, then while I “get it right” without any justification from my own point of view, then I do not count as rational. Without an alternative to the justification internalism constraint, Arpaly cannot distinguish acting *for* reason, and merely acting in accordance with reasons, and she cannot distinguish between a rational response to a reason, and a merely causal response to a feature of the situation. See Mason 2007, Pippin 2007, Stroud 2007.

that *some thing* must be present to consciousness, I will argue that in Arpaly's key examples there is *no determinate thing* which could be conscious or not.⁷ Or rather: I will argue that the lack of determinacy at the level of reasons and motives is reflected at the level of the agent's consciousness, in the form of a kind of registered opacity. The idea is that in these cases opacity is not superficial but pervasive, that it characterizes not only the agent's perspective but her reasons and motives themselves. Because Arpaly posits clarity and determinacy at the level of the agents' real reasons and motives, the agent's own murky perspective can start to seem wholly irrelevant to understanding what she does and why. Against this view, I will try to show that the idea of pervasive motivational indeterminacy that manifests as self-opacity provides a way of maintaining the relevance of the agent's perspective for our understanding and appraisal, even when that perspective is opaque.

As we will see, it is significant that the characters in Arpaly's key case studies are all in the midst of evaluative, moral, or rational change: they are coming to see or appreciate reasons that they hadn't before, over some extended period of time, and through exposure to new experiences.⁸ These characters are facing decisions that will transform not only their circumstances but who they are in fundamental ways.⁹ They are beginning to appreciate different kinds of reasons. It strikes me that such change is crucially relevant for understanding (and appraising) the agent's action and the agent's self-opacity. If we take this dimension of change seriously, we can see the opacity and

⁷ In his own work, Pippin is very interested in the idea that there can be as yet no determinate reason for which an agent acts.

⁸ I will not be able to discuss them all, but here is a list of some of the characters who I think qualify as "changing" at the moment of their opacity: Huck, Oliver, Lynn, Tamara, Candide, Sam, Emily, and Oskar Schindler.

⁹ See L.A. Paul's *Transformative Experience* (2014) for an extended discussion of the idea that certain major life choices must be made "in the dark," without the kinds of reasons available in non-transformative choices, and often result in a transformation of the values of the chooser. Two shortcomings I see in Paul's are: first, she focuses exclusively on choices that the agent knows will change who she is (whereas many transformative experiences are not recognized as such in advance), and two, she focuses exclusively on the idea of single momentous conscious choices (whereas many transformative experiences are more gradual, with discrete choices playing a smaller, less dramatic role).

murkiness of the agent's first-person perspective, not as a failure to track some determinate but unknown reason, but rather as reflecting murkiness at the level of the agent's reasons themselves. That is, in these cases, rational, motivational determinacy is not simply epistemically unavailable: it has not been achieved.

Arpaly's book opens with a character at a moment of crisis and in the midst of major change. Oliver Single of John Le Carré's *Single and Single* is a lawyer at his father's firm which is involved in organized crime. Arpaly notes that "Oliver has always known that his father makes shady deals, but only gradually does he learn quite how shady" (2003, 3). In the relevant scene, Oliver is about to turn on his father and call the authorities. As Le Carré writes:

It is midwinter, and Oliver is a little mad. That much he knows about himself, no more. The origins of his madness, its causes, duration and degree, are not within his grasp, not now. They are out there, but for another time, another life, another couple of brandies... Something amazing is about to happen to him and he is eager to find out what it will be... A few vodkas on the flight...have done no more in Oliver's eyes than bringing him up to speed with the furor already raging inside him. Le Carré 1999, 90.

At the airport, Oliver approaches a phone to speak to a customs officer: "he uses it. That is to say, his arm reaches out for it unbidden, his hand grasps it and puts it to his ear, leaving him with the responsibility of what to say" (Le Carré 1999, 91).

Arpaly wants us to see that Oliver is an unfamiliar character for much moral philosophy: he is not the self-aware, reflective, deliberating agent of DE, and yet he is also not swept along by mere urges or alien forces. The question is whether we should see Oliver as acting for particular reasons that were determinate and present but simply opaque to him, and whether the opacity of his first-person perspective is irrelevant to our understanding and assessment.

Aspects of Le Carré's description point towards Arpaly's interpretation. If "the origins of [Oliver's] madness, its causes, duration and degree, are not within his grasp, not now," this could suggest that the relevant beliefs, desires, and reasons are really *there*, but just not within Oliver's

conscious grasp. Indeed, Le Carré explicitly says that “*they are out there*, but for another time, another life,” indicating that we should imagine Oliver’s reasons as already determinate yet epistemically unavailable.

But there is alternative possibility. The idea that the origins of his madness—that which will constitute his reasons for defecting—are “not within his grasp” could mean that they are not fully formed. These reasons evade Oliver’s grasp not because they are determinate but hidden, but because they are not yet solidified, so there is nothing there to grasp. Consider another moment from Le Carré’s passage: the idea that drinking has brought Oliver up to speed with the “furor already raging inside him.” This might suggest that drinking has helped him see the reasons that are “already” present “inside him.” But if what is inside him is a *furor*, then this means that his internal life is in upheaval and disarray, that his beliefs, desires, commitments, and values are tumultuous and changing. Moreover, that he is “waiting to see” what will happen, that self-understanding is “for another time,” that he is “getting up to speed”—all this suggests that we need to understand Oliver’s motivational state *now* as part of a diachronically extended transformation.

Notice also that Oliver quite consciously experiences himself as “mad,” as painfully aware of the furor inside, as “a stranger in his own country” (1999, 91). Oliver’s first-personally available sense of himself is as profoundly, deeply disoriented, a stranger to a world that was once familiar, orienting, and thereby reason-giving. Arpaly recognizes this, of course; she notes that “he experiences himself as the passive audience to his own actions” (2003, 5). But is the right way to make sense of Oliver’s decision to insist that he acted on a determinate motive and it was just his access to it that was murky?

Recall that one of Arpaly’s complaints against DE theorists of agency was that they insist on more clarity than ordinary life exhibits. Murkiness is the rule and yet these philosophers proceed as

if we human beings were always perfectly lucid and self-aware. But the risk now is that despite her emphasis on doing justice to murkiness and complexity, Arpaly's interpretation actually finds *more* clarity than the case suggests. In de-emphasizing the first-person perspective, Arpaly de-emphasizes Oliver's murky experience of himself but then she reinserts clarity at the level of his reasons and motivations. Yet Oliver's inner furor seems to run deeper than that.

Arpaly claims that her approach to self-opaque agency does better justice not only to characters like Oliver, but to our understanding and evaluation of them. She thinks our intuition is to admire Oliver for his momentous moral decision, and she notes that he "will later act morally and bravely on many occasions" (2003, 6). I want to call attention to two things here: first, it is noteworthy that Arpaly mentions what Oliver will go on to do later, suggesting that our assessment of Oliver's choice to defect at this point is importantly shaped by who he actually *becomes*: that he continues to make such moral decisions is part of what gives us confidence that his decision in *this* case was based on the right reasons. Without that further development, it is unclear what our assessment would be or how we would make it. As John Doris notes, "it will often be difficult to determine whether someone holds a value, in the absence of a temporally extended pattern of cognition, rationalization, and behavior" (2015, 162). But such temporally extended patterns are important not only for identifying some value or reason (which Arpaly recognizes [2003, 98]), but for developing and solidifying values and reasons themselves: "valuing has a temporal dimension...values are developed, sustained, and revised during extended periods" (ibid.). This suggests that, at Oliver's moment of choice, it is not only that we have insufficient evidence to attribute good, determinate reasons, but that we have grounds for thinking that these reasons are in the process of being developed: at *this* moment of inner furor, they are indeterminate. And I think

their indeterminacy is reflected in “our” intuitive response to him: that is, I think that if we admire him, our admiration of him is not unqualified, but is rather indeterminate.

Another character Arpaly describes is Emily, a graduate student who has always thought she should pursue a Ph.D. in chemistry but who gradually begins to feel restless, sad, and ill-motivated as she works through the program. “These feelings,” Arpaly writes, “are triggered by a variety of factors which, let us suppose, are good reasons for her, given her beliefs and desires, not to be in the program” (2003, 49). Eventually, “on an impulse,” Emily quits the program, castigating herself as lazy and yet feeling relieved, though she does not understand why. Emily is self-opaque: she cannot make sense of herself before, during, or after her decision. But again the question is, first, is the best way to understand this scenario is to think that Emily had clear reasons on which she acted, but that she simply failed or refused to see them? And second, is the best way to understand and assess Emily by de-emphasizing the deliverances of her first-person perspective?

Notice that with the determinate yet hidden reasons picture, Arpaly’s underlying assumption is that Emily’s beliefs and desires—the beliefs and desires that gave her reasons for leaving the program—remain more or less the same over time, from when she enters the program to the moment of quitting. But there is another way of conceiving of this case, according to which Emily is developing a new understanding of herself during the time that she spends there. It seems possible, indeed likely, that Emily herself would change over the course of her studies, acquiring new beliefs and desires and hence new reasons, and responding to the world in new ways. If this were so, then Emily’s self-opacity is not a matter of simply lacking epistemic access to established reasons; rather her self-opacity and felt restlessness expresses and reflects her as-yet-unsettled volitional life.

Like Oliver, Emily is not radically self-ignorant. Instead, like Oliver, she is quite consciously out of joint with respect to herself and senses that something is off. Like Oliver, Emily is changing,

such that a temporally extended perspective will be required in order to understand her reasons for quitting. So do we understand Emily best by *de-emphasizing* her first-person perspective in order to focus on her reasons? I would argue, to the contrary, that attending to the deliverances of her first-person perspective—i.e., her experiences of sadness, lack of motivation, and restlessness—is indispensable for understanding Emily and what she does, and for any assessment we might make. Whereas Arpaly maintains that “to the extent that moral psychology emphasizes the first-person perspective, it risks misunderstanding these cases” (2003, 18), I am suggesting that in cases like Emily and Oliver, understanding such cases *depends* upon giving the right (though not naively credulous) emphasis to their first-person perspective. While their self-understandings are surely opaque, this does not entail that just for that reason they should be de-emphasized.

Finally, consider Huck again. Arpaly notes that “during the time he spends with Jim, Huckleberry undergoes a perceptual shift” (2003, 76) and comes to see Jim as a person, just like him. Arpaly emphasizes that while Huck does not conceptualize or consciously reflect on this shift, he is “aware” of and “viscerally experiences” Jim’s humanity, and that “Jim’s humanity” is Huck’s reason for acting as he does. Huck acts for good reasons, even though “he does not *know or believe* that these are the right reasons.” Arpaly even goes a step further when she postulates that Huck “does not have the belief that what he does is right *anywhere* in his head” (2003, 77). Again the question is whether this is the best way to make sense of Huck.

The first thing to note is that it is hard to understand what it means to act *for* some reason *because* it is a *good* reason, yet without having this belief “anywhere in one’s head.” The idea spatializes and locates reasons somewhere “in the head” in a strangely literal way,¹⁰ but worse, it

¹⁰ I am thinking here of Wittgenstein’s urging that we reject the picture of thoughts as inner mental events. In Part II of the *Philosophical Investigations*, Wittgenstein writes: “If God had looked into our minds he would not have been able to see there [what] we were speaking of” (Pt II, p. 217). This is obviously an obscure philosophical thought (or “reminder”), but my point in referencing it is just to call attention to his critique of

suggests that Huck acts because of a reason he does not actually have (“anywhere”). But even more curiously, Arpaly also describes Huck as “dimly *aware*” that he is acting for such reasons, indicating not only that the belief is “in his head,” but that he is himself *aware* of this, however dimly. So not only do Arpaly’s cases seem to involve more unclarity at the level of the will, there are moments when they seem to involve forms of self-awareness—uncertain and struggling no doubt, but awareness nonetheless.

One reason Huck’s awareness might be only dim or dawning is because, at the moment of his action, Huck is in the midst of major evaluative change. So rather than insist that he had a good reason for which he acted wholly unknowingly, it might be more accurate to say that Huck has something close to what Agnes Callard calls a “proleptic reason,” a reason which is “provisional in a way that reflects the provisionality of the agent’s own knowledge and development: [his] inchoate, anticipatory and indirect grasp of some good [he] is trying to know better” (2016, 132). For Callard, we need the category of proleptic reasons to account for the special kind of rationality involved when we act on reasons that we do not fully grasp or appreciate, paradigmatically, the reasons to which we are responsive when we are in the midst of changing. Crucially, proleptic reasons and motives are not definite reasons that are just inchoately grasped; rather they are themselves inchoate.¹¹ Put in the language we have been using here, an agent’s proleptic reasons are murky, not just at the level of access, but, as it were, all the way down.

the idea that there is *something there*, which, even if we cannot know it, *God* could know it, as though it were just a problem of access. Wittgenstein wants us to move away from the picture of the mental—thinking, believing, valuing, etc.—as taking place “inside the head.” So when Arpaly says that some belief is nowhere in Huck’s head, Wittgenstein would suggest that she is looking in the wrong place to find evidence of such a belief.

¹¹ To be clear, the topic here is moral psychology, not metaethics. So with the idea of provisional or inchoate reasons, our concern is with an agent’s relationship with her reasons and motives, not the metaphysical status of value or reasons themselves.

Callard's category of proleptic reasons helps us understand the idea of motivational indeterminacy. Proleptic reasons are provisional and inchoate, where murkiness is not just epistemic, but runs to the level of reasons and motives themselves. However, I say that Huck's reason is only "close to" a proleptic reason because Callard is specifically interested in what she calls "aspiring agents," agents who are *trying* to get to a place where they can fully appreciate the reasons that are as yet provisional and inchoate. But this picture is too intentional and aspirational to apply to Huck, who is not *trying* to become other than he is; indeed, in many ways he is trying to resist such change.

That Huck is not best understood as aspirational is also importantly connected to the fact that he is changing not only relative to earlier views and not only relative to the social world in which he lives, but in the sense that he is *growing up*¹²: Huck is barely thirteen years old. Recalling Wittgenstein's critique of the idea that learning one's first language could be understood on the model of learning a second language as an adult, we need to bear in mind here that Huck is not changing his views from one set to another (as an adult might), but developing into a position where he can have moral views or moral reasons at all. *Contra* Arpaly, I do not think that the relevant point here is that Huck "does not conceptualize" his reasons (2003, 10) or is "not a very clear abstract thinker" (2003, 77). What is important is that he is a child struggling on the cusp of adolescence and thus still developing his rational and moral capacities in the first place.¹³ Because of this, it is only in a loose way that we can compare the way Huck changes to Emily or Oliver. The latter characters

¹² And, for what it is worth, in seriously deprived and traumatic conditions. I mention this not because I think this prohibits a person from acting rationally, but because, when we are trying to understand a vulnerable child's reasons for acting in such conditions, the story will have to be more complicated than just: "he is praiseworthy because he did the right thing for the right reasons."

¹³ I am not suggesting that at a certain age we stop developing these capacities. In fact, I believe that the idea of adult moral change involves the idea that our moral capacities continue to change and develop in ways specific to mature moral psychology. My point now is just that Huck's capacities are at the earliest, most fragile stages of development.

are growing in the sense of shifting evaluative stances, whereas Huck is growing in the sense of developing the capacity to take an evaluative stance in the first place. Each transitional position comes with its own unique challenges, yet that each is a *transitional* position suggests that the reasons acted on are not wholly settled.

So where has all this character study gotten us?

My approach to Arpaly's case studies has followed a single interpretive strategy, emphasizing that in these cases murkiness runs all the way down and that the agent's murky, opaque, or confused first-person perspective does not mask or fail to reflect good, clear reasons, but accurately expresses motivational, moral psychological indeterminacy. This suggests that we cannot dismiss an agent's first-person perspective and self-conception as simply getting the facts wrong. In fact, in these cases, being murky and self-opaque is a way of getting things right. So while Arpaly has rightly introduced greater murkiness and complexity into our philosophical moral psychology, at times she does not go far enough. And, somewhat paradoxically, the idea of murkiness running "all the way down" allows us to retain our emphasis on the first-person perspective.

Let me now say more about what it means to emphasize or de-emphasize that perspective.

§3. Emphasizing an Opaque Perspective

On Arpaly's interpretations, that these characters are self-opaque means that their self-conscious assessments of themselves and their situations are wrong relative to their real reasons. These characters are, to use Arpaly's characterizations, not good at "abstract thinking" or are not very articulate or not very reflective or are too defensive or hold bad views about what path they should pursue or hold bad views about morality.¹⁴ In all these cases, the first-person perspective is wrong

¹⁴ Arpaly's idea that someone could hold "bad views about morality" (2007) and still be morally praiseworthy is confusing. I understand this idea if it means something like: "Huck didn't really know what he was talking

relative to the actual state of her will, and only the latter is relevant for responsibility and moral assessment. As Arpaly affirms this view in *Merit, Meaning, and Human Bondage*:

facts about the agent's motivations and reasons make praise or blame warranted or unwarranted. There is no need to switch on some special stance called 'the reactive attitudes' in order to glimpse these facts. A criminal's ill will is 'out there,' evident. 2006, 16.

When she describes the first-person perspective as opaque and failing to track what is "out there and evident," Arpaly suggests that an opaque first-person perspective simply fails to correspond to the facts and to that extent should not be emphasized. Whatever they reflectively endorse or believe or self-consciously commit to, the facts about their motivations and reasons tell otherwise. These characters have false views about their reasons and so we need not emphasize their self-understandings in our efforts to understand why they act.

It is important to bear in mind the *moral* motivation guiding Arpaly's philosophical project. She thinks that philosophy has been wrong—theoretically and morally—to put so much emphasis on self-conscious deliberation, reflection, articulate reasons, and self-knowledge. Many people act for good reasons and are rightly praiseworthy, and yet would fail to meet the demanding intellectualist criteria set by demanding intellectuals. Because philosophical accounts that lean intellectualist tend also to be accounts that privilege the first-person perspective, in rejecting the former Arpaly thinks we also need to reject the significance of the latter.

The problem is that Arpaly has accepted DE's particular conception of the first-person practical perspective as narrowly deliberative and unrealistically lucid. But if this is so, then there might be conceptual room to reject the narrowly deliberative, overly reflective, and sometimes

about but he did the right thing for the right reason." But if we imagine that Huck was an adult and refused to acknowledge that he did the right thing and continued to spout racist views, then I do not see how we can distinguish his "bad views about morality" from his actual moral orientation. That is, in the case of an adult, one's bad views about morality just seem to be one's morality. Here again I think Arpaly's main target is intellectualism about morality, which means that one's moral views can be poorly articulated or one might not use the right words.

naively credulous picture of the agent's self-conception while retaining our sense of the importance of the agent's perspective.

A reason to think that the first-person perspective is important for moral philosophy and moral psychology is that this is the perspective through which we understand and interact with one another, and this is true even if we recognize that people are routinely opaque to themselves, that there is more to their actions than what they themselves can or will see. That Huck feels as confused, that Oliver experiences himself as mad and in need of a drink, that Emily feels restless and sad but doesn't know why—all of this is registered first-personally and all of it crucial for our understanding and assessment. In conceiving of these characters as opaque, we are not conceiving of them as wholly ignorant or just out to lunch about themselves or what they are up to. Rather, we think that what they can say or are willing to avow casts important light on how we understand their actions, even if it does not tell the whole story.

Now Arpaly would grant that a person's avowals can be part of the story. She describes an agent's "best judgment"—the judgment reached through reflective deliberation—as not normatively special but just "one belief amongst others" (2003, 42). So presumably an agent's self-understanding or avowed commitments or any other deliverance of the first-person perspective would enjoy that same kind of status: it is just one belief amongst others, of "minimal significance" and enjoying no "special normative force" (2003, 44). The problem is that Arpaly is committed to aligning her account with our ordinary understanding and appraisals of each other, yet her deflationary approach to an agent's self-understanding is radically revisionary. She is right, of course, that many of our intuitions about her murky cases would not be well captured by deliberative endorsement theorists. But I do not think that our intuitions follow her all the way in regarding the first-person perspective as normatively insignificant, just one among any of the agent's beliefs.

Notice that, on her deflationary approach, it is not clear what it would mean to actually *hold* a person responsible or *ask* her to answer for what she did. That we engage with people in this way, asking them to give an account of themselves, does not commit us to naïve credulity about what they say, where their own self-reports simply carry the day; there is ample room to correct someone's faulty or murky self-understanding. The point is that we need to emphasize the first-person perspective in order to engage in any such correction at all.

Arpaly tends to focus, not on how we interact with each other, but on judgments of praise and blame, which on her view are not practical stances but “belief-like epistemic attitudes” (2003, 173): “to hold someone blameworthy is... to hold that a certain attitude toward him is epistemically rational: there was ill will, there was a wrong act, thus blame is warranted” (ibid.). If the way we think about one another is constrained to such detached evaluations, then a person's self-understanding will begin to seem unimportant, just one more belief. The risk, though, is that such abstract assessments look like “a pointless assignment of moral grades” (Scanlon 2008, 127), and this is *especially* so if the will can only be seen from a God's eye view. For all this moral assessment, there seems to be little traction with actual human life, where an agent's perspective and self-understanding starts to matter again.

Consider Arpaly's claim that a person's moral worth tracks the depth of her moral concern, where such concern is manifested in a person's emotional and felt reactions to moral phenomena, or in noticing “morally salient things,” or in being generally “morality conscious” (2003, 87). Now I think Arpaly is right that a person can notice and respond to features of her environment without being explicitly or reflectively conscious of this, or can feel guilt or pleasure in response to moral phenomena without understanding those feelings, or can be morally conscious without being especially articulate. And yet these all concern how a person experiences the world from her own

practical perspective, and—as Arpaly insists—they are central to our understanding and assessments of one another. Emphasizing the first-person perspective is essential to getting these phenomena in view, though emphasizing or privileging conscious reflective endorsement need not be.

So it seems we do treat the first-person perspective as normatively special in our ordinary social and moral interactions with and assessments of one another, without thereby treating a person's avowals as infallible or canonical. De-emphasizing the first-person perspective would be revisionary with respect to these ordinary moral practices.

We also need the idea that the first-person perspective is somehow special in order to make any sense of the intense conflict and acute restlessness evidenced and experienced by characters like Oliver, Emily, or Huck. That is, we could not make sense of why it should matter so much that they are about to make a decision that challenges their sense of who they are, a challenge felt so deeply that they cannot even bring it into view, cannot properly think it. If Oliver's self-conception was just one belief amongst others, why did he need to drink so much in advance of making a decision that would challenge it?

Thus it actually seems that our propensity to self-opacity and all the ways we avoid knowing ourselves is a *testament* to the normative significance and specialness of the first-person perspective and the judgments it issues: the ubiquity of murkiness and self-opacity prove the point that our self-conceptions are normatively special, not just one belief amongst others. Put otherwise: it is only because how we understand ourselves is profoundly significant that we experience self-opacity—both ours and others—as a practical, even existential problem, where this has nothing to do with

intellectualism or one's capacity for abstract thinking (recall: even Huck is bothered by his own action, because of how he understands himself).¹⁵

In *Unprincipled Virtue* Arpaly works to elicit shared intuitions about her examples in order to motivate her theory of opaque agency. I have tried to show that we do not yet have good enough reason to accept her theory, with its high cost of de-emphasizing the agent's own sense of what she is up to and taking recourse to a God's eye view. By offering different, hopefully more compelling, and less revisionary readings of her examples, I argued that in many cases, the opaque first-person perspective can be seen as expressive of and continuous with the agent's motivational indeterminacy. And by pointing out the important differences between taking the first-person perspective to be normatively significant, and privileging reflective endorsement, conscious deliberation, and facility at abstract thinking, I have shown not only that we can retain an emphasis on the former, but that doing so even explains the ubiquity of opacity: it is *because* it is so normatively significant, because challenges to it are felt to be so unmanageable, that it tends to be so opaque. I think we need to retain a qualified, realistic, critical sense of the importance of the first-person perspective even for self-opaque agents. Thus my view can be seen as a kind of middle way between DE and Arpaly.

§4. Indeterminacy or Indecisiveness?

In closing, I want to address the objection that in emphasizing the connection between indeterminacy and opacity, I have actually provided a view very close to Moran's. As we saw in Chapter One, Moran claims that self-opacity involves, not epistemic blindness to a fully formed mental state, but conflict or ambivalence at the level of the state itself. As he writes, "even when we

¹⁵ It is open to Arpaly to offer a normative recommendation that we work to grant agents' self-conceptions—our own and others—less significance, but then she would have to revise her commitment to being non-revisionary.

can imagine a situation of theoretical blindness to one's own intention, it will be hard to see this as something like mere ignorance, a gap in awareness which reflects no (practical) conflict in the intention itself' (2001, 57). And again: "uncertainty about what one intends to do is normally a matter of one's not having yet fully *formed* an intention" (2001, 55-56). In this chapter I argued against Arpaly's conception of self-opacity as kind of error or epistemic block, in favor of the idea that self-opacity is often expressive of deeper motivational indeterminacy. My idea, to use Moran's language, is that uncertainty about one's reasons is often a matter of one's reasons being not yet fully formed. But now it seems I have returned to the very view I claimed to be rejecting.

While I agree that my view seems similar to Moran's, there is an essential difference. There is an important normative connotation attached to Moran's account that is not attached to mine. Because Moran advocates a deliberative conception of agency and self-knowledge, he regards self-opacity as a failure or privation of deliberative agency. While he deals thoughtfully with various forms of self-opacity and while he acknowledges the pervasiveness of self-opacity in human life, he is committed to understanding self-opacity—*qua* defect of agency—as something to be fixed through greater deliberative resolve. On his picture, an agent's self-opacity reflects uncertainty, conflict, ambivalence, indecisiveness and/or irresolution, where what the agent ought to do is make a decision, make up her mind. Self-opacity reflects "practical and not theoretical uncertainty, and the resolution of it is a *decision*" (2001, 56-my emphasis). So there are two normative claims at work in Moran's picture: 1) self-opacity is a *failure* of deliberative agency, and 2) this failure *ought* to be resolved through a decision.

While I have argued that self-opacity can be tied to indeterminacy, I do not share Moran's generalized normative evaluation of self-opacity. Indeterminacy is not necessarily a form of failure

and there is need be no general normative pressure towards greater decisiveness.¹⁶ Moreover, indeterminacy is not necessarily a form of indecisiveness. Of course, sometimes indeterminacy might be bound up with a kind of indecisiveness which, for various reasons particular to the situation, it would be better for the agent to resolve. But whether and to the extent that this is so will vary from case to case.¹⁷ There may be some times or areas of life where it is not just psychologically difficult to be decisive and resolved but where the latter are not the right goals.

Recall that in the kinds of cases we have considered, indeterminacy is crucially tied to change: these characters were making choices at moments in their lives when they were becoming different people. In these cases, indeterminacy should not be conceived as indecisiveness, because the agent is not *failing* to decide between available determinate options. Rather, the person is changing such that what counts as an option for her (a reason, a motive, a value, a possible course of action, a mode of life, etc.) is changing as well. In Emily's case, for instance, it is not that she occupies a stable position and simply cannot decide between staying or leaving; it is that her sense of who she is, what she wants, and so what counts as a reason is changing, and so the options of staying or leaving are taking on new significances for her. Sometimes change is intentional or aspirational; here we want to become different and we do what we can to change. But sometimes we do not yet know who we are becoming, and our reasons and motives are indeterminate in a deeper way.

To conclude: I have argued, against Arpaly, that there is often no fixed fact of the matter regarding an agent's reasons and motives, that this is connected to motivational and rational change, and that self-opacity in such cases can be conceived as an expression and reflection of this

¹⁶ See Buss' "The Superficial Unity of the Mind" for an argument against the imperative to achieve maximal psychic unity and coherence. See also Calhoun (2009).

¹⁷ For philosophical accounts of why ambivalence is sometimes appropriate, even expressive of agency (rather than expressive of a lack of agency), Benson (2005), Calhoun (1995), Lugones (1987), Poltera (2010).

indeterminacy. I also tried to show that insofar as her account radically de-emphasizes the first-person perspective, Arpaly loses the intuitive thought that we experience self-opacity *as* a practical problem. So Arpaly is certainly right that moral philosophers should attend to the murkiness that is the rule in human life. But murkiness often reaches further than Arpaly herself allows, and there are ways of giving the right kind of emphasis to the internal, practical perspective, even when it is opaque.

Conclusion to Part One

The Deliberative Endorsement theorists were right to insist on the methodological constraint of internalism, which requires that we conceive of agency “from the practical point of view” in such a way as to reflect the essential reflexivity of the self or agent. However, I argued that they were wrong to interpret the practical point of view as the narrowly deliberative. So while self-opacity ought to be conceived as a *practical* problem—a problem for us persons—they were wrong to think that self-opacity should be analyzed as indecisiveness where what is needed is greater deliberative resolve. In focusing on reflective deliberation, DE theorists overlook the human propensity to engage in what I called “defensive” deliberation (and what Kant called our propensity to self-conceit). I also showed that insofar as DE reduces the practical to the deliberative, it cannot make sense of *self-critique* as a properly practical stance. DE is right to insist that agency be understood “from the inside” but wrong to insist that agency should be understood as essentially deliberative.

In Chapter Two I engaged with Nomy Arpaly as an example of a philosopher wants us to see that self-opacity is “the rule” in human life and that moral philosophy needs to take this into its accounts of agency and moral psychology. While I think she is right to insist that we need to pay philosophical attention to complex cases, I tried to show that emphasizing self-opacity does not require that we *de-emphasize* the first-person perspective. Sometimes we can see the “murkiness” of the first-person perspective as manifesting indeterminacy at the level of the agent’s reasons and motives. Moreover we need to keep the agent’s first-person perspective in view, *even* when opaque—this is the only way we can appreciate self-opacity as, not merely some empirical fact about

us, but as something that vexes us, as practical-existential problem which can only be navigated from the practical perspective.

In Part Two I work to show that self-opacity is ineradicable and essential to human agency *and* that grasping self-opacity's significance, its role in our life, involves understanding it from within the practical perspective. I analyze self-opacity as tied to two mutually informing aspects of human life: the fact that human life is *bodily* (Chapter Three) and the fact that human life is *dependent* and *social* (Chapter Four). As we will see, this means we should understand self-opacity as the reflexive dimension of our finitude and non-self-sufficiency. Nagel maintains that as agents we aspire to "encompass ourselves completely," Korsgaard instructs us to "pull ourselves together," and Moran sees no proper limit to the project of self-avowal. The next two chapters examine the ways in which these aspirations will be essentially, necessarily frustrated *and* the ways in which that frustration is registered from the agent's practical perspective (and in the final chapter I argue that such frustration can actually be a good thing).

Part II

*

Chapter Three: The Conflictual Unity of Rational Animals

In her critique of DE, Arpaly insisted that central to the task of articulating a different picture of agency would involve “giving the unconscious its due,” though herself did not engage with Freud. In this chapter I will give an account of why self-opacity is ineradicable for us, but rather than turn directly to his idea of the unconscious, I will instead develop his analysis of drives and sexuality, which I think we can understand as an attempt to articulate how our animality essentially informs human agency. No doubt it is not immediately obvious why sexuality should have anything essential to do with agency. But if human beings are rational animals I will work to show that drives and sexuality are crucial for understanding the nature of the relationship between our rationality and our animality, and how do *we* rational animals experience this relationship, our being at once rational and animal.

My aim in this chapter is to develop an internal conception of human agency which, on the one hand, does not conceive of human beings as metaphysically riven between our animal and rational natures—as though our animality were wholly external to our rationality and could only be viewed from without—and yet on the other hand, is not so unified as to fully discount the idea that there *is* something difficult in trying to hold together our view of ourselves as self-conscious, reflective, often rational agents *and* our view of ourselves as animals. And, I will claim self-opacity is internal to our sense of ourselves as this conflictual kind of agent.

*

In *Mind and World* (1996), John McDowell’s central aim is to lead us away from a picture according to which “an animal endowed with reason would be metaphysically split, with disastrous

consequences for our reflection about empirical thinking and action” (1994, 108). Against a picture which would separate reason from nature, McDowell proposes that being rational is *our way* of being a natural animal: it is our “second nature.” While human beings are born with a “first nature,” we also learn and develop, through participation in a community, a “second nature,” comprised most crucially of our rational and evaluative capacities. McDowell calls the relevant social and educative processes *Bildung*. Thanks to the second natural capacities we develop through *Bildung*, we become responsive not merely to our immediate needs and the demands of the environment, but to the claims of reason and ethics and culture. When we remind ourselves that our rational, evaluative, and moral capacities are second natural, then they can be understood as an expression of, rather than metaphysically split from, our creaturely nature. Moreover, McDowell thinks that this is how we can understand *ourselves* as rational animals, both natural and normative.

Expanding on McDowell, Matthew Boyle has recently introduced the idea of a “transformative theory of rationality,” an account of how we must conceive of the relationship between animality and rationality if we are to make sense of their unity. According to Boyle, we cannot understand our rational capacities as added on to our non-rational animal nature, as though rationality was just one extra capacity amongst others. On such an additive picture, rationality merely supplements what is already there in the animal, and functions as a second order supervisory capacity that can evaluate the deliverances of the animal capacities. Yet on such a picture, there is no way to make sense of *how* the animal and rational capacities could meaningfully interact, how, say, a perception or desire could figure for a person as grounds believing or acting.

Against the additive picture, Boyle proposes that our being rational transforms our very animal being, such that *ours* is the specific kind of animality proper to rational life. Rationality makes an essential, pervasive difference to the nature of our animal capacities. Only as transformed by

rationality can the deliverances of our perceptual and desiderative capacities present to us *as* reasons. So while *Bildung* names the actual, historical-developmental, cultural processes through which our second nature is developed, the idea that rationality transforms animality provides the conceptual framework that makes sense of this possibility: only *because* our animality is essentially transformed can it be educated and developed through *Bildung*.¹

With the concepts of transformation, second nature, and *Bildung*, McDowell and Boyle can be understood as working to clarify how it is that our human capacities are at once continuous with the rest of nature and yet allow us to move beyond mere responsiveness to an *environment* and become intelligibly in touch with the properly human *world* of reason, ethics, and culture. And again, it is meant to be an account of human nature that can be grasped from within.

Yet while they provide important philosophical frameworks for how we should think about human nature, McDowell's and Boyle's accounts are intentionally formal and abstract, and as such do not yield much determinate insight into what a creature so transformed by rationality is actually like. In light of this, it has seemed to some that, despite McDowell's insistence that our second nature is continuous with first nature, it is hard to see how our free, rational, and uniquely human kind of agency could ever emerge from unfree, instinct-bound, non-rational animal life (see Lovibond 2008). Animals and humans are both natural, yet if our nature is transformed, they seem to have no nature in common. If "their" merely animal life involves reacting to their environment in light of their instincts (McDowell 1994, 117) and "our" life involves "rising above" the pressure of instinct into the free space of reasons and culture, how are we meant to understand the continuity from the former to the latter?

¹ The point, to be absolutely clear, is that "transformation" in Boyle's sense is not a developmental process. That our animal capacities are "transformed" is not a claim about something that has happened in time, but is a conceptual claim about what they essentially are. *Bildung* names the educative processes through which our animal and rational capacities are developed.

One approach to this challenge involves pointing out that there is more “rationality” in non-human animals than the transformative picture acknowledges. To take one example, Sabina Lovibond calls attention to what she calls “transitional phenomena” (2008, 120), forms of proto-rationality in animals, so that we can see their practical engagements as more than brutally instinctual in response to a kind of a proto-world (rather than a mere environment). Lovibond’s aim is to find the points of continuity and gradual differentiation between the species, in order to demonstrate the ways in which we share a nature. I want to take a slightly different tack. Rather than argue either that there is already rationality in non-human animals or that there is more of their animality in us, I want to call attention to a different kind of “transitional phenomenon,” a kind of transitionality *in us*. This is a form of embodied mindedness (Lear 1999, 88) that does not bridge the divide between “us” and “them,” but that marks the transition between animality and rationality *in the human being*. That is, I am interested in the dimension of continuity and differentiation within our form of life. To anticipate, the merit of such an account is not only to give a more filled-out picture of our second nature. It will also allow for an essential form of what Rüdiger Bubner calls “productive unrest” (2002, 211), which we need in order to account for a relationship with our culture that is at once natural and critical.

I will argue that this is the conceptual work to which we can put Freud’s concepts of drives and sexuality. Thinking of human being in terms of drives rather than instincts is a way of thinking of our animality as transformed by rationality, as McDowell and Boyle propose, yet where this transformation is more complicated and more conflictual than they imagine. While of course a rational animal must be metaphysically unified, I will argue that the unity of a rational animal is an essentially conflictual unity, where this conflictual unity characterizes the nature of rational animality in such a way as to reflect what it is like to *be* such a creature.

Through a close reading of Freud's *Three Essays on Sexuality*, I will show that the ideas of second nature and transformative rationality help bring out the philosophical stakes of Freud's account, and that in turn, Freud's picture challenges and deepens our understanding of what it is to be at a rational animal. In this, I take myself to be following Hans Loewald's conception of how psychoanalysis approaches the question of the human, which very much anticipates the idea of and need for a "transformative" account of our nature. As he writes:

Psychoanalysis deals with man within the full range of his human potential. As to the somatic events and levels of human functioning, traditionally considered the domain of biology and physiology, psychoanalysis attempts to deal with them from a different viewpoint or within a larger context or framework—as being integrants, constituents, of the psychological organization of human beings, and as such affected by that organization... If the psychology Freud created is a natural science, then we are dealing with concepts of nature and science that include man's moral nature, no less than his biological functions and processes, as topics for the scientific study of "nature." 2000, 535.

I begin by briefly discussing what it means to read Freud as making properly philosophical claims about human nature, before turning to his analysis of drives and sexuality as discussed in the *Three Essays*. I propose that we understand drives and sexuality as constituting a unique order or form of rational animality that is both distinct from and yet bound up with what I will call "the vital order" and "the normative order," the source and arena of *Bildung*. As we shall see, while the sexual order complicates our relationship to the normative order, it also accounts for our being emphatically bound to it. In this way, the concept of drive contributes to our understanding of how our creaturely being is bound up with and transformed by culture and the space of reasons, and yet in such a way that this relationship retains an agonistic dimension. I will also argue that that the agonism that is constitutive of our kind of unity as rational animals allows us to see how critique and creativity with respect to culture are basic to our second nature.

§1. Some Remarks On Method

Before analyzing Freud's theory of drives and sexuality as contributing to and challenging our philosophical understanding of rational animality, we should first briefly raise the question of method and the nature of philosophical claims. Given that McDowell adheres to a rather specific understanding of what philosophy can do, there is a question as to whether and how Freud's discourse can make a properly philosophical contribution to the project of understanding our rational animality and second nature.

McDowell's approach to the idea of second nature is Wittgensteinian in spirit.² On this approach, human conceptual capacities count as second natural when they are not given at birth as part of our first nature, but must be developed, taught, and learned through initiation into a community. These capacities are robustly social and wholly intramundane. While they are nothing other than natural, they cannot be exhaustively explained by the natural sciences but must be understood within the space of reasons, where phenomena are not explained according to laws but justified according to reasons. The space of reasons constitutes the logical space of second nature.³ Paradigmatic instances of second nature for human beings are shared norm-governed practices: from patterns of emotional responsiveness, to reasoning and evaluating, to moral understanding, to sophisticated forms of culture. McDowell writes, "human beings are intelligibly initiated into this stretch of the space of reasons by ethical upbringing, which instills the appropriate shape to their lives" (1996, 79). He calls this process of upbringing *Bildung*. Notoriously, McDowell insists that he is using the concept of *Bildung* only as a philosophical "reminder" meant simply to signal that there is conceptual space for a non-reductive notion of nature. Following Wittgenstein, the concept of

² Other Wittgensteinians interested in second nature include Stanley Cavell, Alice Crary, Cora Diamond, and Sabina Lovibond.

³ Since *Mind and World* McDowell has clarified that nonrational animals can have a second nature. As he writes, "the idea of second nature belongs on both sides of the distinction I am chiefly concerned with, between what can be made intelligible by placement in the space of reasons and everything else" (1999, 99). So we cannot equate second nature with the space of reasons.

second nature is meant to be diagnostic and therapeutic; it should help us recognize captivating but problematic pictures of human beings (as either wholly first natural or as metaphysically exceptional) and see our way of out of them.

An important feature of McDowell's method involves the idea that philosophical progress must be made immanently to what it seeks to understand. If we are trying to clarify something in human experience, we cannot seek to achieve an external standpoint or what he calls a "sideways on" view of ourselves or our relation to the world. So, for example, we cannot step "outside" of rationality, or morality, or—as we are about to see, sexuality—in order to make philosophical sense of them. Instead, philosophical progress must remain "internal." Thus, any account of rationality must be given from within the perspective we have *as* reasoners, any account of morality must be internal to the perspective of a moral form life, and, I will argue, a properly philosophical understanding of sexuality cannot radically depart from our experience and understanding of sexuality *as* sexual beings.

I recall McDowell's philosophical methodology in order to raise the question of what Freud's method is, and whether their discourses can be usefully connected. This question is especially pertinent given that it may not be obvious whether Freud can be understood as making philosophical claims at all: perhaps he is instead an empirical psychologist, or a speculative scientist, or a revolutionary theoretician, or an unconventional clinician. In any event, it might seem that the nature of his work is too far from philosophy to make a properly philosophical contribution. So what then is Freud's method for understanding human beings, and sexuality more specifically, in the *Three Essays*?⁴

⁴ While I am confining myself for the most part to this text, I do think that my arguments about Freud's method can be extended more generally.

Freud begins the *Three Essays* by situating his work with respect to two available discourses on human sexuality, the biological sciences and what he calls “popular opinion,” which conceive of sexuality as, respectively, an empirical datum and as a matter of social convention. Freud begins his text by claiming that these discourses are not adequate for understanding human sexuality. In contrasting his approach to that of biology, Freud is suggesting, as we shall see, that the sexual drive cannot be understood in reductively first natural terms, as, for example, an innate instinct for procreation that operates in us more or less automatically, like hunger. In contrasting his concept of sexuality with what he calls popular opinion, it might seem that Freud is contrasting his account with our ordinary, pre-theoretical understanding of sexuality. This might suggest that Freud is offering a novel substantive theory, one that would be revisionary or “sideways on.” The quietist objection here would be that if Freud is offering a substantive, revisionary theory of sexuality, such a theory would be too “external” with respect to *our* experience of sexuality and its role in our form of life. Of course, no one who advocates quietism is proposing that philosophy simply mirror the claims of popular opinion; a quietist approach to philosophy is not subjectivist or parochial. The worry about Freud distancing his claims from popular opinion is simply that if he departs too radically from the manifest image of sexuality, then his claims will remain tangential to sexuality as it appears in our form of life.

But in fact Freud is not at all indifferent to what we know of sexuality precisely by *being* sexual. In describing his work as “stretching” or “enlarging” the concept of sexuality, Freud is presenting his claims not as wholly external to but as essentially anchored in the deliverances of our experience while also challenging them: his approach is internal but critical. Freud grounds his concept in pre-theoretical, pre-psychoanalytic experience while pressing us to see what we could not or would not see before. Throughout the *Three Essays*, Freud’s rhetoric indicates that he thinks that

whatever is involved in coming to a better understanding of sexuality, it will not be like learning about a topic of which we were simply ignorant, the way we might learn about, for example, our digestive system. So while Freud's approach to human sexuality is no doubt revolutionary, it is meant to remain essentially grounded in the internal knowledge we have of sexuality *as* sexual beings, where sexuality is understood not on the model of an empirical fact about human beings but is rather analyzed as an essential and pervasive *way of being* human.

In this way, Freud's theory of sexuality represents not an empirical discovery regarding the workings of our physiology, but is rather a philosophical clarification of a pervasive aspect of our form of life, one that will prove crucial for understanding how we come to be initiated into the space of reasons and culture. Biology can discover facts about human reproduction, knowledge that ordinary, non-scientifically-educated humans might never possess or understand. It is conceivable to *be* a human without knowing anything about the nature of one's own biological functions. But sexuality is not like this: one might not know the right terms or one's knowledge might be unclear—or, of course, repressed, deflected from or defended against, sublimated, etc.—but to be human is to know about sexuality and its various roles in our form of life precisely in virtue of *being* sexual. In this way, to follow Freud's arguments is to come to know something, not about the species *homo sapiens*, but about oneself. This is not empirical knowledge of an object one also happens to be, but is a kind of self-knowledge.

In the next section I turn to Freud's effort to elucidate human sexuality through the concept of drive. Analyzing the structure of the sexual drive prepares us to see how sexuality both participates in and is crucially different what I will call “the vital order” and “the normative order.”⁵ To anticipate: the concept of drive is meant to occupy the conceptual space between the

⁵ I separate these for the sake of analysis, but do not mean to suggest that they are ever neatly distinguishable in fact.

physiological and the psychological, to capture the idea that for rational animals, the two can neither be neatly separated nor neatly unified.

§2. Sexual Form

In this section I argue that Freud presents a *formal* analysis of sexuality, where being sexual constitutes a mode of being in relation to both self and world. While it will turn out that sexuality must be understood in its connection with and difference from both the vital order and the normative order, we will not be able to understand these relationships until we appreciate the structure of the sexual drive.

While his method obviously involves empirical, clinical observation, Freud's introduction of the idea of drive should not be conceived as the discovery of a heretofore-unknown fact. Rather the drive is a concept used to make sense of certain pervasive dynamics in human life that can be seen as sexual in narrower and wider senses, and yet which serve no biological purpose and do not fall in line with sexual conventions. We can understand Freud's question as the following: *given* that that various acts, fantasies, and states of affairs register as in some sense sexual, or at least charged, and yet do not apparently have anything to do with reproduction or heterosexual sexuality, what must human sexuality be like? Which is also to ask: what must we be like?

Sexuality must be understood formally because, rather than define sexuality with reference to particular objects, Freud defines sexuality through the concept of the drive (*Trieb*), which, as we will now see, has an *object*, an *aim*, a quantity of *force*, and a *bodily source*. The sexual drive is the form or structure that organizes these terms. I will move through these terms rather quickly, in order to get the structure into view.

Freud introduces the idea of the sexual drive's *object* in his discussion of sexual perversions. One of the main innovations of Freud's analysis of homosexuality—a deviation with regards to object—is the idea that the sexual drive is not essentially directed at any specific object (the way that the instinct of hunger, for instance, is essentially directed at food). As he puts it:

we have been in the habit of regarding the connection between the sexual drive and the sexual object as more intimate than it in fact is. Experience of the cases that are considered abnormal has shown us that in them the sexual drive and the sexual object are merely soldered together. 2000, 14.

The point here is not that the drive has no formal object or that drive and object are only contingently connected. To the contrary, the drive always takes *some* object, not necessarily in the sense of a physical thing, but in the sense of a “focus of passion,” as Jean Laplace puts it (1985, 11); the drive's object can be a thing, a person, a part of the subject's own body, a fantasy, a practice, a value, and so on. While there is always some focus of passion, the particular object itself is contingent.

Regarding the drive's *aim*, Freud shows that while popular opinion tends to think that the “normal” or “natural” sexual aim is heterosexual sex, what is essential is actually not the act itself but what the act facilitates, namely “a release of sexual tension” (2000, 15). Thus the drive's formal aim is reduction of sexual tension. (In the next section, we will see what makes tension, and the pleasure that comes from reducing it, properly sexual).

That the drive aims at tension reduction indicates, *third*, that the sexual drive is felt *as* tension by exercising some quantity of *pressure*. This can be understood as the drive's essential phenomenology, the way we register it. Freud writes that the sexual drive lies “on the frontier between the mental and the physical [and is]...the psychical representative of an endosomatic, continuously flowing source of stimulation...a demand made on the mind for work” (2000, 34). Later, in “Instincts and their Vicissitudes,” Freud adds that the characteristic of exercising pressure

on the mind is the *essence* of the drives, and that this pressure made on the mind “for work” is a *consequence* of its connection with the body. So the drive is the way the body put pressure on the mind, which means that the mind is—to paraphrase Freud himself—first and foremost a bodily mind.⁶ As Jonathan Lear puts this point, “the idea of a drive...helps us to understand what it is to be an embodied mind” (1999, 88). Again, whereas an animal instinct specifies its proper object and is experienced as a need *for* that object, hence as the demand to do work to achieve that object, a drive puts pressure on the mind to engage in whatever work, with respect to whatever object, that will function to release tension. So a drive is a demand made on the mind that does not specify *what* it needs, only *that* it needs.

Finally, if the drive makes a demand on the mind in consequence of its connection with the body, this means that the drive has its *source* in the body, in some bodily organ, process, or zone. Freud’s purpose here is not to give a causal account of the actual physiological mechanism underlying the drive (though he often lamented the fact that he could not provide a scientific account of the nature of this mechanism), but is to make perspicuous that the drive is *of* the body. Again, if the drive is a demand made on the mind in consequence of its connection with the body, then drives can only be predicated of a bodily mind (where, to anticipate, the mind in question is a mind of *Bildung*, a cultured mind).

We see now how the drive describes a formal relationship holding between four terms: object, aim, pressure, and source. While his method is based on the study of concrete cases, we need to also appreciate that Freud is not only providing an account of the human animal gleaned from empirical observation. Rather he is working to articulate the *form* of the drive, which is not an empirical discovery but a philosophical claim about the conditions for certain dynamics of

⁶ In *The Ego and the Id* Freud states that the ego is first and foremost a bodily ego.

experience. This form tells us little of what sexuality will look like in any given case: all we have is a structure. But the idea of sexuality as structure is exactly what is needed in order to account for the fact that human sexuality manifests in various and unexpected ways.

Part of what is shown by thinking of sexuality in terms of structure rather than substance is that something can only count *as* sexual insofar as it stands in the appropriate formal nexus; put otherwise, there is no preexisting or essential sexual content, only sexual form. Once we understand this claim, we can see that the error common to both biology and popular opinion was to mistake (contingent) sexual content for sexuality as such. While of course procreation and heterosexual sex can be sexual, those specific activities do not in fact define human sexuality. So, for Freud, sexuality is not best understood as an act or object or state of affairs but instead describes a mode of relating to oneself and the world that is characterized by an emphatic form of attachment, what we could describe as an opaque needfulness.

That the drive is at once emphatic and opaque explains two features of human life: first, it explains why sexuality can be so effectively saturated with the values and meanings that come from the normative, cultural order; second, it explains why we can be so emphatically attached to value and meaning. So while McDowell characterizes second nature as what allows us to rise above the pressure of need into the free space of reasons and culture, I hope to show that the notion of drive indicates that we are not exactly free and distanced with respect to culture but emphatically, opaquely attached to it, subject to a new kind of pressure (“new” in the sense of: not first natural). Put differently: that we are initiated into culture through certain configurations of the sexual form means that we gain not only a set of moral and evaluative concepts through which to understand the world; we also gain a *moral body*, a physiology that both sustains and complicates our orientation in the space of reasons.

I will call this the sexual order, which intersects on the one hand with the vital order and on the other hand with the normative order, the order of *Bildung*. As we will see, the sexual drive is informed by and inseparable from both our first-natural animality⁷ and our rational capacities and culture. And yet while it is essentially both—or *because* it is both—it does not constitute a straightforward unity. Rather, the sexual names the productive, lived tension between animality and rationality for a being that is both.

Let me say something in anticipation of the discussion of the normative order. Part of what the concept of drive helps clarify is the basic idea of *mattering*. Freud saw that attitudes, ideals, actions, relationships and so forth matter to human beings in ways that cannot be explained either in terms of biological necessity or in terms of rational justification. We are *invested* in the world (and the world in us), and the concept of drive is meant to explain the nature of that investment, to explain why the world matters (and how it can fail to), in a twofold sense: the world matters in the sense that it is significant or meaningful, *and* it matters in the sense that it is experienced not only as subjectively mattering but as substantially, robustly real: it has matter. The world—including the world of values, ideas, and ideals—matters in this twofold sense because we are invested in it (and *into* it, via *Bildung*), and for Freud this cannot be understood only at the level of self-conscious responsiveness to reasons but must involve what I called opaque needfulness, a form of felt bodily investment. This, however, is not the vital body of first nature but is the transformed body of

⁷ On a transformative picture, do we even *have* a first nature? Hasn't everything been transformed by reason? No: only those animal capacities whose deliverances could figure in a judgment are transformed, so primarily perception and desire. Boyle (2011) contrasts the way in which my desire for X presents me with an ostensible reason, from the way my feeling of nausea presents me with some situational fact to cope with. The first nature of a rational animal is constituted by those natural capacities and dispositions that present, not *prima facie* reasons, but obstacles to be dealt with. Drawing these distinctions might raise interesting questions concerning the nature of intensity: while my hunger gives me a *prima facie* reason to get up from my desk and go get food, is there a point at which it will no longer present as a reason but instead as a brute need or situational fact that I must cope with by any means necessary? It seems that, for a rational animal, the same (transformed) capacity could issue in both brute needs *and* reasons, and that the difference may depend on situational factors.

second nature, the body for which things matter, always in both senses. Let us now turn to the vital order so that we can better understand the difference between these two ways of being bodies.

§3: The Vital Order

Freud characterizes the pressure exerted by the drive as “a demand made *on* the mind for work *in consequence* of its connection with the body.” So how should we think of the drive and sexuality vis-à-vis the first-natural body?⁸ What kind of connection is this? As we will see, Freud’s idea is that the sexual is rooted in our animality and yet is more than *merely* animal (and in fact we already know this, insofar as Freud understands sexuality in terms of drive rather than instinct). So the sexual is bodily yet not first natural. Indeed the development of the sexual drive out of first nature can be conceived as one of the most ground-level achievements of *Bildung*, and hence as a manifestation of our second nature.

Freud introduces the idea of the first natural, vital order in connection with the idea of infantile sexuality, when he describes the sexual as “propped on” or leaning on (*Anlehnung*) our vital somatic functions. In his discussion of thumb sucking, exemplary of infantile sexuality, Freud writes:

The satisfaction of the erotogenic zone is associated in the first instance with the satisfaction of the need for nourishment. To begin with, sexual activity attaches itself to functions serving the purpose of self-preservation and does not become independent of them until later...The need for repeating the sexual satisfaction becomes detached from the need for taking nourishment. 2000, 47-8.

So, Freud regards the process of feeding as essentially satisfying, but it is satisfying in two distinct ways. Feeding satisfies the vital, first-natural hunger instinct by providing it with its proper object, nourishment: the instinct to feed is the instinct to eat, the demand is transparent to its object. But

⁸ McDowell thinks of infants as first-natural, because non-rational: “human infants are mere animals, distinctive only in their potential, and nothing occult happens to a human being in ordinary upbringing” (1994, 123).

feeding is also satisfying in a way that is tangential to this vital need: feeding is satisfying in the sense that it sensually pleasurable, where this pleasure is irreducible to the satisfaction of instinct. Initially, this second satisfaction is attached to, is perhaps coincident with, the satisfactions of the hunger instinct. But the drive for satisfaction *qua* pleasure can also operate independently from the instinct for nourishment, and when it does so, the drive is properly sexual. So for example, the pleasure of sucking can be achieved in the absence of actual nourishment (through thumb-sucking), which means that pleasure is possible—and sought—without serving any vital purpose. Importantly, this means the capacity for pleasure detached from need is not an innate instinct but something that must be *learned*. Thus, while the sexual has essential footing in the vital or first-natural order, it is not reducible to or expressive of that order: sexuality is second natural. But to anticipate: sexuality also has an essential footing in the normative, cultural order, so at this moment in the argument, while we have connected and distinguished sexuality from the vital order, this is not enough to characterize human sexuality proper.⁹

Two more points are relevant about the vital and sexual orders.

In the conclusion to the *Three Essays* Freud writes that “sexual excitation is a by-product, as it were, of a large number of processes that occur in the organism, as soon as they reach a certain level of intensity” and “pass beyond certain quantitative limits” (2000, 99, 70). Again, the vagueness of Freud’s reference to organic “processes” indicates that he is not trying to provide a causal account of the genesis of sexual excitation, but is rather making an economic-phenomenological claim about the sexual. What is important for Freud’s purposes is not the physiological source of the sexual drive (a “source=x,” as Laplanche puts it) but the drive’s distinctive economic and

⁹ One reason for the caveat is also to forestall the objection that animals can take pleasure in activities that are unrelated to their vital needs. Because, for Freud, sexuality is essential *both* animal *and* normative, the non-rational animal that experiences pleasure without any connection to its basic vital functions is thereby not a sexual animal (in the human sense of sexual). Rationality, construed, broadly, is necessary for *our* sexuality.

phenomenological characteristic, namely its intensity “passing beyond certain limits,” its being felt to be *too much*.¹⁰ This suggests that the sexual is essentially excessive: a sexual being is a rational animal that cannot everywhere manage the intensities of its life.¹¹ But this does not mean that the sexual is simply *more* life, as though the sexual constituted a merely quantitative difference that remained within the same vital order. Instead, the quantity of excitation is registered not just as more but as *too much*, where this involves a shift to a different register, from the merely vital to the emergently sexual.¹²

The second point I want to make concerning the vital and the sexual is structural. Alenka Zupančič claims that “every satisfaction of a need brings with it the possibility of a supplementary satisfaction” (2008, 18). If the tension of vital need reaches a certain level of intensity, the site of this tension makes possible a secondary, non-vital need and satisfaction: sexual need and sexual satisfaction. Zupančič argues that if, as Freud claims, any vital function has the potential to produce this supplement—i.e. can support an erotogenic zone—this indicates that there is no first-natural human need that is absolutely complete and wholly self-contained. If any of our vital needs can become too much, and if any need can thereby provide an occasion for sexual need and satisfaction, then the sexual order should not be conceived as additional to the vital order, with the latter unchanged by the presence of the former. Rather, the very nature of the vital order is transformed by the sexual. So, borrowing from Boyle’s account of how to understand rational animals, we can

¹⁰ On the idea of “too muchness” see also Jonathan Lear (2000) and Eric Santner (2001). Later Freud will move beyond the idea that pleasure and pain can be thought in purely quantitative terms, and will begin to consider the issue of quality.

¹¹ A full version of this account would have to include a discussion of the role that the period of infantile helplessness plays in psychic formation. This is relevant because the infant experiences needs that it is physically unable to satisfy, needs which then become too much, and then take on a sexual characteristic. This would also be part of the story of the role of wish and phantasy in psychic structure.

¹² Taking a phrase from Wittgenstein, this is a case of the transition from quantity to quality. See Footnote 9 for some thoughts about the role of intensity in distinguishing first-natural, *untransformed* needs and dispositions from the kinds of animal capacities that are transformed by rationality.

say that it is not that human beings share a set of first natural instincts and capacities with all other animals—for example, the capacity to see, the instinct to eat, etc.—to which the sexual instinct is added as a one more capacity. Rather, as I have been arguing, sexuality involves a wholly different order or mode of being, its own proper form, which transforms the vitality on which it relies. The structure of propping means that on the one hand, sexuality is an expression of bodiliness that departs from the vital order, while on the other hand, sexuality saturates the vital order, such that any activity of the latter can become a site of activity for the former. The sexual order makes a difference to the vital order from which it initially finds its support.¹³

But articulating the sexual drive's footing in and departure from the vital order is incomplete as an account of human sexuality. What is needed is an account of how the sexual order and the normative order—the rational and moral capacities that orient us to the space of reasons and culture—are likewise inseparable yet distinct. Because of the relationship with the normative order, everything I've said here about the relationship between the vital and the sexual orders is incomplete and must be heard with the qualification that we need the idea of the normative in order to fully understand the sexual.

§4: The Normative Order

We are trying to understand where human sexuality sits with regards to animality and rationality, which is to say the place of sexuality in human being and the difference it makes to our understanding of the human form of life. We have seen that the sexual order leans on and departs from our first nature, but we do not yet understand sexuality's relationship with the normative order, which is to say, reasons, values, morality, culture, that into which we are initiated through *Bildung*.

¹³ See Matthew Boyle's "Additive Theories of Rationality: A Critique" (2016)

We have seen that human sexuality is not just a brute non-normative, first natural force. We can appreciate this is not a bit of Freudian theory but a pre-theoretical truth by considering that any account of human sexuality that simply did not square with our own experience of sexuality would be to that extent inadequate. Whereas a natural scientific account of blood types is not answerable to human experience, the same cannot be said for sexuality. Sexuality calls for an internal analysis, which means that sexuality is, as Freud saw, not a bit of first nature but is somehow caught up with our self-understanding *as* sexual. So what is its relationship to the normative order? On McDowell's account, the only alternative to causal relationships explicable through the language of science are claims that can be justified in the space of reasons.¹⁴ Yet while it seems clear that sexuality is not a normatively-neutral force, something that merely happens *to* me like a hiccup, its relationship with the normative order is not obvious.

In fact, there is a tension pulling at Freud's own thought throughout the *Three Essays* as to how to understand the sexual vis-à-vis the normative, a tension which can be presented in the form of a question: is the sexual order organized by its *own* norms and standards *or* are sexual norms imported from the normative order, as articulated by a community at a time?

On the one hand, Freud uses ideas of normalcy and perversity throughout his discussion of sexuality. He sometimes describes the sexual drive as organized according to a *telos* that culminates in genital sexual organization, where the opposite sex is taken as the proper sexual object, and penetrative sex is pursued as the proper aim. As he writes, "the sexual instinct of adults arises from

¹⁴ Commentators who criticize McDowell for providing a false alternative between rationally justified claims and brute causation are Finkelstein (2003), Dreyfus (2005), and Bäckström (forthcoming). Writing on the idea of force in German Idealism, Christoph Menke (2013) writes: "the modern dualism of nature and spirit compels us to conceptualize everything in man that is not subject as nature. Here nature is taken in the sense of the modern natural sciences...The critical thesis contained in Herder's concept of obscure force is that if man is not—entirely—subject, it is not because he is nature in the mechanical or biological sense" (46). What Menke says here about Herder is very close to what I want to say about Freud.

a combination of a number of impulses of childhood into a unity, an impulsion with a single aim” (2000, 98). This suggests that he takes there to be a governing norm that is proper to the sexual order itself.

On the other hand, many of Freud’s insights challenge any strict distinction between the normal and the perverse when it comes to sexuality: for instance, the idea of the contingency of the sexual object, that there is a universal disposition to perverse sexuality, that the status of the perverse is, as he says, “purely conventional.” Further, Freud tends to qualify his references to “normal” sexuality: he speaks of that which is “regarded” as normal or of what “passes” as normal, and he often puts “pathological” in scare quotes, suggesting that he is borrowing ideas of normalcy from available popular opinion, rather than deriving them from sexuality itself. So while Freud does sometimes make normative claims about normal and perverse sexuality, there is much in his thought that suggests that the very idea of sexuality or the sexual order rejects any neat distinction.

I think that the tension running through Freud’s thought in fact *cannot* be dissolved, because I think it tracks a tension in the sexual itself. This tension is based in the fact that the sexual order is not self-sufficient: it leans on the vital order, and as we will now see, it is saturated and structured by the claims of reason and value: judgments and values, permissions and prohibitions, ideas of good and bad. But here one might ask: doesn’t the *aim* of the drive—satisfaction—provide the relevant internal norm for the sexual order? (or using slightly later language, doesn’t the pleasure principle provide the sexual standard?). Yes, satisfaction is the standard proper to the sexual drive, but what counts as pleasurable or satisfying will turn out to be more complicated, with the standards of the normative, cultural order directly informing how the drive seeks and what can count as satisfaction.

In the chapter on perversions, Freud begins a section titled “Sexual use of the mucous membrane of the lips and mouth” with another reference to popular opinion, noting that while lip-genital contact is “regarded” as a perversion, lip-lip contact is not.

Those who condemn the [former] practices...are giving way to a feeling of *disgust*, which protects them from accepting sexual aims of this kind. The limits of such disgust are, however, purely conventional... Here, then, our attention is drawn to the factor of disgust, which interferes with the libidinal overvaluation of the sexual object but can in turn be overridden by libido. Disgust seems to be one of the forces which have led to a restriction of the sexual aim... The sexual aim in its strength enjoys overriding this disgust. 2000, 17-18.

I quote this passage in full because, as the pleasure of thumb-sucking exemplifies the propping relationship between the vital and sexual orders, disgust exemplifies the relationship between the sexual and normative orders.

Freud begins this passage by noting that those who categorize certain bodily acts as “perverse” are yielding to the feeling of disgust, one of several inhibitions Freud discusses in the *Three Essays*. He describes inhibitions as “mental forces which impede the course of the sexual instinct and, like dams, restrict its flow” (2000, 43). The central inhibitions Freud discusses here are shame, pity, aesthetic ideals, moral ideals, and physical pain. All can exhibit some judgment of good and bad: while the first four inhibitions express evaluative, rationalizing judgments, physical pain can be understood as a kind of vital standard of badness. In this way, inhibitions articulate the human being’s engagement with self and world by marking some acts, thoughts, feelings, judgments and so on as good and others as bad. For Freud, inhibitions introduce normative, evaluative structure to the body itself.

Inhibitions thus play an essential role in human development. As Freud writes, they restrain the drive within the limits that are “regarded” as normal (2000, 28) and facilitate the growth of a civilized individual (2000, 44). Inhibitions are an essential dynamic of the processes of *Bildung*; they constitute, to use McDowell’s language, one of the “plain facts” of human upbringing (1996, 109),

“a normal part of what it is for a human being to come to maturity” (1996, 84). Both Freud and McDowell agree that *Bildung* and its inhibitions are essential for the shaping of character and moral orientation (though Freud saw that when too forceful or too premature, inhibitions inform the construction of perversions and neuroses).¹⁵

So what are these inhibiting forces and how do they function? Understanding inhibitions as counter “forces” or “dams” that restrict flow suggests that they operate like quasi-physical constraints. On this picture, of all the inhibitions listed above the paradigmatic one would be physical pain. Manifestations of sexuality would meet with eruptions of pain or other counter forces, thereby habituating the suppression of sexual inclinations or producing a reflexive association between sexual inclination and physical pain. The language of dams and the idea of pain as paradigmatic suggests a “force” model of inhibition, and thereby a conception of *Bildung* as a process of behavioral conditioning.

But Freud also describes inhibitions as determining what can be “accepted” as a sexual aim and what kinds of objects can be “valued.” This suggests that the sexual drive is inhibited not by a mere counter force but by a forceful counter *claim*, a normative claim about the acceptable and the valuable. Of all the inhibitions listed above, a “moral ideal” would thus be paradigmatic. While the force model of inhibition pictures the relation between inhibitions and drives as sub-personal and involuntary interactions between causes and effects, the claim model conceives of inhibitions as normative evaluations and thus as operating in culture. As Lovibond notes of these contrasting models, “the distinctively human thing that we get from our upbringing, the thing that goes beyond mere habituation to various routines, is an ability to find our way around within ‘the space of reasons’ established by the culture into which we are being initiated” (2008, 115). The claim model

¹⁵ In *Civilization and its Discontents*, Freud writes that the process of civilization brings about a transformation in the disposition of the drives, where part of such transformation goes into the construction of character. (1962, 43).

provides an account of *Bildung*, not as the conditioning of responses to pain and pleasure, but as the process of shaping the bodily mind or minded body through reason and value, developing our capacities to orient in the space of reasons.

If inhibitions are claims to value and acceptability, then it might seem like we are trying to imagine a conflict taking place at two levels: on the one level, there is the drive, and on the other level, there is a judgment. So long as we imagine drives as in terms of force, we will not be able to make sense of how they can be inhibited and shaped by normative claims. But of course, to imagine the drive as force would be to imagine the drive as reductively first natural, like an instinct, and we have already seen that the sexual drive is second natural, irreducible to first nature and first natural explanation. So if drives are not first natural, and if drives and inhibitions *qua* value judgments can meaningfully interact, what must drives be like for this to be possible?

The idea of a “transformative theory of rationality” (2012, 2016) is essential here. An *additive* theory pictures rationality as supplementary to and separate from merely animal capacities. On this picture, while all animals have capacities for perception and desire, rational animals have an extra, special capacity. Rationality is, as Boyle puts it, “tacked on” to an animal mind that forms an intelligible, environment-responsive system apart from this addition. The problem is that such a theory cannot make sense of the interaction between perception and desire on the one hand, and reason on the other. That is, it cannot make sense of how some perception or desire could provide even *prima facie* reasons for judging or acting.

According to a *transformative* theory of rationality, by contrast, a rational animal’s animal capacities are transformed by rationality. For us rational animals, perceptual and desiderative episodes must be structured such that we can reason about and evaluate those episodes. This means that a rational animal’s perceptual and desiderative capacities will be essentially, structurally different

than the non-rational animal's. On this picture, my seeing some object gives me a reason to judge, my desiring some object gives me ostensible reason to act. "Being rational transforms the nature of being an animal, and thus constitutes a new way of being a living thing" (2012, 410). Can this framework help us understand how drives and inhibitions interact? It can, with some qualifications.

If we are to make sense of the idea of inhibitions not as forces but as normative claims, then the sexual drive must itself be transformed by rationality. The point here is not that our sexual nature goes through some temporally extended process of rational transformation—that is, at the moment we are not talking about *Bildung*. The point is this: if drives can be informed or inhibited by judgments of value, acceptability and unacceptability, this means that the drives must proffer their own claims to acceptability and value, good and bad. Consider that salivating when hungry or jerking away from something painful are reflexive, first natural responses that don't offer prima facie reasons so much as situational factors to deal with; moreover value judgments will be powerless to inhibit these bodily responses. By contrast, if sexuality can be shaped by value judgments and claims, this means that the sexual drive must itself be already evaluatively structured and claim making. Or, more precisely: the idea is that the vital capacities of a rational, evaluating, minded animal will be transformed by conceptual and evaluative capacities *such that they become sexual in the first place*.

So, in the first place, the sexual drives are essentially transformed by rationality. And in the second place (though this is no part of Boyle or McDowell's account), value judgments or inhibitions *qua* claims to acceptability must themselves have a kind of force, they must be charged or invested (*Besetzung*). For a value judgment to inhibit the sexual drive it must be, as it were, a forceful or "emotionally cathected" (2000, 30) idea. In these ways, drives and inhibitions must be symmetrically structured.

So the sexual drive must have conceptual structure and the inhibiting idea must be libidinally charged if we are to make sense of their interaction (and note that we are concerned here with a philosophical, even transcendental question, not an empirical question about psychological operations). But at this point one might picture the relation holding between drive and its inhibition as one of sheer opposition, where the latter restricts the former from without. On such a view, there is the sexual drive apart from and prior to its inhibition, which means that an uninhibited sexual drive is conceivable, if rare in actuality (this would be the idea of a drive without vicissitudes). And again, this would result in a picture of the drive as only contingently related to inhibition, which is to say *Bildung*.

To see whether this was Freud's idea, let us return to the quote above. Freud writes that while disgust—or any other inhibition—can interfere with the sexual drive by making its aim or object unacceptable, this inhibition “can in turn be overridden by libido. Disgust seems to be one of the forces which have led to a restriction of the sexual aim...The sexual aim in its strength *enjoys overriding* this disgust.” What this suggests is that the drive and its inhibition are not in fact externally and contingently opposed but rather stand in an internal relationship, with the drive actualized or expressed *by* overriding, even if only temporarily or minimally, its inhibition. If satisfaction is the drive's formal aim, and if the drive finds satisfaction in overriding its inhibition, and, then the drive realizes its aim *by* challenging its inhibition.¹⁶ Thus there are two logical moments to the sexual drive: first, the sexual is propped on and departs from the vital order; second, the sexual is organized by and transgresses the inhibitions of the normative order. But if this is so, then while drives are transformed by rationality, this does not make them wholly reasonable in the way that Boyle takes “ordinary desires” to be “presumptively reasonable” (2016). In fact, if drives are characterized by

¹⁶ A similar conception of the drive as essentially transgressive can be found in Nietzsche. See Paul Katsafanas' *Agency and the Foundations of Ethics: Nietzschean Constitutivism* (Oxford, 2013) for an elaboration of this idea.

their tendency to challenge conventional inhibitions, then drives seem to provide something like a natural site of resistance to norms which, while bodily, are thoroughly second natural.¹⁷

I have been trying to show how it is that drives and inhibitions interact as contesting claims to value. Only in this way can we make sense of how judgments of value might be effective in inhibiting the expression of the drive. But of course one should want to protest here: “but such value judgments often are *not* effective! Our sexual urges are often recalcitrant and insistent in the face of efforts to inhibit them!” Calling attention to the drives’ recalcitrance may make them seem much more like instincts than normative claims.

First, note that to say that value judgments can—*structurally* speaking—inhibit the sexual drive is not to say that just any such judgment can do so or that it will always be spontaneously effective. The point about inhibition is conceptual, and can be brought out negatively: given that judgments of value and acceptability *can* inhibit the drive, given that *Bildung* involves not mere physical conditioning but orientation into reason and culture, then neither drive nor inhibition can be mere forces.

But second, to say that claims to value and acceptability can inhibit the drive is precisely *not* to say that they thereby extinguish it. Rather, it is the drive’s very responsiveness to inhibitions that accounts for and demonstrates its creative recalcitrance. As we’ve seen, it is a basic characteristic of drives, unlike instincts, that they have no proper object and that they operate by making opaque, non-specified demands on the mind for work. This means that if one object or activity is inhibited, it is proper to the drive to *find* an alternative: this is the sense in which the drive is at once creative

¹⁷ In the Afterword to *Mind and World*, McDowell writes that it “would be absurd to suppose that *Bildung* effects a transfiguration, so to speak, of everything that happens in a human life.” And he continues in a footnote: “even those aspects of mature human life that are shaped by *Bildung* show unassimilated residues from their evolution out of mere nature (first nature). That is a way of putting a central thought of Freud” (1996, 183). I am saying something very close but importantly different: the real sites of resistance to *Bildung* are not residues of *mere* i.e. first nature, but are the drives, which are shaped—if perhaps incompletely—by *Bildung*. I will return to this thought of McDowell’s below.

and recalcitrant in virtue of its responsiveness to inhibition. In “Instincts and their Vicissitudes,”

Freud writes:

although the ultimate aim of each instinct remains unchangeable, there may yet be different paths leading to the same ultimate aim...Experience permits us also to speak of instincts which are ‘inhibited in their aim’, in the case of processes which are allowed to make some advance towards instinctual satisfaction but are then inhibited or deflected. 2008, 76.

The aim of a drive is “in every instance satisfaction,” but inhibited drives find different routes to satisfaction: they find different, less directly sexual forms of expression. Inhibitions re-route the sexual drives “wholly or in part, from their sexual use and direct them to other ends” (2000, 178). Inhibitions thus do not extinguish or silence the demands of the drive, but enable it to invest in other objects and activities. So a fundamental dynamic of *Bildung* is inhibition, *not* in the sense that it “shuts down” the drives, but in that it creates alternative and apparently more acceptable objects and modes of expression for the drive. The drive *learns* to be satisfied in new ways. This diversion of instinctual forces from sexual to new, non-sexual aims is what Freud calls sublimation, a vicissitude of the drive which allows it to find satisfaction in mental and social activity, which is to say, in culture and the space of reasons.¹⁸ With the drives, then, Freud provides a way of understanding, quite simply, our *investment* in the space of reasons and culture.

§5. Resistance and Opacity: Essential Forms of Our Second Nature

In this final section I want to clarify how Freud has provided a unique account of our investment and orientation in culture, hence a unique account of agency. On McDowell’s view, as we’ve seen,

¹⁸ It seems that sublimation is a vicissitude particular to the sexual instincts, and may not be applicable to the aggressive instincts, what Freud will later refer to as the death drive. In his essay on Leonardo Da Vinci, Freud notes that “the sexual instinct is particularly well fitted to make contributions [to professional and creative activity] since it is endowed with a capacity for sublimation: that is, it has the power to replace its immediate aim by other aims which may be valued more highly and which are not sexual” (1910, 78).

the space of reasons is the site of rational agency and remains logically (though not metaphysically) distinct from our first nature. Reason is authoritative for us because we have been “emancipated” from need, such that “the shape of our lives is no longer determined by immediate biological forces” (1996, 115).

McDowell saw that we human beings are oriented by and make sense of ourselves and our world through reasons, and yet as natural beings, we are also determined by natural forces that do not lend themselves to the same internal self-understanding. While I, *qua* agent, am responsible for my beliefs, desires, and actions, and can provide grounds for them, I am not similarly responsible for what I am *qua* merely natural being. While I enjoy self-knowledge with regards to my beliefs, desires, and intentions, I know myself *qua* natural being only observationally, and I relate to that dimension of my life as a situational factor or force to be dealt with.

Considered in light of McDowell’s account of natural, rational agency, Freud’s theory of drives and sexuality reveals an essential way of being human that cannot be wholly recuperated within the space of reasons and yet cannot be relegated to the realm of normatively-neutral causation. On the one hand, while the sexual order is intimately connected to and grounded in the vital order, it is not reducible to that order; as such, it cannot be understood as a kind of natural constraint or force which I must manage but that remains essentially external. On the other hand, the sexual order is intimately connected to the normative order—it is shaped by and gives shape to that order—and yet it does not lend itself to full agential identification or authorship. For Freud, being sexual is not just one more capacity but is essential, and essentially transformative, to our rational animality. It represents a way in which we are invested in the space of reasons, and yet for all that, it is also a source of opacity for our self-conscious, rational agency.

In the previous section I argued that we need to understand the sexual order rationally transformed in order to make sense of the interaction between drive and inhibition, and also to make sense of our investment in non-sexual objects and activity. Because the sexual describes a formal relationship rather than any specific content, because its object is contingent, because its responsiveness to judgments of value is at once recalcitrant and creative, all of this allows us to make sense of our emphatic attachment to our rational and cultural forms of life.

But perhaps this all renders us *too* attached, thus closing the space for rational reflection and critique. If our very bodily being is formed through norms, how can we “step back” and subject them to rational assessment? While McDowell offers the image of Neurath’s boat to characterize the critical gesture—where we don’t step *out* of culture in order to critique it, but rather move around the boat, occupying certain corners of it in order to criticize others—Freud seems to be offering an image according to which, we are the boat itself! How could we come to take any reflective distance what are actually the very conditions for our thinking and being? Where is the space for critique?

I do think that Freud’s brings out not only how difficult critique can be, but how easy it can be to “perform” critical distance while achieving none—so, for example, engaging in the gestures of critique without ever letting one’s defenses down, without ever really risking one’s sense of self in the way that real critique would seem to demand. That is, if the object of critique is not a norm or purported reason conceived as something “out there” but the *self as* constituted by such norms and reasons, then critique must always be self-critique, where this not only shrinks the space of possible reflective distance but increases the likelihood that we will be inclined to avoid and defend against such a practice (this is the very problem of defensive reflection from Chapter One).

I think that in fact for Freud the drives can be seen as providing a capacity for critique and resistance that takes the form, not of stepping back and reflecting, but of an internal disruptiveness to our second nature, in the relation between drives and culture: this is the source of the “productive unrest” that Bubner was concerned went missing on McDowell’s account of second nature. On Freud’s account of drives and inhibition, our orientation through the sexual order means that our orientation and occupancy of the normative, cultural order will be essentially fraught, rather than only contingently so (though this is not to say that we live in a state of permanent crisis). It is through drives that we are invested in the space of reasons, and if drives seek satisfaction not in spite of but *by* contesting their inhibitions, then we rational animals are funded with a positive, non-accidental source of potential resistance to the inhibitions and values of culture. The creativity and recalcitrance of drives provides a source of resistance that is essentially different than reason’s capacity to step back: it is not that we can step away and reflect whenever we choose. The idea is that our occupancy of the space of reasons and culture is both sustained and troubled by something in us which is neither wholly rational or acculturated, nor brute first nature.

In the “Afterword” to *Mind and World*, McDowell notes: “it is part of what I want to insist on that we are animals too...it would be absurd to suppose that *Bildung* effects a transfiguration, so to speak, of everything that happens in a human life.” And then in a footnote he continues: “even those aspects of mature human life that are shaped by *Bildung* show unassimilated residues from their evolution out of mere nature (first nature). That is a way of putting a central thought of Freud” (1996, 183, footnote).

I think this is *almost* a way of putting Freud’s thought. I have tried to show that it is not that we harbor unassimilated residues of *mere*, first nature, but that whatever in us that is “unassimilated” is still informed by second nature. If drives name whatever unassimilated residues are not wholly

transfigured by *Bildung*, this is not because they are other or prior to the dynamics of second nature. As libidinal, we are thoroughly second natural and yet resist full assimilation to that very nature. Such resistance is not a way of remaining “outside” culture but a recalcitrant, creative, and bodily way of being in it. By finding a dimension of second nature that is not wholly assimilated without thereby being simply *unassimilated*, Freud locates a second natural site of productive unrest.

Regarding reason’s capacity for critical self-assessment McDowell notes that “reflection can break out at any time” (2002, 190). He means that nothing rational is immune to reflection, so we can step back “at any time.” Freud’s model allows us to say something different: *resistance* can break out any time, not because we can always choose to step back, but because *our* form of animality might, at any time, resist.

*

In *Civilization and its Discontents*, Freud writes: “sex is a biological fact which, although it is of extraordinary importance in mental life, is hard to grasp psychologically” (1962, 52, footnote 3). Sexuality is extraordinarily important to us when it is understood not biologically but as something that both makes opaque demands on us and at the same time expresses us.¹⁹ And yet while it is of such extraordinary importance to us, in mental life, it is hard for that very mind to grasp, which is to say, it remains opaque: not as something external to the mental but as an ungraspable dimension *within* the mental.

From everything I have argued, it should be clear that Freud is not suggesting something interesting about sex that is nevertheless circumscribed in its scope: i.e., concerns some subsection of human acts and related activities (sexual acts, fantasies relationships, etc.) that seem important but

¹⁹ To clarify: while Freud thinks that sexuality cannot be understood according to the natural sciences, his point is not that there is way to occupy a scientific perspective on sexuality. Like most everything else in human life, it is possible to give it a natural scientific description. His point is that such a description does not get at what sexuality is *for us*. Or rather, such a description would concern physiology not sexuality.

are “hard to grasp.” On this view, Freud is telling us about a particular, strange subject matter. If this were all Freud were saying, it would be hard to see why anything Freud says about the sexual order should bear on our broader understanding of our form of life. But we have seen that the sexual is not a thing but a form, where sexuality is not an extra capacity or activity but a *way* of human being, and so it gets things clearly wrong to hear Freud as making a claim about a special, limited subject matter. Borrowing an insight about ethics from Wittgensteinians like Alice Crary, Cora Diamond, and Raimond Gaita, we should see the sexual as concerning not some limited part of life but as concerning a *dimension* that pervades our form of life. As Gaita writes, “our mortality, our vulnerability to misfortune, our sexuality, are not merely matters upon which we exercise our sensibility. They define its forms. We see the world *as* mortal, vulnerable, sexual beings” (2013, 283).²⁰ So when Freud notes that sex is extraordinarily important yet hard to grasp, he is characterizing, not an important, difficult topic, but a sense of importance and difficulty that characterizes our lives as minded creatures, a way we experience the human world.

We have seen that the process of *Bildung* functions to constrain and shape the expression of our drives, where this process binds us to the normative order, where, again, this is only possible because these are *drives* rather than instincts. Our investment in the space of reasons and values gives us a world that matters to us; that this essentially involves inhibitions means that our relationship with these values will be fraught and conflictual. But this should not be taken to mean that such conflict gets in the way of our caring about moral and other values. I would suggest

²⁰ In *A Common Humanity*, Gaita suggests that any properly ethical relationship with a person or group of persons depends upon (amongst other things) appreciating the way that sexuality is of extraordinary importance to their lives, that sexuality could *mean* to “them” what it does to “us.” Gaita thinks that appreciating the importance of sexuality in a person’s mental life is internal to appreciating that person as meriting ethical attention. I think this is an extremely sensitive, humane way of pointing to the subtle ways in which sexuality and ethical life are internally related, and the way that a distorted view of someone’s or some group’s sexuality (as deficient or excessive or otherwise “unlike ours”) constitutes a kind of ethical blindness, even an ethical wrong.

instead that on Freud's account such conflict is an expression of the profound depth of our cares and attachments.

The intimate relationship between the vital, the sexual, and the normative also helps explain how negative moral judgments matter like they do. If morality is our second nature, and if this means that through *Bildung* we develop a moral body and a moral world, then threats to our moral sense of ourselves will have the power to be felt as threats to the self and its hold on the world. That moral judgments can be this powerful, this visceral, would then make sense of our propensity to all the defensive mechanisms Freud came to recognize, which can be understood as efforts *not* to know ourselves in order to hold ourselves together.²¹

The extraordinary importance of the sexual *and thereby* of judgments of good and bad is “hard to grasp” because the sexual is neither rational nor vital but essentially, conflictually both. At a deep level, our basic attachment to the values that give structure and significance to our lives will not be fully articulable or fully recuperable by the agent as “her own” in any straightforward sense.²² This will mark a limit to the project of self-avowal (cf. Gardner 2004, 266-267). Writing on second nature and ethical formation, Sabina Lovibond notes that the goal of these processes is to bring a person to the place where she can be the real *author* of her judgments and actions, endorse them wholeheartedly as her own. However, she notes, in some cases human action and judgment retain “something of the foreignness with which moral demands and prohibitions naturally present

²¹ See Carla Bagnoli's paper on self-deception for a similar idea. She writes: “Self-deception is a phenomenon typical and distinctive of animals that hold ideals and representations of themselves. *It is because we are self-reflective animals* capable of designing representations of ourselves *that we are liable to self-deception*. The self-deceptive agent is concerned with the coherence and stability of her emotional and epistemic system as any rational agent would be.” I would simply add that while cases of full-blown self-deception involve discounting specific and available evidence that would directly challenge the agent's self-conception, ordinary self-opacity involves the more subtle ongoing work of interpreting oneself and one's world in ways one can live with.

²² This is not the same point as McDowell's idea of the “uncodifiability” of ethics. He thinks that virtue cannot be codified in a set of rules or principles, where such a set could be articulable in advance of any particular situation or encounter.

themselves to us as learners” (2002, 100). In these cases, for Lovibond, we have not fully inhabited the second natural judgment or activity, it is not yet properly *ours*, we cannot get behind it as its author. In these cases our judgments and action retain a kind of “foreignness.” What I am suggesting is that the sexual order provides a fund of essential foreignness, which characterizes not an unusual or rare subset of cases, but a dimension of our form of life as rational animals, which can become more or less pronounced. That this extraordinarily important dimension is hard to grasp is not accidental; I would venture to say that this nexus of opaque needfulness and value derives some of its felt importance for us precisely *because* it is so hard to grasp.²³

In emphasizing our conflictual being, I have tried to provide a counterweight to a tendency in philosophy that emphasizes our unproblematic straightforward unity. While of course we rational animals are and must be metaphysically unified natural beings, the risk is that in focusing only on this, we dismiss the ways in which we are split or conflicted or self-opaque as merely contingent privations, inessential to what we most basically are or as irrelevant to philosophy. I have tried to show, with Freud, the way in which a dimension of opacity and productive unrest is actually essential to our form of life, in virtue of our being essentially rational sexual animals, where this dimension both troubles our orientation in the space of reasons, and makes that orientation properly vital. This form of life is neither supernatural nor spooky, as McDowell insists. But it is, as Freud says, extraordinary.

²³ On the idea that opacity and importance are not accidentally related, Iris Murdoch makes the following claim: “in order to tell the truth, especially about anything complicated, we need a conceptual apparatus which partly has the effect of concealing what it attempts to reveal” (1999, 221). Murdoch is talking about art, though I think her words could also characterize the method of psychoanalysis as a clinical practice.

Chapter Four: “The All but Unending Vulnerability of Human Action”

In Chapter Three I worked to show how to understand our animality as a condition of our investment in and attachment to the normative, cultural world and space of reasons, where—*qua* animal—this condition outstrips our capacities for reflective self-understanding. The point of our deepest investment in norms and values marks a limit to the project of self-avowal. In this chapter I will explore the relationship between dependency and self-opacity from a different angle, from the perspective of our engagement with and vulnerability to the social world.

What difference should our essential dependence on the world and others make for how we think about human agency? There is a dilemma here, in that whichever way the answer tends, agency threatens to shrink to an “extensionless point” (Nagel 1991, 35). If agency is made too independent of the world, the risk is that it becomes unrealistically purified and disappointingly powerless, unable to make traction with reality. But if agency is made too thoroughly dependent on the world, it becomes indistinguishable from mere events and thereby lost; of this latter possibility Thomas Nagel notes, “the self which acts... is threatened with dissolution by the absorption of its acts and impulses into the class of events” (1991, 36). If it is too pure, it is not *real* agency; if it is too impure, it is not real *agency*.

This dilemma sets us with the following task: we need to be able to picture agency as at once appropriately active *and* appropriately passive and dependent, such that agency involves both the capacity to make the world yield to my efforts and the capacity to yield to the world. Additionally, we must conceive of our dependency in such a way that does not wholly undermine our sense of ourselves as agents, and yet—as we will see—appropriately tempers our sense of or aspiration to self-sufficiency.

Contemporary Kantian conceptions of agency, and in general those animated by the wish “to be able to encompass ourselves completely, and thus become the absolute source of what we do” (Nagel 1989, 118), have been charged with neglecting the passive side of agency, and so failing to integrate dependency. The charge is that theirs is an overly purified conception of agency. Yet in a little-discussed section in *Self-Constitution*, Korsgaard analyzes what she calls a *standing fact*, namely that our agency is essentially dependent on and vulnerable to the natural and social world. She writes: “it is always true, in every case, that the effects of your actions depend not just on your will but on the force of nature and the actions of other people” (2009, 86). So while critics complain that Kantians do not take the right kind of account of dependency, Korsgaard here seems to be insisting on exactly this. What needs to be seen is how Korsgaard analyzes such dependency and whether her picture is satisfactory.

I will ultimately conclude that it is not. Korsgaard develops a philosophical picture on which our agency is dependent on and *just thereby* threatened by the world, taken “hostage” by it, as she puts it. That is, the natural and social world is conceived as threatening to agency, not in this or that case, but simply in virtue of its being external to the will. Rather than properly integrating dependency as essential to human agency, Korsgaard draws a line between real agency and everything else, where “everything else”—i.e. the world and other people—becomes logically secondary and threatening simply *qua* external.

Seeing why Korsgaard’s picture of agency cannot accommodate our dependency is instructive for what any appropriately realistic conception of agency must look like. In brief, we will need to integrate dependency into agency and understand it as essentially double-sided: while it is in virtue of our dependency that our agency can be threatened, it is also only in virtue of such dependency that it can be realized and real. Considering not only the moment of self-conscious

initiation of action—as Korsgaard does—but more extended scenes of interaction will help us conceive of dependency not as *necessarily* threatening and compromising (though it can be) but as constitutive and as we shall see, an occasion for creativity in action.

§1. Korsgaard's Standing Fact

In *Self-Constitution*, Korsgaard examines what she calls a “standing fact” about human agency, namely that our actions depend not just on our own will but on the world and other people. This fact prompts her to wonder whether agency is really possible at all. So she turns to what she calls the problem of the *efficacy* of human action—its being realizable in the world.

For Korsgaard, we constitute who we are through what we do: “you constitute yourself *as* the author of your actions in the very act of choosing them” (2009, 20). The essential characteristics of an agent are *autonomy* and *efficacy* (2009, 82). An agent is autonomous when her attitudes and actions are rationally self-determined, and she is efficacious when she succeeds in realizing her intention. While Korsgaard does not specify this, it is important to notice that efficacy involves two components: I must *succeed* in realizing my intention, and it must be *me* who does so, the difference I make in the world must be “genuinely” my own as she puts it. While much of the critical literature responding to Korsgaard has focused on her criteria for autonomy, which many find too demanding as criteria for agency in general,¹ I will here focus on the issue of efficacy, as it turns out that Korsgaard's own criteria for efficacy puts pressure on her own conception of agency.

Upon introducing the criterion of efficacy, Korsgaard stops to wonder whether it is actually possible to *be* an agent. She worries that our fundamental dependency on the world and persons

¹ John Kekes puts the criticism in refreshingly frank terms: “Speaking for myself, I act from time to time, have reasons for my actions, and at least sometimes do what is morally right. Yet I have never in my life gone through the process Korsgaard claims is necessary for action, for having reasons, and for doing what is right... I think that what is true of me in this respect is true of the overwhelming majority of people...who do not cease to be human beings on account of having fallen short of Korsgaard's requirements” (2011, 338).

might threaten our agency and our sense of ourselves as agents. With respect to the former, the problem is that if our action's efficacy depends on the contingent grace of the world, then our agency becomes contingent, hence not real agency at all; with respect to the latter, the problem is that if we apprehend the world as threatening in this way, we cannot coherently think of ourselves as agents. The latter is important for Korsgaard because of her commitment to understanding agency from the first person, deliberative perspective. Which means that any threat to agency must be a threat registered as such from the perspective of agency, an existential threat to our *sense of ourselves* as agents. That Korsgaard is concerned with what would threaten our most basic sense of ourselves suggests that her target worry here is not this or that possible threat, but a broad threat with global implications for agency.

There are, Korsgaard writes, two basic threats to agency as such. On the one hand, our autonomy seems threatened by determinism, which Korsgaard calls the problem of “prior causes,” the fact that of any apparently free choice it seems possible to explain it in terms of a prior cause, and thereby undermine its being expressive of one's own agency. We are formed by our genes, our culture, and our character, and this deep shaping seems to threaten the idea that any of our actions or choices are truly autonomous.² But in this section Korsgaard is interested in “the causes that loom up in front of us” (2009, 85) and thereby threaten our efficacy and sense of ourselves as agents. Action must be efficacious: to count as acting I must realize my intention in a way that is genuinely my own. And yet whether my action actually *is* efficacious depends on factors that are

² Korsgaard does not think that determinism ultimately threatens agency or freedom, because, she argues, agency and freedom are practical stances, not theoretical facts the existence of which could be disproved by the truth of determinism. Freedom is not a fact or property but rather describes our condition, as seen from our perspective as agents, insofar as we find ourselves having to decide what to do. Korsgaard maintains that the idea that scientifically-discoverable facts will restrict our freedom and agency is a kind of category mistake. See Korsgaard (1994, 94-97)

outside my will and not up to me. So it seems that agency, the capacity for self-determined, efficacious action, is threatened at the most basic level.

Before we get into the details of the account, it is important to notice the tenor of the set up: Korsgaard frames the very idea of dependency in terms of threat and danger. She worries that such dependency means that “our actions are like children ... there seems to be no telling what might happen to them out there in the dangerous world” (2009, 87). While the possibility of threat certainly comes with our being dependent and vulnerable in the ways we are, framing dependency or non-self-sufficiency in terms of profound, child-like risk evidences an especially fearful way of thinking about this aspect of our life, as though our lack of total control were tantamount to a total lack of control.³

It is because she immediately equates dependency with danger in this way that Korsgaard is led to confess that dependency presents an intractable problem for her account, one she says she “does not know how to solve” (2009, 84). As we go forward, it is important to bear in mind that dependency will seem threatening to the point of paradox *only* on certain pictures of agency: how you think about agency shapes how you think about dependency, and vice versa. If agency involves a demanding conception of self-determination, exemplified in free decision, then dependency will be alarming; but if agency is more socially situated, exemplified in ongoing interaction, then dependency can be more easily accepted as constitutive for acting. This means that we will be faced with a choice: either we should retain Korsgaard’s picture and accept that, since efficacy cannot be guaranteed, we can only “cross our fingers” and hope for the best, *or* we should revise our account,

³ In *Shame and Necessity*, Bernard Williams notes that the Kantian picture of agency involves a fearful (and in this case paranoid) conception of contingency as tantamount to mind-control: “if my values are mine simply in virtue of social and psychological processes to which I have been exposed, then...it is as though I have been brainwashed: I cannot be a fully free, rational, and responsible agent” (2008, 158).

such that the standing fact of our dependency can be accommodated in a more viable, realistic, humane picture of agency. But let us now turn to the task of articulating her account.

Here is how Korsgaard describes acting in the world:

to act—that is, *if* such a thing is possible at all—is to insert yourself—your first-personal, deliberating self—into the causal network. But that inevitably means it is to become hostage to the causal network, to the forces of nature, and the actions of others. 2009, 86.

Such a picture suggests that insofar as our actions depend on conditions not wholly under our control we are, to that extent, wholly *out* of control: we become “hostages.” An agent may choose some action or will some end, and yet insofar as any success or efficacy depends on “the causal network,” then it looks like “the relation between the content of our wills, and the effects of our willings, seems completely contingent” (*ibid.*).

§2. Endangered Agency

Again, Korsgaard’s concern is whether we can coherently conceive of ourselves as agents who can make a difference in the world that is our own *given* that our actions depend not just on our wills but on much else. To see how Korsgaard understands the relationship of dependence between will and world, we need to tease out various elements of her account. Korsgaard operates with what I will call an “intervening forces” conception of dependency: on this model, I set out to act, but then *something intervenes*. While this model makes sense for certain kinds of encounters, I will argue that it cannot helpfully be applied to all cases, and that taking it to be both general and exemplary ends up distorting the nature of our dependency.

In the passage quoted above, Korsgaard claims that in acting one “inserts” oneself into a world and thereby becomes its “hostage,” such that the relationship between the “content” of one’s will and the will’s “effects” becomes completely contingent. Here Korsgaard is concerned with how

the world intervenes between an agent's willing some end and its successful realization in the world. Her paradigm examples of this kind of intervention involve gusts of wind and meddling agents that disrupt agency through brute blockades. These forces "step in, divert the course of the chain, and make the result the opposite of what you intended" (2009, 86). Here the relationship between will and its *intended effect* is undermined by external intervention.

Elsewhere in the same section Korsgaard describes a different kind of case. She writes that "the *meanings* of our actions are determined, not just by what *we* mean by them, but by *how the world takes them up* [such that] there seems to be no telling what might happen to them in the dangerous world" (2009, 87-my emphases). Here the suggestion is that there will be a mismatch between what we mean by our actions and words, and what others think we mean. Korsgaard compares this situation with one in which "the relation between the *content* of your speech and its *effects* in the minds of your audience, becomes (*as perhaps it is already*) completely contingent, or worse" (2009, 86-my emphases).⁴ In this scenario, I say something and yet my words have a different "effect" in my interlocutor's mind than the effect I intended. Here are her examples:

'Let me help you' you say, and the person you say it to backs off in terror, or covers as if expecting a blow. 'You're wonderful' you say, and find you have given the gravest offense. 'Kiss me' you say, and get punched in the stomach. 2009, 86.

While this kind of mismatch between how you mean your words and how someone else takes them up does not always occur, "it sometimes does," and she adds gravely: "acting is like this" (ibid.).

Thus far we have seen two distinct ways in which agency is dependent on and just thereby threatened by the world: the effect of one's will depends on the world, and the meaning or effect of

⁴ Later in the book in her discussion of reasons as essentially public, Korsgaard compares real human communication with what she calls "spin:" "in the case of spin, instead of thinking together when we talk, I think of the meanings that I know certain words have for you as tools that I can use or obstacles I must contend with in the pursuit of my ends. What I say is determined, not by what I think is true, but by the effects that I want to produce on you, the effects I hope my words will have" (2009, 196). Given this definition of spin, it really seems as though the kind of interaction she describes in the section I am analyzing is not conversation but rather spin.

one's words depends on their social uptake (Korsgaard is not clear whether I depend on others to establish the *meaning* of my own words or whether I depend on others for my meaning—established by my intention—to be successfully transmitted. These are different, and I will return to this ambiguity below). Korsgaard also describes a third threat to agency. She writes: “we who must act must cross our fingers and hope that the world, *starting with our own bodies*, will take up our willings in the sense in which we mean them” (ibid.). On this picture, I can will or intend to do something, but my body may not “take up” my willing.

Korsgaard thus presents three ways in which agency is dependent and thereby vulnerable: the world may intervene between will and worldly *effect*, the world may transform the *meaning* of one's actions, and one's will may not be “taken up” by one's own *body*. These are very different relationships involving different kinds of dependency and failure. So, taking Korsgaard's example of trying to help someone: it is one thing to intend to help you and have my efforts sidelined by external forces, whether gusts of wind or other agents; it is another to express an intention to help you and have you react as though you thought I threatened to hurt you (whether that means you've misunderstood me, or I communicated poorly, or something else); and it is yet another to intend to help you and find my body wholly unresponsive. Yet despite these differences, Korsgaard subsumes the latter two under the structure exemplified by the first, treating *intervening forces* as paradigmatic for understanding our dependency on the world.

On this model, there is something I do, and then somewhere down the causal chain, something that happens, which may or may not correspond to my intention, and whether it does or not is out of my hands. Recall again the nature of our relationship with our own bodies: “we who must act must cross our fingers and hope that the world, starting with our own bodies, will take up our willings in the sense in which we mean them” (2009, 87). If, for Korsgaard, I relate to my own

body in the mode of *hope*, then presumably my relationship to the world and to other people will be even more distant, with even more room for intervention. While Korsgaard asserts that really willing an end involves constituting oneself as the cause of the end (2009 68), and is thus different and more than mere wishing, she circumscribes the domain of willing so narrowly as to render it inert: I may will some end, but can only hope that anything will come of this.

Before moving on, I want to recall the idea that our dependency involves not merely a threat to agency but to our *sense* of ourselves. This means that Korsgaard thinks our dependency causes problems not merely at the level of execution, but at the level of our practical self-understanding: it makes a difference to how we conceive of ourselves. But if Korsgaard thinks our dependency can undermine our sense of ourselves, then how does she picture “our sense of ourselves” (apart from or prior to this threat)? This is where her deliberativism shows up. As she writes:

we are conscious of the grounds on which we act, and therefore in control of them...When you deliberately decide what sorts of effects you will bring about in the world, you are also deliberately deciding what kind of cause you will be. And that means you are deciding who you are. 2009, 19.

Or again: “when you deliberate, it is as if there were something over and above all of your desires, something which is *you*, and which *chooses* which desire to act on” (1996, 100). Korsgaard’s standard scene for analyzing agency is the scene of initiating action (inserting oneself into the causal network): it involves an agent becoming “conscious” of some desire or possible course of action, “stepping back” and “reflecting” on it, and then “deliberately deciding” what to do. On this picture, our sense of ourselves as agents involves a sense of ourselves as *over and above*. Such a picture of untethered freedom goes hand in hand with a picture of agents as transparent and lucid about what they are doing and why, where self-awareness is sufficient for self-control (“we are conscious of the grounds on which we act, and therefore in control”). So for Korsgaard our “sense of ourselves” as agents is, as “perfectly self-aware and so perfectly detached and free” (Murdoch 1999, 340): I know exactly

what I am up to and why (“the grounds”), and I can decide what to do (which also means decide who I am), *but then*—in a second, disruptive step—the world intervenes. And if this is the set up, Korsgaard is right to wonder how would it possible to think of oneself as an agent under conditions of dependency and vulnerability.

As we shall see in the next section, I think Korsgaard’s basic structural misstep involves picturing dependency on the intervening forces model, as something like an impinging, external *limit* to our otherwise perfectly detached and free agency. On this picture, our dependency on the world and other people is framed as logically secondary with respect to agency proper, and *thereby* as threatening. While I think Korsgaard is wrong to conflate dependency and threat in the way she does, this is not because I think agency is *invulnerable* or that our vulnerability is, for example, a philosophically inconsequential empirical detail, only relevant when something goes wrong. The point is not to excise dependency and vulnerability from agency but rather to put them together in the right way.

§3. Dualism and Skepticism

In the passages I have been analyzing, Korsgaard appears to waver between two pictures of agency, what Anton Ford (2016) calls “volitionalism” and “corporealism.” On each of these conceptions, a line is drawn between agency proper and agency-by-extension. As Ford writes, on any dualistic philosophy of action, there are two factors at play: “there is what the agent *does* in some emphatic philosophical sense... and there is what is caused by that which the agent *does*” (2016, 6). For Korsgaard, these two factors are, as we have seen, the will and its effect. The question is whether she is a volitionalist or a corporealist.

On a *volitionalist* model of agency, agency is a power to do something even less than move one's own body (2016, 5); it is a kind of inner willing, like the mental setting of an intention or the endorsement of some motive. According to this picture, agency is expressed in willing some content, but nothing else. In describing our practical relationship with our own bodies as one of *hope*, Korsgaard certainly appears to be offering this kind of picture. If acting involves inserting yourself into some external and unpredictable causal network, and if, as she writes, the causal network "starts with *our own bodies*," then indeed it looks as if on Korsgaard's account agency stops short of actual bodily movement. On the volitionalist picture, the threat of intervening forces arises at the level of the body itself.

Yet Korsgaard also provides a different picture, according to which agency extends all the way to bodily movement and speech. When she writes that "normally, I cannot bring about the effect that I intend directly, but rather move to set off some causal chain, which, if all goes well, will lead to the effect that I will," or when she discusses the relationship between what I meaningfully *say* and the *effect* of my words on others, she seems to be adopting a *corporealist* picture of agency. For the corporealist, agency is expressed in bodily movement, nothing less (so not mere willing), but also nothing more. Thus *my* action extends through my body but stops short of the world. That this seems closer to Korsgaard's view is confirmed elsewhere in *Self-Constitution*. As she writes:

actions are not inward or outward; they are, like utterances, intelligible objects in the public world whose very nature is to bridge the gap between inner and outer. They are acts of thinking performed by the body, mental states in the flesh: there is no need to imagine something inner behind them that causes them. 2009, 124.

Action is, on this picture, is not an inner volition that may (or may not) trigger bodily movement; instead agency is emphatically bodily, the very bridge between inner and outer. But notice that, at this point, another bridge appears, that between body and world. So while "inserting oneself into the causal network" now involves self-movement, everything else is contingent and up to chance.

This suggests that on the corporealist picture, the threat arises at the level of the world (perhaps somewhere around the outer edge of the skin).

For my purposes, we do not need to decide whether Korsgaard is a volitionalist or a corporealist. The point is that she draws a line between what I properly do, and everything on which I depend for my agency to be efficacious. This is a picture of human agency as maximally self-aware and capacious within its purview, that stands opposed to everything that would allow it to realize its power in concrete action. The problem, as Korsgaard sees, is that real efficacy matters for agency, and efficacy *necessarily* depends on much that is external to the will. But for Korsgaard, “anything outside of the will”—including an agent’s own desires and other attitudes—is “an alien cause” (2008, 319), which means that anything external to the will—*just* by being external—is sufficient for it to qualify as a threat. Thus Korsgaard wonders: is it possible to be an agent?

But now notice: while Korsgaard attends exclusively to actions that fail in virtue of the world, the threat would seem to hold just as much for the *good* case of successful action. The problem is not merely that world can undermine my sincere efforts and through my actions off course such that they fail. The more basic worry is that because world is essential for realizing any of my intentions, it is *thanks to the world* that my actions succeed when they do.

Recall that Korsgaard defines efficacy as “making a difference in the world that is *genuinely* my own” (2009, 90). There are two dimensions to this definition: first, I must successfully make a difference (successfully realize my intention) and second, the difference I make must be “genuinely my own.” While “genuine” is ambiguous, I take it that what Korsgaard means is that I must be the clear “author,” since she defines action as a movement that is attributable to its author (2009, xi), where such authorship is the basis of responsibility for action (2009, 20). But this suggests that the real problem is that efficacious actions will always involve and depend on more than the capacities

of the self-conscious, self-directing will, the only thing properly my own. So the risk is not just that my actions might fail thanks to the world, but that they might succeed, thanks to the world.

Korsgaard recognizes that agency must be efficacious if it is to be agency proper, which means that agency must involve more than the will. And yet she almost always locates the site of agency—hence authorship and responsibility—in the moment of decision or will or choice: in *deciding* what effects to bring about, you decide who you are (2009, 19), in *choosing* your actions, you create an identity (2009, 20). What I decide to do is “up to me” on Korsgaard’s picture. But once we move beyond the moment of insertion into the causal network, then it seems that my action in the world is not up to me in the same emphatic sense. The difference I make is not “genuinely my own,” but rather seems surrounded on all sides by contingency.

Thus, it seems that if we follow Korsgaard’s logic, we end up heading towards a kind of skepticism: while I have full knowing control over what I will, anything I might actually *do* is, by definition, out of my hands. All I can ever properly do is something less than real worldly action. This dependence on the world is threatening at the most fundamental level, at the level of the point of view of agency (which for Korsgaard, is the only point of view from which these issues matter), not only because it can make things go wrong, but also because it makes things go right. Because, for Korsgaard, I cannot know whether my actions will be realized, I can only cross my fingers and hope.

Before concluding that this picture is skeptical, I want to briefly look at Korsgaard’s “teleological” conception of the world, which she claims is our “natural” way of interacting with the world. On the teleological picture, the world presents as a practical environment articulated by our interests, commitments, cares, purposes, and projects; this world is *our* world. Of course, this does not mean that it is never threatening; it just means that it threatens only in specific contexts, rather

than just *qua* world, just insofar as it falls outside the will. Thus it would seem that on a teleological picture, acting in the world does not involve crossing one's figures and dropping into a causal network, but instead involves transacting with a practical environment. And indeed Korsgaard describes the teleological conception of the world as "essential to our functioning as agents." If this is Korsgaard's considered view, then perhaps we need not saddle her with dualism, skepticism, and all of their problems.

It is hard to know exactly how to reconcile Korsgaard's two pictures of the world of action, the causal network from which an agent is fundamentally distanced and the teleological order with which an agent is fundamentally attuned. It would appear that a teleological conception and the hostage-taking conception are mutually exclusive: if you think of the world in terms of practical affordances, then you cannot conceive of the relation holding between agent and world as "completely contingent;" the whole point of a teleological conception is to show that this relationship describes a kind of practical unity.

Perhaps we can reconcile these pictures by assigning each world-picture to a particular stance: while the teleological world is the world we see as agents, the "causal network" world is the world we see when we *exit* the practical perspective and instead view the world theoretically. When we take such a theoretical perspective on the world, we see it as described by the Modern Scientific Worldview, and precisely cannot, from that perspective, see it as a domain for human action. The problem, though, is that in the "causal network" section I have been analyzing, Korsgaard *is* describing a scene of action *as* apprehended from the deliberative perspective. Recall: "to act—that is, *if* such a thing is possible at all—is to insert yourself—*your first-personal deliberating self*—into the causal network. But that inevitably means it is to become a hostage to the causal network" (my emphasis). In this way, the problem of efficacy is very unlike the problem of determinism: while

Korsgaard thinks that determinism is not a problem for agency because it is a piece of theoretical rather than practical knowledge, efficacy *is* an essentially practical problem: as she emphasizes, dependency threatens *our sense of ourselves as agents*. So Korsgaard is not here concerned with the threat to agency that would arise from taking a disengaged, theoretical view of self and the world, but is rather concerned with a threat to agency that arises *for* agents, from the perspective of agency.⁵ The view of the world as threatening is a way of registering the world of action.

Korsgaard says that she does not know how to solve the problem that dependency poses for her conception of efficacious agency. Given the skeptical direction towards which it points, I agree that the problem cannot be solved, *given* her conception of agency. But this means that if, on her picture, agency as such is threatened by the standing fact of dependency, then the problem might be with her conception. What is needed then is a different conception of agency as more thoroughly dependent, not in a sense that renders us unable to act, but in the sense that our agency does not encounter dependency as something that “looms up in front of us” in a second, threatening step.

Before concluding this discussion of Korsgaard, I want to make explicit why a dualistic picture can seem attractive or compelling, as this will help clarify the philosophical task we are now faced with.

⁵ A further problem concerns the relationship between the teleological world and self-consciousness. Korsgaard is explicit that for self-conscious rational creatures like us, the world—including our own bodies—is not or is no longer immediately normatively significant. In a chapter entitled “Expulsion from the Garden,” Korsgaard claims that self-consciousness introduces a break from our immediate responsiveness to our bodies and the world: “self-consciousness opens up a space between the incentive and the response” (SC 116), a space that makes human beings “essentially homeless” (2009, 118). Because of this space, the world remains “normatively undecided” (2009, 123) until granted authority by the reflectively deliberating agent. As normatively undecided, the deliverances of the world—including the incentives arising from our own bodies—present as “mere psychological states” (*ibid.*), and hence as nothing to me until I authorize them. So although Korsgaard affirms that “our *natural* way of perceiving and conceiving of the world is teleologically,” this must be a nature from which we stand at an essential remove, from which we are, to use her word, expelled. Because self-consciousness fundamentally transforms our relationship with the world, and because she equates self-consciousness with reflective distance, Korsgaard is basically stuck with a dualist picture of agency and world, whether she spells that out in voluntarist or corporeal terms.

We can understand a dualist picture as trying to hold together two elements that must be part of any account of agency: the fact that actions are sourced in the agent—they depend on *her*—and the fact that actions are realized in the world—they depend on *it*. The problem is that the dualist holds these ideas together by holding them apart, framing these components as distinct factors separated by a boundary: there is whatever the agent properly contributes—whatever is up to her in some emphatic sense—and then there is the rest. We should thus see the dualist as trying—unsuccessfully—to provide a solution to a valid requirement for understanding human agency. Part of what makes it unsuccessful is that it is an essentially *defensive* solution, for two reasons: first, it evidences a conviction that impure, dependent agency is no agency at all, and so it defends (pure) agency by walling it off from the world; second, it responds to agency’s dependency on the world in the mode of disavowal, registering it only to reject it as logically secondary. In this way, while the dualist does try to maintain both terms, she does not maintain them on equal footing. Instead the dualist constructs a hierarchy, where the first term is essential and the second term demoted to an inessential add-on: what really matters is what I alone can properly do, and then there is all the rest. In refusing to actually *integrate* dependency into human agency, the dualist lets it drop out of the core picture.

Appreciating how the dualist fails helps us see our task, which is to maintain both components without splitting them apart. Obviously, if we are avoiding dualism we cannot draw the boundary between agency and everything else at some later point (“further out”). Instead we have to give up the very idea of a strong metaphysical boundary. We need to figure out how to appreciate that the realization of agency will outrun conscious intention or will without thereby eradicating

agency or our sense of ourselves. And yet while we are not threatened with practical annihilation it will require adjusting our conception of agency and our sense of ourselves as agents.⁶

Now notice something crucial: if we give up the idea of a firm boundary, this of course cannot mean that agency *as pictured by the dualist* now enjoys unrestricted range of its awesome power and knowledge. If that on which agency essentially depends can no longer be pictured as wholly external, this means that human agency itself can of course no longer be pictured as an unfettered, self-transparent power now with limitless reach.⁷ In giving up dualism and so giving up the metaphysical boundary, we need to reconceive the very nature of agency's power and knowledge, in such a way that what the dualist had ejected from the province of agency—i.e., agency's dependency on the world and others—finds a new place within. This means that Korsgaard is correct to note that the “standing fact” of our dependency puts pressure on our sense of ourselves as agents: making dependency part of agency itself will make a difference to how we understand ourselves as agents. But again, the point is that we can let our dependency make a difference, without that figuring in the form of a threat.

§4. Our Deep Dependency

What would it mean to do the right kind of justice to our dependency, and how might that transform not just our philosophical picture of agency but our sense of ourselves as agents? How can we understand dependency as essential and internal to agency itself, rather than as conceptually secondary? While I will not be able to provide a complete account of dependent agency here, we can start to sketch the outlines of such an account in light of the lessons we've learned from Korsgaard and some insights from Cavell.

⁶ See Krause (2015).

⁷ Ford's way of putting this point is that in giving up dualism cannot mean embracing “triumphalism” (2016, 16).

Margaret Walker points out that *only* if you think of agency as perfectly detached and free will you think that dependency is incompatible with our sense of ourselves as agents. So while Nagel thinks that the very idea of agency is “incompatible with actions being events, and people being things” (1991, 37), it turns out that this sense of brute incompatibility depends on very specific ideas of agency. As Walker writes of “impure agency:”

our human kind of agency [involves] our perfectly predictable entanglements in a causally complex world, with imperfectly predictable results. Part of the normal and required self-understanding of human beings is a grasp of that fact. 1989, 19.

On Walker’s view, our ordinary sense of ourselves as agents is not incompatible with but rather includes a grasp of the perfectly predictable unpredictability of our actions, their dependence on and vulnerability to the larger world.

We can hear Walker as offering an emphatically anti-skeptical response to the dualist. In insisting that the uncertainty and vulnerability of our actions is “perfectly predictable” and that a grasp of this fact is “normal,” Walker’s deflationary phrasing and tone suggests that she thinks Nagel’s and Korsgaard’s worries are the result of confusions that can be dissolved through a better appreciation of our ordinary, impure form of life. In this exchange, the skeptic can sound panicky (“there seems to be no telling what might happen to them out there in the dangerous world!”), and the anti-skeptic responds with reassurances and therapy.

Now, one of Cavell’s greatest philosophical lessons is his observation that in this kind of exchange, the anti-skeptic is not really listening to the skeptic, refuses to hear anything sympathetic or valid in the skeptic’s worries. Cavell argues that even if dualisms and skepticisms should be rejected as metaphysical pictures, still, they give voice to real human experiences that cannot be discounted or dissolved as simply confused. What Cavell finds common to various skeptical pictures is a tendency to transform a real *practical* problem into a philosophical or metaphysical

picture. So, the fact that human agency *can* be compromised by its involvements in the world is translated into a philosophical picture on which human agency is *essentially* compromised by the world, taken hostage by it, to use Korsgaard's phrase.

But if skepticism represents a philosophical response to a practical problem, then our response cannot be: "there's no problem." As Cavell puts it, "there is a fact to be recorded" (2002, 254), and the fact, in the case of dependent agency, is that there is a "*practical* difficulty of pegging the mind to the world, and especially to the social-political world" (1999, 109). The dualist "records" this fact by tracing one (limited) domain in which agency is effectively and knowingly exercised, and then everything else is left to the contingent causal order. Our task of showing how dependency is internal to agency must include an acknowledgment of what the skeptic sees, which is not only that we *are* dependent and vulnerable to worldly threats, but that this fact can be staggering: it *can* feel incompatible with our ordinary self-understanding. The key is to acknowledge that this can throw us without feeling the need to construct an account of agency that deflects from this experience by either quarantining agency from the world or insisting there's no problem.

We have been seeing that how one thinks about human agency determines how one thinks about dependency. If one starts with the moment of self-conscious decision, or with an idealized case cleansed of dependency, then the latter will present as both philosophically alarming *and* as a topic to be dispatched rather quickly: Korsgaard worries that dependency might undermine agency *tout court* and then spends all of three pages on the topic, confessing at the outset that she does not know how to "solve the problem." By contrast, if one *begins* bearing in mind "the all but unending vulnerability of human action, its openness to the independence of the world and the preoccupation of the mind" (Cavell 1996, 87), then dependency on the world and others will be built into one's basic picture, rather than arising in a second step. This means that dependency will not be

philosophically alarming, and that it will be woven throughout one's analysis, rather than treated as a troublesome sub-topic.

How one thinks about human agency will also be exemplified by certain kinds of cases or scenes, where focusing on those cases provides support for certain pictures. Korsgaard focuses on the moment of a self-conscious decision to initiate action, which is then interrupted by intervening forces. Such a scene sets the stage for the world to appear in the guise of a threat. In light of this, it would be worthwhile to consider a different scene of agency to see if it points towards a different conception of dependency. Rather than focus on one-off executions of intention that crash against the world, perhaps the relevant scene to analyze should be a temporally extended, ongoing interaction, a series of reversals of making the world yield and yielding to it, where responsiveness and ongoing negotiation is just as central as the moment of initiation.⁸ This is not to say that choice and initiation of action are not crucial exercises of agency or that philosophy cannot learn from them. The point is just that deliberate choice is not the only way we manifest our agency, and that focusing on the initiating moment can make our dependence on the world seem especially dangerous. Thinking about agency in its more diachronic and dialogic forms is helpful here because it builds dependency right into our most basic ways of thinking about action. If we consider action in an ongoing, interactive way, this might help us see that while our actions and their effects outrun our willed intentions, this need not be taken as a threat to agency itself.

It is noteworthy that in Cavell's philosophy, individual action tends to be submerged within and analyzed as part of protracted interpersonal interactions. This is not only because he thinks that

⁸ John Doris engages this technique of "changing the scene" in his analysis of agency, writing: "agency also has a diachronic character: the exercise of agency emerges not only in discrete, synchronic, actions, but also in temporally extended processes... This realization is not especially evident in philosophical writing on agency and responsibility, which characteristically focuses on more or less discrete behavioral units... rather than complex, temporally extended processes" (2015, 163).

any bit of human action can only be understood within some context and some form of life, but because he thinks that the typical human practical situation involves, not stepping back and surveying one's possibilities, but being always already engaged and responsive, with reflective clarity coming not prior to action, but usually after, if at all. Rather than begin with the moment of launching into the world, Cavell's scenes begin already immersed and underway. This means that in his basic case, we act

not knowing in advance what we may do, what consequences we may be faced with, e.g., that we have been misunderstood, betrayed, rejected; that we have been unperceptive, betraying, rejecting all unknown to ourselves, and without even seeing how. 1999, 325.

Like Korsgaard, Cavell here affirms action's essential unpredictability, the boundlessness of possible consequences, and the ever-present possibility of failure (as he states elsewhere, "*the most characteristic fact* about actions is that they can—in very specific ways—go wrong" [2002, 22]). However, unlike Korsgaard, Cavell's point is not that *we at least know* what we do, but simply cannot see the future or control the world. His point is rather that we act "all unknown to ourselves," where any understanding of what we are doing is developed over time, through what he calls "moral argument" with others. Whereas Korsgaard's agents begin in full awareness of what they are doing and why, for Cavell acting is an unfolding transaction that is always partially in the dark. Is there any way we can use this diachronic, dialogic model for thinking about some of Korsgaard's examples so as to avoid a skeptical picture?

Korsgaard's examples, recall, are the following: 'Let me help you' you say, and the person you say it to backs off in terror, or cowers as if expecting a blow. 'You're wonderful' you say, and find you have given the gravest offense. 'Kiss me' you say, and get punched in the stomach.

In her scenes, I knowingly and with control make some statement, and then *something intervenes*, rendering me inefficacious, end of scene: "things do not turn out as you had set yourself to

make them turn out. Could you still see yourself as a speaker? Of course, it does *not* always happen that way—and yet it sometimes does” (2009, 86-7). In these scenarios, it is as though a gust of wind swept my words wholly off course, such that that they could not realize their intended effect in the mind of my interlocutor. But if we want to understand the role of dependency in human interaction, why use these rather cartoonish examples? Moreover, if, as Korsgaard insists, other human beings are co-reasoners to be regarded with respect, why characterize a failed conversation on the model of intervening forces? (and why describe human dialogue in the rather manipulative terms of eliciting “effects” in someone else’s mind?⁹).

Korsgaard uses these examples to illuminate what she thinks is a standing problem that we already face as agents. Yet it seems to me that these truncated scenes function only to make it more difficult to appreciate how dependency and vulnerability really do shape our lives and bear on our sense of ourselves. So what kind of context would we need in order to bring the “you’re wonderful”/grave offense scenario down to earth, so that we can recognize in what ways acting is already “like this”?

Well, perhaps we are in the middle of a serious fight and at one point I say, “you’re wonderful.” I sincerely mean it—I want you to know that, even in this tense situation, you are extremely special to me. And then imagine that you react with offense, and that I am wholly baffled (we can even imagine that it makes me feel as though the relationship between what I say and what you hear was “completely contingent;” after all, in fights we do sometimes wonder if we are even having the same conversation).

The question is, why should analyzing “the exercise of individual agency as an essentially interpersonal phenomenon... help illuminate the path away from skepticism”? (Doris 2015, 103).

⁹ See footnote 4

How does contextualization change how we understand Korsgaard's scenario, and how does it point towards a different picture of dependency? To begin: if we can only understand the meaning of a word in the context of a sentence, we can really only understand the meaning of a sentence in the context of some larger whole, in this case, an ongoing conversation. Of course, the phrase "you're wonderful" carries with it certain obvious associations that might make us think that we know exactly what it means. But when the recipient responds with offense, then one thing we might think is not that the relationship between the speaker's words and the other's mind is completely contingent, but that we don't actually understand what that phrase means here or how the speaker meant it.

Korsgaard wants us to recognize that something has clearly gone wrong. Probably something has, but she relies on a narrow conception of what it means for a gesture to go wrong and of how we should understand this. According to this intervening forces model, the intended, presumably affectionate meaning was *there* in the phrase "you're wonderful," but, to use her phrasing, something stepped in, diverted the course of the chain, and made the result the opposite of what was intended.

Recall that there was ambiguity regarding Korsgaard's conception of the way meaning depends on uptake. I noted in Section 2 that it was unclear whether I depend on others to *establish* the *meaning* of my own words *or*—a less controversial idea—whether I depend on others for what *I* meant to be successfully communicated, i.e., I depend on them to hear me. Korsgaard writes that "the meanings of our actions are determined, not just by what *we* mean by them, but by the way the world takes them up" (2009, 87). This is a stronger claim of dependency and I think this is right, but I don't think she has actually allowed for this idea. According to her scenarios and her model, the world may take up or fail to take up what we *already meant*. As she writes, "we who must act must

cross our fingers and hope that the world...will take up our willings *in the sense in which we mean them*" (ibid.). The world takes up or disrupts the intended meaning of what I do, where it is not that the *meanings* depend on the world, but rather that the already-present meaning depends on the world in order to realize its intended, meant effect.

But if the meanings of our actions and words really are determined by the way the world takes them up, this is a very different, and much deeper, kind of dependency than the kind suggested by Korsgaard when she asserts that we depend on the world to take up our willings *as we mean them*. This deeper kind of dependency is more robust than mere facilitation or non-intervention with respect to what I already mean; and it requires grasping, as part of our normal self-understanding, that we depend on others to understand what we ourselves are actually doing, and that in sufficiently complex situations, we cannot grant our own initial self-understanding or avowed intentions canonical status.¹⁰

Let us return to the developed scenario: in the context of a serious argument, I say "you're wonderful," and you take offense. Has the world intervened in my action and threatened my sense of agency? Whether it has depends on what my words mean. But if the meanings of our actions depend, not just on what we mean by them but by the way the world takes them up, then what we need to ask is: what happens next? Imagine that you say that my complimentary language is belittling, or that I am being manipulative, just trying to achieve some effect in your mind. Maybe you say: "you can't just say 'you're wonderful' and expect everything to be fine." If the meaning of what I say depends on the world, then given your response, how should we understand the meaning of what I said? How am I supposed to understand myself?

¹⁰ On this point, see also Raymond Geuss: "Just as other people can describe my action in a variety of ways, so, too, I can describe my action in a number of ways. No one description, not even one I honestly give of my action myself, can claim absolute canonical status. So it is a *political* struggle how the particular description of some relevant action is produced, who produces it for what purpose, and what its exact content is" (2009, 37). Geuss takes this to suggest an inherent indeterminacy in the very nature of human action.

In arguing that we are as responsible for the unanticipated and unknown implications of our utterances as much as for their explicit and intended claims, Cavell notes that such implicit meanings and understandings could never be eradicated from what we say, “otherwise the only threat to communication would be acoustical” (2002, 12). In a way, Korsgaard’s brief presentation of the dialogue involves something like a merely acoustical threat: the meaning of what I said was all there as intended, but due to intervening circumstances, it didn’t travel sufficiently far so as to achieve its effect in your mind. Yet as I have developed the scenario, while it remains true that “my words are taken up in a sense different than how I meant them” (2009, 86), this is not because something—acoustical or otherwise—has thrown their meaning off course. Instead, it has turned out that the content of what I said has multiple meanings and implications, not all of which I intended and not all of which I can discount as wholly accidental or due to sheer ignorance, as though I just discovered an idiosyncrasy in our language (“I had no idea a compliment could be taken that way!”). Some of the different ways my words can mean will be local to my character or to our relationship; some of the meanings will be broader, saturated with power and significance from the social and political world: if, for example, there is a differential in our social standing, then paying you a compliment in the middle of an argument will mean something different depending on where each of us stand.

Once you tell me that I’m being manipulative, while I may or may not accept your take, I—we—cannot go on as before. If we keep talking, then we won’t just be moving the conversation forward, but will be also be striving for a “fuller articulation of the positions in question” (1999, 303). It might seem, following Korsgaard, that my action has not been “efficacious” and thus my agency as been “threatened,” but only if we stick to the one-off, ping-pong exchange she describes. On a wider view of the dialogue, it might turn out that I do end up effectively communicating my

feelings to you; it might even turn out that your initial resistance allows me to communicate even better than I could have had you just accepted my initial “compliment.” Or it might turn out that, for reasons unknowable to me, you cannot or will not hear my words as I sincerely meant them. Korsgaard is right that sometimes acting *is* like this, but it still does not seem that this should threaten agency as such.

The point of extending and contextualizing Korsgaard’s dialogue in any direction is of course not because it needs to be understood in any particular way. It is rather to show that extension and contextualization undermine the “bruteness” of the intervening forces model and show that Korsgaard’s narrow scope functions to obscure rather than illustrate our kind of dependency. On her model, I intend my words to have a certain effect in your mind, and when such an effect is not realized, it is as though my words hit a wall and my agency simply fails. But when I see that my words have affected you in ways I did not intend or anticipate, we need not conceive of this in terms of brute failure or inefficacy. While it is true that things have not gone as I initially intended, the idea that this should be the sole criterion for success suggests a picture of agency as surprisingly inflexible and unresponsive.

On the more diachronic and dialogical approach, that the world and other people respond to what I do in unanticipatable ways is not yet a reason to throw one’s hands up. Instead, it seems reasonable to conclude that one’s agency is limited and dependent, liable to failure but not thereby lost. In fact, that our action is positively liable to forms of failure is why Cavell is so attentive to what he calls *elaboratives*, the excuses, justifications, and explanations to which we have recourse when our actions come to grief. Elaboratives make action tolerable *given* how wrong it can go and how often it can go wrong (Cavell 1999, 296; 325). Without elaboratives, our situation might indeed more closely resemble Korsgaard’s, with single efforts to act terminating in utter failure, threatening

our sense of agency. I think the amount of space a philosopher spends on the theme of excuses—as well as apology and forgiveness—is itself indicative of their basic conception of action. Cavell’s interest in this theme signals that, for him, human action will not be understood if modeled only on the initiating decision; it will not be understood unless it is seen as essentially involving ongoing and collaborative elaboration.

Writing on Cavell, Stephen Mulhall proposes that “a reconceptualization of human action under the aspect of passivity is essential to overcoming a skeptical understanding of human relationships with others and with the world” (1999, 293). If Korsgaard’s deliberativist account, with its emphasis on the initiating moment of decision, points towards a skeptical understanding of self and world (we can only cross our fingers and hope), I have been trying to show that an account that figures action as interaction—which is to say both temporally extended and essentially responsive, which is to say passive—reveals dependency and failure to be not only ineradicable but essential for agency. Indeed I would suggest that this very condition of limitation and dependency makes possible a form of learning and creativity in action and interaction, in a way that Korsgaard’s picture of efficacy does not really allow. This is the creativity involved, not in initiating some new action, but in responding to your action’s becoming different than was anticipated or intended or even desired, to the world taking your deeds and words other than how you meant them.

§6. Dependent and All Unknown

As we have seen, for Cavell, it is essential to conceive of agency as diachronic and dialogic not only because “our responsibilities, the extensions of our cares and commitments, and the implications of our conduct, are not obvious”—Korsgaard could agree with this much—but because we are all unknown to ourselves, “because the self is not obvious to the self” (1999, 312). Why does he insist

on opacity when trying to think about human action? What is the connection between our being deeply dependent and our being all unknown to ourselves?

In a sense, Cavell's insistence on self-opacity is not controversial. Given the indeterminacy of human action and interaction, given that our actions take place in a space of meaning complex enough to support serious disagreement about the meaning of what anyone says or does, given that it is not easy to figure out what we are really doing and that we can work to keep ourselves in the dark: given all this, there is no way that any person will be wholly lucid about what she is doing and why. This is a standing fact. Cavell's further proposal is that if you have a live sense of your own vulnerability, dependence, and opacity—if an acknowledgment of dependency transforms your sense of yourself as an agent—then your sureness of self will be checked. This is how our “sense of ourselves” as agents can be productively challenged and transformed by our dependency and vulnerability, by building into it an awareness of opacity and sense of fallibility about the self.

Much philosophical debate about agency and dependency concerns questions of efficacy and control, with some maintaining that we aim “to be able to encompass ourselves completely” and be “the absolute source of what we do” (Nagel), while others insist that our “responsibilities outrun control” and that we would not wish it otherwise (Walker). Cavell helps us see that our dependency concerns not only the reach and strength of our control, but has an epistemic dimension as well.¹¹ The issue is not just that we cannot wholly control what we do, but that we cannot wholly know it. Hence the under-appreciated “epistemological problems involved in our knowing what we, or someone else, is doing” (1999, 293).

Recall that on Korsgaard's picture we are lucid about our intentions and meanings, and then faced with a boundary beyond which everything is wildly unknown and out of control. There is no

¹¹ See also for example Arendt (1998), Geuss (2009), Pippin (2008, 2011).

epistemological or descriptive challenge for comprehending either what I am doing or the world's responses; the meaning is given, there is just success or failure. On this picture, our relation to the world and others would have to be managerial, strategic: "I know what I am about, I just have to figure out how to make the world take my actions up as I mean them." In this way, Korsgaard's scenes of interaction are not only overly fearful about dependency, but uncreative, where the unpredictable or negative responsiveness of the world is either threatening or disappointing but never an opportunity. When you take offense to my compliment, this apparently *ends* the exchange: there seems to be nothing to learn.

But if, following Cavell, we understand dependency to name not just the limit of our control over the world (including our own bodies) but the limitedness of self-knowledge, then acting is not just an opportunity for successfully or unsuccessfully executing one's intention but is an opportunity for learning, where such an opportunity is predicated upon a certain degree of failure.

This can sound obstinately redemptive, as though all our failures can be transformed into learning opportunities. Whatever version of this sentiment I am advocating, it is not that there is never any real loss or failure, or that any loss can be turned into something the agent can endorse ("I know it *looks* like a set back, but really, I've learned so much about myself and now I know going forward how to do things differently..." or "when one door closes, another door opens," and so on). My point is only that failure and passivity need not be pictured as either mere interruptions of agency from the outside or lamentable facts of human life, but can also be seen as sites for forms of self-understanding and change that would not be otherwise available (I discuss this at length in the next chapter). Whereas Korsgaard thinks that in the face of dependency we must hope that the world takes up our willings *as* we intend them, I am suggesting that the world's *not* always taking up

our willings *as* we intend them is no less crucial or important for agency. Things going otherwise than strictly intended would seem to be essential for creativity of any kind.¹²

Again I don't want to be naïve about dependency and vulnerability, or redeem it in all of its forms. The suggestion is not that dependency is always fecund but that, in a philosophical climate that sees *only* threat and danger, it is important to remember that it can be. I would also add that in cases when dependency is *not* fecund—when, to recall Korsgaard's example, you say “kiss me” and then get punched in the stomach—then surely the conclusion is not that since we cannot immunize our actions against the world, we must cross our fingers and hope. I think we should instead see such cases as indicating the limits of trying to understand human action exclusively on the model of the individual. Indeed, the entire question of efficacy and dependency already indicates that any effort to understand agency in individualistic terms will be essentially incomplete.

But if dependency and failure have been integrated into our picture of agency rather than excised from it, how can we distinguish between success and failure? Korsgaard's examples are meant to illustrate defeats of agency, forces that undermine agency from the outside. So on my more thoroughly dependent picture, when is agency defeated? When am I properly responsible for what I do and when am I excused?¹³

The question of responsibility is too large an issue to decide here. What I have tried to do in this chapter is clarify the kind of agents we are, in a way that could be seen as setting the stage for questions of moral responsibility. But let me point to something Korsgaard says elsewhere (1996b) about responsibility, in which I am agreement.¹⁴ She argues that attributions of responsibility are not

¹² Gould (2003) sees this as the reason why Cavell is so interested in improvisation and chance.

¹³ Thanks to Dan Brudney for raising this issue to me.

¹⁴ I don't think her account of responsibility in *Self-Constitution* is satisfying at all. She argues that what we are properly responsible for, when we are responsible for “bad actions,” is failing to self-constitute: “what we are going to blame you for is not that other force that was working in or on you, but for the fact that you let it do that, that you failed to pick up the reins and take control of your own movements. And the reason we are

theoretical judgments we make about persons when certain observable factual conditions hold, but are practical stances we adopt, forms of relation with the person in question. When we hold someone responsible, we are asking the person to give an account of herself, where this means that we stand in the kind of relationship with her that gives us standing to ask. Responsibility is not something we read off the world, but is a way of organizing our lives together.

While Korsgaard does not specify the norms we would need for these practices of responsibility, she insists that holding someone responsible, including ourselves, requires “generosity of interpretation” (1996b, 211). That issues of responsibility involve us in the difficulties and ethics of interpretation, in the “epistemological problems involved in our knowing what we, or someone else, is doing,” means that we need to adopt exactly the kind of protracted, elaborated view that I have been recommending. It means appreciating that human action does not come “with a visible tag attached, an explicit statement which clearly specified the nature of the action in question and the reason the agent had for performing it” (Geuss 2010, 36), and that “the process of interpretation by which a description is assigned to a particular action is by no means a trivial or unproblematic one” (ibid. 35). So while I think Korsgaard is wrong to conceive of dependency as she does, I am sympathetic with her approach to how we should approach questions of moral responsibility. How then should we understand the requirement of interpretative generosity? Beyond the obvious need for humane moral imagination, I want to propose three components.

Recall that on the more radical picture of how the world determines the meanings of what we do, we are not only dependent on the world to take up our meanings as we intended them but

going to do that is that making yourself into an agent, giving yourself an identity, becoming a person, is your *job*” (2009, 175). I think this is too revisionary to our actual practices of blaming, where it does not seem that the target of blame is a failure to self-constitute, or that you allowed some “force” to work in or on you. Moreover, this account of responsibility and blame, wherein I blame you for failing to do “your job” and “become a person” is even *more* moralizing than more straightforward accounts wherein I blame you for your action. That Korsgaard’s account of agent-versus-inner-forces points towards such an account of responsibility seems to be a kind of *reductio ad absurdum*.

depend on the world and others to understand what we ourselves meant. This suggests, first, that interpretive generosity calls for a level of *epistemic humility* about oneself. This involves an openness to learning, for example, that we have been unperceptive, betraying, rejecting all unknown to ourselves, and without even seeing how (this is not the same as learning that our actions were “taken up” in those ways, despite what we intended or meant). And when directed towards another person, epistemic humility requires an openness to learning that how you “took up” their action might be skewed in ways of which you were not aware. This means that we require not only humility, but a stance of *non-defensiveness*. This is not because we should never defend ourselves, but because any competent defense against a competent accusation or protest depends on being able to consider the charge seriously, whereas being defensive means dismissing the charge out of hand, refusing the very idea that one might have been doing something other than one avowed, that there is anything to discuss. Finally, stressing the importance of humility and non-defensiveness suggests that interpretative generosity will often involve tolerating *not knowing*. This is not the same as resting content with ignorance, as though one’s efforts to understand or interpret simply hit a wall. Rather, allowing for not knowing means allowing that when it comes to sufficiently complex, ethically significant human actions in an unpredictable world, what someone did and so what that person is responsible for can be indeterminate, admitting of various competing interpretations where these can serve competing purposes.

Given the depth of our dependency and the indeterminacy of any complex human action, I thus think Korsgaard is right when she notes that “respect for someone’s humanity is not always best expressed by holding him responsible for each and every action” (1996, 221). Taking dependency seriously would, I think, require critically reconsidering various aspects of our responsibility practices. While I am not convinced that we should give up the participant stance

entirely, I do think the “objective attitude” can be more ethically productive than is often appreciated. Against those who worry that the objective attitude is cold, sterile, even inhuman, Tamler Sommers, for instance, shows that taking up a such an attitude would allow us to “rid ourselves of high-toned self-righteousness” (2007, 336), where “rather than relentlessly judging the actions of [our] friends and acquaintances, [we] would try to appreciate them in all their complexity” (2007, 329). This stance of appreciation rather than judgment would be no doubt difficult to cultivate, but I do think that taking the depth of our dependency seriously means that we must start thinking about how to do that.

*

In this chapter I have tried to show that any satisfying account of human agency must integrate our essential dependency into agency itself. While Korsgaard is right to consider the ways in which our dependence on the world and others and our own bodies renders us vulnerable, I argued that she is wrong to draw a firm line between the will and everything else, since this prepares a picture of the world as *essentially* threatening, that is, threatening just in virtue of being external to the will, i.e. just in virtue of its being a world. This, I argued, sets one on the road to skepticism about agency, and I suspect this is why Korsgaard affirms not only that we must, can only, cross our fingers and hope for the best, but that she does not know how to solve the problem of dependency. *That* these are her conclusions indicates that we need a different account of human agency, where overcoming skepticism about our relationship with the world and others requires not more power and activity, but a more integrated conception of dependency and passivity. Part of what this involves, I proposed, is a different scene or exemplar of agency: rather than focus on the decisive moment of initiation, we should look at diachronic and dialogic scenes of interaction. This will allow us to see the ways in which passivity and dependency are best conceived, not as intervening forces, but simply

as dimensions of human agency in the most basic sense. Whereas Korsgaard sees dependency as getting in the way of real human action, I have tried to show that the ongoing negotiation of dependency provides us with the best picture of human action.¹⁵

It is surely painful to fail in our efforts or meet resistance in other people or be corrected in our own take on what we are doing. But this is different from the idea that resistance undermines our agency and sense of ourselves. Such an interpretation of resistance misses the fact that friction gives *reality* to our agency and ties it to the world. We gain this kind of reality by being resisted and corrected, in being shown that we are not just as we take ourselves to be, and hence is a reality we gain *only* insofar as we are dependent on others and “all unknown” to ourselves.

Of course, we also need to experience ourselves as real through the success of our actions, by their being taken up by the world as we mean them, and by being seen in a way that tracks our own self-understanding. But it would be wrong to conclude that experiences of failure and the recalcitrance of the world serve *only* to flout or undermine our sense of ourselves. They can also tether us to the world. They are basic ways in which our essentially dependent, vulnerable, and opaque agency is realized, that is, real.

¹⁵ The reference here is Wittgenstein’s claim that “the human body is the best picture of the human soul” (PI p. 178), against the skeptic who claims that the body blocks our access to the human mind, rendering us unable ever to know whether there are other minds.

Conclusion to Part Two

In this section I have argued that we need to appreciate that self-opacity is an ineradicable dimension of the human condition insofar as we are thoroughly dependent rational animals. That we are non-self-sufficient is not a mere fact but is registered from the practical perspective in the form of self-opacity, as the inability to encompass oneself from the deliberative perspective. In Chapter Three I argued that Freud's concepts of drives and sexuality explain how we are bound to and oriented in the space of reasons, and yet for all that, the drives also constitute a limit to what we can rationally avow. I also argued that the creativity and recalcitrance of drives means that our occupancy of the space of reasons and culture is essentially conflictual.

In Chapter Four I argued that we cannot think of our vulnerability and dependency on the world and others as kind of external limit to our agency, but must think of our agency as constituted by such dependency. Korsgaard worries that such dependency threatens to undermine "our sense of ourselves" as agents, and while I rejected the idea that we are undermined, I tried to show how our practical sense of ourselves is transformed by taking such dependency seriously. I also argued that dependency does not just concern the reach of our control, but involves, as Cavell puts it, "epistemological problems involved in our knowing what we, or someone else, is doing" (1999, 293). That we are dependent in this way means that we will need others to know ourselves.

In the next and final chapter, I explore some of the ways our self-opacity can contribute productively—if disruptively—to our lives.

Part III

*

Chapter Five: The Difficulty and Value of Self-Opacity

In this last chapter I want to ask simply whether it is possible to see our self-opacity as capable of playing a positive role in our ethical life. This would mean thinking differently not only about the value of self-knowledge as avowal and endorsement, but also about what it means for something to make an ethically important contribution to our lives. So much contemporary thinking about morality and agency—with its emphasis on reflective, deliberative, decisive agency—is tilted against this very idea. As we’ve seen, self-opacity is regarded for the most part as an occasional lapse of ordinary self-knowledge or as evidence of ambivalence or some failure of resolve, where the more or less explicit normative atmosphere surrounding these observations is that one ought to *do* something about this as quickly as possible; you ought to, as Korsgaard puts it, “pull yourself together.”

In the last two chapters I have argued that our self-opacity is connected to our non-self-sufficiency, to our animality and to our dependency on the social world (where these are, as we saw, intertwined and mutually informing). Because of this, self-opacity cannot be conceived as just one more empirical fact about us, like being bipedal,¹ but as an aspect of our practical lives to be navigated from within the perspective of agency. It can be tempting to picture our being “vulnerable” or “finite” on the model of an outer limit, as though our agency meets some kind of

¹ Though even bipedalism may resist wholly third-personal, theoretical characterization. It is possible that any aspect our morphology could call for an internal analysis, such that there is almost nothing in human life that is adequately understood solely on the model of a mere fact. Fichte, for instance, tied uprightness and bipedalism to self-consciousness and freedom: “through its daring upright gait...the species, in maintaining its balance, also maintains freedom and reason in constant practice; it remains perpetually in a state of becoming” (*Foundations of Natural Right*, 77). Regardless of whether Fichte’s argument is successful, my point is simply that it is a philosophical decision what of human life calls for an internal analysis and what is adequately understood externally.

threshold, as though we were wholly capacious *up to a point*. Part of what I have tried to bring out is that our finitude, vulnerability, and animality pervade and determine our human kind of agency, not as external limits but as internal, formative conditions. And I have tried to show that for self-conscious beings like us, these dimensions inflect and determine not only what our agency is like, but what it is like for us, how it informs our practical sense of ourselves. Self-opacity is in this sense the reflexive dimension of our finitude. What I want to suggest in this final chapter is how our self-opacity and our being only ever incompletely self-encompassed might be essential to a good human life.

Self-opacity is a strange idea. It is not something one can get clearly into view, let alone something for which one can straightforwardly advocate as valuable or good. This is not only because standard philosophical frameworks are structured so as to obscure this possibility, at descriptive, normative, and methodological levels, and not only because it does so often take the form of a problem or difficulty. The conceptual difficulty is that self-opacity is a dimension of the very self-knowledge that the concept is supposed to rule out. Self-opacity is not a state of brute ignorance about some clearly demarcated domain of facts, but is the opacity of a self that also knows itself. We are self-conscious beings, straightforwardly and non-observationally aware of much of what we do and why. That we can be, as Cavell says, “all unknown to ourselves” is only possible *because* we are self-knowing. Emphasizing self-opacity must not involve denying this. And in fact, it is only and precisely because we are self-conscious, self-knowing creatures that our failures to know and understand ourselves can be so unsettling, both philosophically and practically, in both positive and negative ways. Given all of this, it makes sense that the importance and value of self-opacity has been missed, hardly been seen as something worth considering.

My aim in this chapter is to try to make room for the possibility of appreciating the potentially positive roles self-opacity plays in our lives as finite, dependent, rational animals. I will proceed by exploring some of the ways it shows up, again not as a state of brute ignorance (of which we are wholly unaware) but as disrupting our ordinary self-understandings and ways of going on, as contributing to the “productive unrest” of the critical individual (as discussed in Chapter Three). Some of these phenomena may not immediately appear to have straightforward ethical significance, largely because they are not immediately “practicable” in ways that philosophers sometimes expect moral concepts and experiences to be: there is not always something obvious that we should *do* with or about them, and they are not experiences one can straightforwardly pursue. My hope is that by moving through this territory, I will be able to generate an appreciation for how self-opacity inflects our lives in various ways, and why and how this matters to us.

What is common to these cases is that they present encounters and experiences that cannot be straightforwardly comprehended from the ordinary first-personal deliberative perspective. Central to their ethical significance is that they disrupt a person’s sense (or fantasy) of self-sufficiency or self-certainty. These experiences are registrations of our animality and our dependency that do not take the form of reflection or endorsement but rather actively resist those exercises of first person authority. Yet while Korsgaard could only regard a challenge to our sense of ourselves as agents as a threat, one that must be managed and recuperated by endorsing it, making it mine again, this would be a way of avoiding their ethical import—it would be a way of “deflecting” (a term I will discuss at length). The reason that efforts to endorse or authorize are inappropriate in these cases is that, as I will argue, the *form* and *content* of these experiences are internally related: *what* we learn (that we are not self-sufficient) is inextricable from *how* we learn (through unwilling disruption).

Before proceeding I want to note the difference between these kind of potentially productive disruptive encounters and trauma, a difference that is essential to draw if we want to recognize the value of the former without thereby attributing value to the latter. In “A Late Adventure of the Feelings” (2009) Gregg Horowitz argues—by way of Freud and Klein—that while certain forms of loss are essential to human development, we must be careful to distinguish healthy, productive, manageably painful loss from pathological, destructive, unmanageable trauma. Psychoanalysis helps us see that ordinary losses as well as the pain that comes with them can be occasions for creative response, which involves weaving the loss into the texture of livable experience (though not by bringing it into “reflective view” so as to endorse or “consciously incorporate” the experience). Trauma, by contrast, is by definition not ordinary (regardless of its statistical prevalence), and—again by definition—is not an occasion for healthy creative response, plays no productive role within the weave of experience. This is not to say that victims of trauma cannot respond creatively; it is only to say that they do so *in spite of* and not thanks to what they suffer. The kind of experience I am interested in are thus more like an ordinary, manageably painful losses. Making this distinction allows us to appreciate the value of difficulties of reality without requiring that we tell any kind of redemptive story about trauma.

I want to show that making sense of the significance of these experiences depends upon the idea of self-opacity, and moreover, that the very opacity that blinkers and unsettles us in these experiences is essential to their value: this is to say that greater self-knowledge and self-possession, in these cases, would undermine their value and significance.

§1. Missing the Significance of Self-Opacity

It will be helpful to begin by again making explicit the philosophical picture of agency that makes an appreciation of self-opacity so difficult. We saw in Chapter One that philosophers advocating for what I called the Deliberative Endorsement conception of agency and moral psychology construe self-opacity as a failure of agency conceived as authorial authority. This picture of agency privileges deliberation and decisiveness for two main reasons: first, reflective self-consciousness is seen to ground our freedom and authority (Moran 2001, 142), and second, deliberative concerns are taken to exhaust the realm of the practical; as Wallace puts it, “the practical point of view [is] defined by our interest in the deliberative question of what we ought to do” (2006, 153). Thus agency is defined by our capacity to reflectively deliberate about what to think or do and to execute to our decisions.

Moreover, these capacities are regarded not just as essential for agency (again construed narrowly) but for psychic flourishing and wellbeing, not just as what we need to act but as what we most deeply want. So, for instance, Moran maintains that self-knowledge in the form of avowal, endorsement, and commitment is importantly expressive of psychic health, and that a failure to know oneself in these ways constitutes a form of psychic impoverishment and incapacity (see Moran 2001, 8; 35; 60; 107). Thanks to this conceptual framework, self-opacity can really only be understood as a form of agential failure and psychic malady. Being clear and committed about who one is, what one believes, and what one wants—and why—exemplify of human agency and moral psychological health. This is the state we ought to strive for, and ought to want.

I wish to challenge neither the idea that in many cases practical self-understanding—acting for reasons one understands and endorses—may be the thing to strive for, nor the idea that being opaque about oneself can constitute a failure or malady. But notice that championing self-knowledge as avowal as necessarily good, as always the thing to aim for, involves accepting 1) that

our agency is primarily and best expressed in self-conscious resolution; 2) that such expression is necessarily conducive to flourishing; and 3) that this value is not context dependent. In tying self-knowledge as avowal and endorsement to rational agency, freedom, and psychic health, this picture comes with a generalized *normative* recommendation that one ought always strive to achieve this kind self-knowledge and self-possession. All of this indicates that to so much as see the possibility that self-opacity might make a positive contribution to ethical life, one will have to resist this no doubt compelling idea that the good human life is best exemplified in self-conscious, deliberatively-arrived-at, resolute decisions, behind which the agent can stand with her whole, undivided heart.²

These pictures of agency are oriented in such a way as to frame self-opacity as definitionally pathological (an “infirmity” as Moran puts it), as a deviation from good agency and psychic health. This philosophical orientation operates at the level of the explicit values championed: rationality, reflective consciousness, deliberation, control, decisiveness. It operates at a more implicit level in terms of the background framework that reduces the broadly practical to the narrowly deliberative, where any non-deliberative self-relation is *just thereby* a theoretical relation, leaving little room to recognize modes of practical engagement that are not immediately directed towards choices to act. And this orientation often operates at the level of style and rhetoric, in the confident voice used to articulate intuitions about agency and ethics that “we” all hold or ideals we ought to strive for, as when Korsgaard issues the apparently general exhortation to “*make up your mind*, or even better *pull yourself together*” (2009, 126). This kind of picture can be captivating; it can be hard to get outside (Wittgenstein 2001, §115).

² Frankfurt’s idea of volitional necessities might seem to provide a counter example to the ideal of deliberative decision making, since the whole point and value of them is that they are precisely *not* chosen, not up to me. But Frankfurt insists that such volitional necessities must be wanted and identified with. So even though they lack the rational deliberative aspect emphasized by Korsgaard and Moran, volitional necessities involve wholehearted endorsement. Indeed, they can be understood as one of the ways we “make the contingent necessary” which for Korsgaard is the mark of the good human being.

The question is: what does such a picture exclude? What parts of our practical, ethical life does this deliberative picture obscure, or (explicitly or implicitly) characterize as defective or peripheral or as philosophically insignificant? Might there be important aspects of our lives that are opaque and obscure, that resist avowal and endorsement, that are not directly concerned with determining what to do, and yet are for all that both practically and ethically significant?

The remainder of this chapter involves a series of efforts to make concrete the idea that self-opacity plays an important, productive role in ethical life. I begin by analyzing a scenario described by Barbara Herman in order to make clear why the Deliberative Endorsement model cannot do justice to the nature or value her experience and why we need the concept of self-opacity in order to appreciate it. Next I explore Cora Diamond's concept of "difficulties of reality," experiences that involve "the mind's not being able to encompass something which it encounters" (2003, 2). I work to bring out their essentially reflexive and self-implicating structure, such that the unthinkability of reality necessarily involves a certain incomprehensibility of the self; and I clarify why unthinkability and incomprehensibility are ethically important modes of experience. In the final section I make perspicuous the structure common to these kinds of encounter: they involve an unwilled and unexpected encounter with something we cannot comprehend, where the significance of *what* we learn (that we are not self-sufficient) is inextricable from *how* we learn (through unwilled disruption).

§2. Stumbling Over One's Orientation

Let us begin by exploring a scenario that would initially seem to lend itself well to the Deliberative Endorsement kind of analysis. I will use a short scene from Barbara Herman. She writes:

When, for example, I hear my mother's parental anxieties in my own voice as I criticize my son, I cannot resolve the problem I discover just by abandoning some end or disowning some desire, however much distress I feel about what I am doing. Part of what I discover in these moments is who I am—or who I am as a parent. I listen and find out what I desire.

But then it is not enough to say that I do not want to act on these desires, and also not enough to say that I do not want to have them. I want them not to have a place in the complex of desires and thoughts that constitute myself as a parent. This may be no easy thing, for the very desires I would disavow may hold together things I like about myself as a parent. 1993, 195.

How should we understand Herman's anxieties and her relationship to them? As I read this scene, when Herman hears her mother's anxieties in her own voice, what she hears is a repetition, in two senses: first, she hears herself repeat her mother's worries, and second, she hears something that she has herself been repeating in the form of a habitual response or orientation (a concept I will discuss in a moment). This does not seem to be the first time that Herman has felt or expressed these fears, though it does seem as though this is the first time she has really "heard" them as such: this is a new encounter with something old.³

With this we have already ruled out categorizing the anxiety as a trivial discrepancy or circumstantial outburst that have no bearing on Herman's sense of self or her sense of her relationship with her son. Herman is not describing a case of snapping at her child after a bad day at work. Whatever these anxieties are, they touch upon something central to who Herman is: while her mother's anxieties had been orienting and framing her relationship with the world, and her son in particular, she is only now becoming aware of this. But if the problem is not a mere outburst, what is the problem?

Herman's anxiety might initially seem to be a case of what Korsgaard and Frankfurt call an "alien" force or desire. Herman is unable and unwilling to "unequivocally endorse or support" (Frankfurt 2005, 163) or identify with these anxieties or the desires they express, and so they may

³ It is important to notice the kind of complexity involved in hearing a repetition in a single expression or instance. A repetition, after all, is an element in an extended pattern, and is thus always secondary to something that came before. But in Herman's case, she hadn't been noticing the anxiety animating her criticisms, which is to say she didn't hear it the first time around. So to suddenly hear herself repeating (herself; her mother) is not only a complex but a properly uncanny kind of experience, a coming-to-consciousness of a pattern one had been in the midst of carrying out without realizing it.

seem to that extent “external” or as needing to be made so through Frankfurtian acts of self-organization. But Herman is careful to note that what she discovers when she hears this voice is something about herself, about *who she is*, as she puts it. She does not relate to this experience the way the unwilling addict relates to her craving and it is not “something merely working on her” (Korsgaard 2011, 76). In his critique of deliberative endorsement conceptions of agency, Bernard Williams suggests it is absurd to think that “any ethically significant part of me that belongs to me simply as a result of the process by which I was contingently formed” is thereby *not me*, “as though I had been brainwashed” (2008, 158). In this same spirit, Herman importantly does not describe her experience of her own anxiety as *not her*, or as an experience of powerlessness in the face of some wholly external attitude that is breaking out from within. In proposing that, whatever else she does, part of her response must involve “listening” and “finding out what she desires,” Herman is suggesting that these anxieties, no matter how strange, *do* express some desire of *hers*, rather than a desire she happens to have as a result of mere empirical fact.

Perhaps Moran would point out here that insofar as Herman cannot avow her anxiety, does not see reasons that speak for it, she thereby takes up a *theoretical* stance with respect to this attitude. Taking a theoretical stance “reduces the emotion to an interior occurrence to which she is passively subject” (2001, 58), as though Herman were simply a witness to her own inner life. Discovering that her mother’s anxiety animates her criticisms of her son would then be analogous to discovering that, for example, she has a heart condition, or to bring it closer to Herman’s scenario, that she has the same heart condition as her mother. On Moran’s conception of a theoretical relationship, Herman’s relation to her own unavowable anxiety would be structurally akin to the kind of observational epistemic relation that anyone else could take up. She would know her anxiety “third personally,”

where “this information may as well be about some other person, or about voices in her head” (Moran 2001, 93).

Yet while it is true that Herman’s problem involves an attitude that she cannot endorse, this only shows that deliberation and resolution may not be appropriate here: deliberation cannot lead Herman to endorse her anxiety as the appropriate feeling to have, and she also does not spontaneously change her mind in the absence of rational support. But notice that we are steered towards the idea that she *therefore* occupies a “fully *theoretical*” stance *only* if we accept Moran’s conditional: *if* one’s stance is not deliberative, it is thereby theoretical or third personal. But if our options involve either endorsing one’s attitude (as reasonable) or disavowing it (as unreasonable), *or* reporting on it impersonally it as if it were a bit of evidence, we should ask whether these are really our only possibilities. It seems that our philosophical picture is prejudicing us to see *less* than is actually here. Herman’s point in fact seems to be that because what I discover “in these moments” is “who I am,” taking either a deliberative *or* a theoretical stance will be inappropriate.

Perhaps Moran has secured another option: sometimes we have to “manage” a recalcitrant attitude with a view to eventually achieving the right kind of internal, deliberative relationship.⁴ So while Herman may not be able to exercise her properly first-personal authority with respect to it now, the goal should be to move the attitude out of the theoretical and into the deliberative realm, such that she can endorse or reject the attitude based on the reasons that support it (or do not). “Managing” the attitude thus constitutes a relationship of supervisory control that that is external and manipulative rather than internal and rational, and it is, to that extent, a defective form of agency on Moran’s account. But if some attitude cannot be avowed and is not responsive to

⁴ See McGeer (2007) for a critical discussion of the role of such self-management in Moran.

reasons, we can take up a managerial stance as the next best thing, engaging in a kind of transitional relationship that can move us from theoretical to properly deliberative self-knowledge.

There is, however a more basic problem with the idea of management as something that can move us from the theoretical to the deliberative. For if Herman really did occupy a fully theoretical stance and see her anxiety as a kind of given fact, then it is not clear how could her attitude could present as the kind of mental item that could be even *potentially* avowable.⁵

Imagine that I take medicine that increases my heart rate and makes me anxious. If I know that the anxiety I feel is due to the medicine, then it is not an even potentially avowable attitude, because this anxiety is not the kind of thing that can be either responsive to reasons or expressive of who I am. I might find it difficult to remember that this anxiety is merely physical and thus be tempted (erroneously) to interpret the world and myself in light of it; for example, it might be easy to think that my social interactions are strained and anxious, and hard to remember that it is “just the medication.” But insofar my anxiety is an empirical fact, knowable theoretically, then it is not even a candidate for avowal. I can manage it by, for example, taking my pills only at certain times, and I can talk myself down when on the brink of interpreting my day through the uncomfortable feeling (“remember, it’s just the pill”), but I cannot do any of this with a view to avowing or disavowing it.

Herman’s anxiety is very different from this. True, she sees some of its roots in her relationship with her mother, so there is a kind of causal or etiological story to be told here. But again Herman is explicit that what she discovers in these moments is *who she is*. Whereas my

⁵ If I understand him correctly, Gardner makes a similar point about Moran, writing: “if, as realism implies, the objects of self-knowledge have a real existence that extends beyond their being avowed or identified with, how do they come to present themselves as practically amenable, as occasions for deliberation, as surrounded with practical possibilities? How, once I have grasped something as a fact about me, can I go on to grasp it as ‘practicable?’” (2004, 262).

chemically induced anxiety does not and cannot tell me about who I am,⁶ Herman's anxiety does. And this is because it doesn't so much speak *to* her, as would a "voice in her head;" rather she hears *herself* expressed *in it*.⁷ Perhaps she doesn't fully understand herself here, perhaps her anxiety—and who she is, if that is what it discloses—remains opaque. But it is important to see that these relationships are not well described as "theoretical," i.e., as a distanced, disengaged position on an extant object that, in principle, anyone else could take. No doubt Herman can neither avow nor wholly disavow her anxiety, no doubt she registers a disconnect between what her attitude is responsive to (her son) and the attitude itself, no doubt she cannot quite see its rationale, but none of this entails that she is relating to her anxiety "theoretically," as if it were a brute fact that she can observe like anyone else.

The philosophical picture that pushes us to describe this relationship as theoretical is one that reduces the practical to the deliberative and then pictures every other relationship as thereby theoretical.⁸ But this makes it seem as though we are as persons either freely and rationally endorsing or rejecting our actions and attitudes *or*, if not, we are gazing at ourselves as spectators, as if all non-avowable attitudes were structurally akin to facts or alien forces (and note that even these

⁶ I am not certain of this way of distinguishing medically induced anxiety from psychological or even existential anxiety. Presumably how I react to even medically induced anxiety will tell me something about who I am (i.e., I am someone who can or cannot handle constant low levels of anxiety). This does involve some kind of self-discovery and yet still, it seems different from Herman's, because the attitude does not *express* anything of me, rather my response to it does. Herman's response to her anxiety is also self-disclosing, but again, the crucial difference is that her anxiety itself is self-disclosing as well.

⁷ In this way I think Herman's case is also different from the kind of therapy-scenario discussed by Finkelstein (2003). In these scenarios, the analysand learns from her analyst that she resents her dead mother and even believes this about herself, but cannot *express* this resentment. To this extent, she relates to it theoretically. But in Herman's case, she does express her anxiety—she does not learn about it or accept its existence based on testimonial or behavioral evidence. Yet she still does not understand it.

⁸ I think a major source of the problem in Moran is that he seems to treat all non-deliberative self-relations as of a single kind, and uses the following terms virtually synonymously: theoretical, attributional, empirical, factual, third personal, perceptual, observational, evidential, explanatory, lacking authority. Moreover, he sometimes suggests that these forms of self-relation are also all thereby alienated and/or opaque. But this conflation of terms obscures important differences, for example, the possibility that I might observe something about myself without thereby being alienated from it.

are importantly different from each other!). And yet what Herman's scene brings out is that there are significant, central stretches of practical life that involve grappling with *practical* problems (i.e., do not involve spectatorship or evading responsibility), and yet cannot be approached by deliberating and making a decision, whether that means determining what to do, or whether to endorse or outlaw some desire. What I now want to suggest is that, in this case, the deliberative and theoretical stances are inappropriate precisely because the attitude at issue is not discrete or isolated but is an integrated part of Herman's general orientation.⁹

Herman herself notes that making a decision is precisely what cannot be done in her situation. As she writes:

Encountering (or more often *stumbling against*) such a *complex*, we may be enlightened or transformed, or moved to therapy, or despair ... We cannot just choose to care (about some things), and we cannot just prize out an unwanted desire by identifying and rejecting its object. 1993, 196-my emphasis.

What is significant about this passage is that it describes these moments in terms of “stumbling against” not this or that attitude but a *complex*. This means that Herman's anxiety is not an isolatable attitude to be endorsed or not, but is rather an interwoven dimension of a broader orientation, where such an orientation, as Cavell puts it, “affects everything, and which I may or may not be interested in discovering about myself” (1999, 360). So Herman has stumbled against something that she could not see before, cannot bring fully into view now, cannot avow, and may be resistant to facing.

⁹ Angela Smith expresses a similar sentiment when she writes: “If we view any attitude that we do not reflectively endorse as something that simply happens in our mental history—as ‘merely psychic raw material,’ as Frankfurt recently put the point—then we will not be open to the possibility that these attitudes might actually teach us important things about ourselves and about our most basic values and commitments” (2004, 347). I would only add that sometimes learning both from and about ourselves does not take the form of acquiring more contentful knowledge; sometimes morally-relevant learning involves being open to discovering that one knows less about oneself and one's basic values and commitments than one thought.

Daniel Brudney glosses an orientation as “one’s general stance towards the world” (2010, 301). An orientation is general, first, in that it has steadiness and consistency, showing up across various experiences. There is a pattern to how the world is organized and filtered. Second, an orientation is general in that it encompasses and informs various dimensions of human experience: it is epistemic, cognitive, conative, evaluative, and practical. Finally, an orientation is general in the sense that it is pervasive, which is to say: an orientation is not something one normally encounters *in* experience; it is rather the general frame of one’s experience. So an orientation is relatively consistent across diverse experiences, it concerns various dimensions of such experiences, and it provides the broad frame through which any experience is filtered.¹⁰

Importantly an orientation should not be conceived in abstraction from larger socio-historical and political patterns, orientations, and habits. The relationship between orientation and world runs both ways: one’s orientation patterns one’s engagement with the world and the world patterns one’s orientation. So for example, one will be oriented by and towards gender in various ways and these ways will be expressive of one’s individual character and upbringing, as well as various larger social and political cultures. As Judith Butler notes, “there is no ‘I’ that is not implicated in a set of conditioning moral norms, which, being norms, have a social character that exceeds purely personal or idiosyncratic meaning” (2005, 7). This is to say that while an orientation will be an orientation *for* an individual, there need be nothing individualistic or narrowly psychological about the concept of orientation. The different orienting threads are often tightly

¹⁰ This can make orientation sound a lot like character or a set of character traits. John Doris (2003) has criticized the very idea of character (especially as it appears in virtue ethics) based on empirical research that suggests that *situations* determine how a person morally acts and reacts *more* than what we might think to be a person’s character traits. Doris concludes from this research that, *contra* virtue ethicists, character traits do not predict behavior and hence the idea of character has little explanatory value. But there seems to be room to maintain that, for instance, my appreciation of “who you are” in some broad sense guides my understanding and assessments of you, even if I know that your concerns and commitments will not result in wholly predictable behavior across all situations. For a similar response, see Arpaly 2005.

woven and may be difficult (epistemically and metaphysically) to pull apart. These are the threads that bind a person to the world in ways that make the world make sense, where, crucially, these patterns of attachment and coherence are not themselves objects of reflective or deliberative consideration but instead provide the framework for any such consideration.

An orientation thus cannot be conceived as a kind of object, not only because it cannot be straightforwardly brought into view, but because it is not a static thing. An orientation is more like a kind of ongoing, habituated activity: it is learned through interacting with others, developed over time, something one unthinkingly and continuously does. Even though it is neither the result of nor itself a kind of action or choice, it is still a stance one actively *engages* and thereby falls within the realm of the practical, even while not narrowly deliberative. An orientation is also for these reasons difficult to notice *as such* because it is, as I said, the active frame of one's experiences rather than any particular action or object, it is a pattern of one's attention rather than an object of attention. This means that an orientation operates for the most part in the background, behind one's back.

If an orientation essentially functions behind one's back, out of one's view, then being oriented *means* being self-opaque in various respects and to various degrees; while I may be able to bring some dynamics of my orientation to consciousness, others will be more resistant, and others still will be properly unconscious. In this way, self-opacity makes a valuable contribution to agency and ethical life insofar as it allows one to engage meaningfully with the world and others. So how exactly is self-opacity involved in being oriented?

First, in understanding any particular experience one will be oriented towards it in some particular way, interpreting it in such a way as to make some facts salient and others irrelevant, and in such a way as to knit it into the broader understanding and experience. Understanding a series of events or stage in one's life or a relationship will involve adopting one interpretation to the exclusion

of other, competing, plausible interpretations, such that having any meaningful grasp of a situation will thus involve *not* grasping much that is present. As Marcia Baron notes,

we interpret what we see and experience, trying to make it make sense, and since there are many ways of interpreting it and usually no right way (though certainly many wrong ways), we sometimes adopt the interpretation that is the one we can best live with. 1988, 442.

Thus, in any given situation of sufficiently complex meaning and stakes, one's orientation will guide one's interpretation and this will necessarily leave much out of the picture (not in the sense that it is recognized and discarded, but in the sense that it will be barely registered at all). This not-seeing that is part of seeing does not mean that one is delusional or willfully blind or self-deceptive; it is just that having any take on or relationship to a given situation means that there will be features of world and self that will not be visible, and some will not even be available for conscious reflection.

The second way that an orientation involves self-opacity is that insofar as one's basic orientation will involve patterns of attention and interpretation that are general and pervasive, *that* they are patterned in any particular way will not be realized. Instead, an orientation's output simply seems "normal, the way things are" (Brudney 2010, 219), where this requires that one does not notice either one's ongoing and active contribution to "the way things are," or the ways in which this contribution is shaped by various normative frameworks provided by the larger world. While configurations of a person's orientation will be continually manifesting in how she engages with others and the world, and so will often be displayed *to* those others, these configurations will for the most part remain opaque to the person herself.¹¹

In "stumbling upon" a complex, Herman is, as it were, stumbling over some part of her orientation that had remained, as it must, in the dark. That the anxiety to which Herman gives voice

¹¹ This idea that one's orientation will "show continually" while being unseen by the agent herself can be found both in Murdoch, and also in Hannah Arendt, who defines action and speech as modes of "disclosure of who somebody is," where "the 'who,' which remains so clearly and unmistakably to others, remains hidden from the person himself" (1958, 178-9).

is not an isolatable attitude but part of a larger complex means that her response to it cannot involve *either* straightforward endorsement or rejection of that attitude, *or* theoretical observation of it. There are three reasons for this. First, the attitude is implicated in a larger, opaque complex and is thus difficult, epistemically, to isolate in order to pull out for reflective consideration. Second, insofar as it is so implicated it is impossible, evaluatively, to simply quarantine and rationally reject (“the very desires I would disavow may hold *together things I like* about myself as a parent”). In these ways the mind cannot be broken up into bits for reflection and evaluation; it precisely cannot come up for reflective review. Virginia Woolf gives vivid, first-personal expression to this inability in *Orlando*:

every single thing, once he tried to dislodge it from its place in his mind, he found thus cumbered with other matter like the lump of glass which, after a year at the bottom of the sea, is grown about with bones and dragon flies, and coins and the tresses of drowned woman. 1973, 101.

The third reason Herman cannot extract and rationally assess her/her mother’s anxiety is because the complex, of which her anxiety forms a *working* part, constitutes the pervasive orientation guiding her understanding of self and world. It is not just that the anxiety is “cumbered with other matter,” but that all of this is *her*. Whereas the deliberativist sees activity and passivity as neatly separable, such that one simply suffers anything that one cannot avow, an orientation involves active and passive dimensions woven tightly together. This is why one doesn’t so much reflect upon such a complex as stumble over it; she *expresses* anxiety that she recognizes as *hers* and in the same moment notices it precisely because she is only properly hearing (and recognizing) herself now.¹² Herman is *more* active than she realized insofar as she has been contributing to how she orients in the world,

¹² Lear (2011) is interested in exactly these moments of “uncanny doubling” wherein one not only expresses oneself but also “looks over” at oneself as weirdly unfamiliar in one’s familiarity. And he notes, importantly, that these experiences of self are “urgent and practical,” confronting one in the first person (59). I have not engaged with him more directly because he is concerned more particularly with the phenomenon of “unconscious,” which would take us into more specifically psychoanalytic terrain.

and also *less* active than the deliberative picture allows, insofar as this form of activity is something other than rational decisiveness.

Herman is describing the disorienting, even uncanny experience of glimpsing, not some object or other, but the frame thanks to which any object can appear. And again, while this may not be something one can endorse or avow, this form of encounter is also wholly unlike a “theoretical” report about another person or fact in the world or voices in one’s head. The essentially practical and first-personal experience Herman describes is both more strange (than a mere report) and less strange (than alien voices) than what Moran envisages.

But now an objection one might be tempted to raise at this point is the following: what Herman’s scene describes is not a *failure* of self-knowledge at all but is rather an instance of it. Does not the language of self-discovery precisely mark this as a moment of self-*knowledge*? Why think of this scene in terms of self-opacity at all?

While I have provided reasons for thinking that Moran’s conception of self-knowledge as deliberative and resolute is too narrow, I agree with him that self-knowledge must be analyzed as a practical and moral psychological, rather than strictly epistemological issue. If we broaden Moran’s conception of the practical, one of the crucial lessons we learn from him is that self-knowledge might be better understood as self-understanding, as making sense to oneself in light of how one makes sense of the world. On this construal, self-opacity marks the absence of such self-understanding: one does not make sense to oneself, something remains unclear, hard to understand, obscure (recall here the three definitions of opacity cited in the Introduction). Such self-opacity may be total—as when I do not even know that I lack such self-understanding (i.e., Herman before the stumble)—or it may be partial—as when I recognize that I do not understand my own actions or

attitudes, do not make sense to myself. What this latter possibility shows is that self-opacity can actually be compatible with certain forms of self-awareness.¹³

In Herman's case, for instance, she recognizes the anxiety she expresses as both her own and as her mother's, and so even though this involves a kind of self-awareness, it immediately implies—and implicates her in—a broader stretch of self-opacity: along with this new awareness comes an inability to practically understand why she does what she does, feels what she feels, sees her son the way she does. All of this involves some degree of self-awareness, but what Herman lacks is self-understanding, and to this extent she is self-opaque.¹⁴

Notice also that the kind of self-relation involved in Herman hearing her mother's anxiety in her own voice also involves a dilated temporal dimension (and this is what we should expect, given the connection between self-opacity, upbringing, and orientation). Part of what is revealed in this moment is that Herman had not realized, until now, that she *had* been oriented by such anxieties. In stumbling upon a component of her orientation, part of what Herman learns is how much of who she is has been and remains out of her sight, and how much of her experience of the world has been shaped by those very blind spots. The stumble intimates, as it were, how much of *her* is *not up to her*,

¹³ Moran himself acknowledges this idea, for he thinks that the incapacity to avow one's attitude is itself a form of self-opacity. And though I object to the idea that self-knowledge needs to take the form of *rational avowal*, I agree that one can be aware of one's attitude or action while it remains nonetheless opaque. David Finkelstein makes a similar same point in *Expression and the Inner* about the difference between conscious and unconscious mental states. On what he calls the "simple account" of consciousness, a mental state is conscious if you know you're in it, and unconscious if you do not know this: to be aware *of* a mental state makes it a *conscious* mental state. Against this conception, Finkelstein argues that "to have an unconscious wish or fear is not essentially a matter of being ignorant of something. It is a matter of being unable to *do* something—of being unable to express one's state of mind in a particular way" (2003, 122).

¹⁴ In another essay, Herman writes about self-opacity in terms of mysteriousness rather than total self-blindness: "Aspects of our own behavior may appear mysterious to us ... Beyond the issues of ordinary self-opacity, there are barriers to acknowledgment and insight that are set up to protect the integrity or felt decency of the self ... no one can see everything; features of our character really do blind us from some things; it may even be necessary, in some deep practical sense, that we not attend to everything" (2008, 93).

how much of her own mind she has not reflectively made up. So there is an opaque backward-looking dimension to this self-relation.¹⁵

There is also an opaque forward-looking dimension: insofar as Herman recognizes that this is not a problem that can be solved by an act of will, part of what she discovers is how her orientation will continue to inform, to some degree, her relationships going forward. If she knows she cannot just extract this tendency, both because it is not subject to the will *and* because she knows it is playing a role in holding her together, then she knows that this blind spot and others like it will continue to do their orienting work in the dark.¹⁶ Again, this need not involve taking a theoretical or spectatorial relationship to herself, let alone a stance of resignation or a repudiation of responsibility, as is suggested in Moran's Sartrean moments ("I guess that's it then, that's just who I am"). Just as rushing to make a decision can be a way of evading responsibly, pausing and listening can be right way to take up some responsibility.¹⁷ Whatever happens, Herman is willing for the moment to be unknown to herself, certainly she has not—and I think cannot—"decided" who she is. Indeed it seems that she is learning something by *not* understanding herself.

I have tried to bring out that reflective and deliberative models of agency and self-knowledge will not do justice to the significance of Herman's experience, because their conceptual resources are

¹⁵ In his discussion of backward-looking reasons and motives, Anselm Müller takes pains to distinguish such reasons and motives from mere causes. Do opaque motives operate on us like mere causes? Given that her anxiety hardly counts as a consideration that speaks in favor of her criticism, doesn't it seem that Herman's anxiety merely causes her reaction to her son? While it is not a reason or a consideration, her anxiety does seem make her reaction intelligible. But again, isn't this a kind of causal intelligibility? "Given the way my mother raised me, I developed some anxieties that now show up when my son does certain things..." While this helps us understand Herman's reaction, it doesn't justify or speak in favor of it. I think the key for appreciating that Herman's anxiety is not a mere cause (though not, for that matter, a full on reason), is her idea that it "holds together" things she likes about herself as a parent. That it does so means that she cannot simply reject or wish herself to be rid of it. The function of "holding together" also means that the anxiety is conceptually and evaluatively connected with aspects of herself that she avows or values; it is part of how she interprets the world.

¹⁶ Thanks to Kyla Ebels-Duggan for discussing this point with me.

¹⁷ See Pippin (2015) for a discussion of the ethical significance of "not knowing" what one is doing, holding off on the need for clarity, and acting without decisiveness.

too rigid: either her attitude is responsive to reasons and her relation to it is deliberative *or* she relates to it theoretically, as a given fact. Against such a picture I have tried to show that in tarrying with an interwoven dimension of her orientation, Herman remains in a properly *practical* stance, if we allow that the domain of the practical extends further than the narrowly deliberative and immediately action oriented. While Herman describes her encounter with herself as a “problem,” it is not the deliberative problem of deciding what to do in some confined and clear context of action. Engaging with this kind of problem involves relating to oneself as opaque, accepting, as Iris Murdoch proposes, “a darker, less fully conscious, less steadily rational image of the dynamics of the human personality” (1989, 43-44) and engaging in a struggle that is “more obscure, more historically conditioned, and usually less clearly conscious” (1989, 38).

Herman’s scene helps us appreciate that the authorial and epistemic self-relations that characterize self-knowledge and self-opacity are complicated by the idea of an orientation. It is no doubt true that many of my attitudes can be made available for reflective rational consideration, and yet it seems that the more an attitude is embedded in one’s orientation, the less available it will be for either extraction or reflection. It is for this reason significant that Herman’s encounter did not involve “stepping back” and reflecting on herself or her son or her mother; rather, she stumbled, which is to say: there was something involuntary, unexpected, and disruptive about this encounter.

In the next section I turn to Cora Diamond’s idea of “difficulties of reality,” in order to clarify the value of opacity and disorientation. For Diamond, “difficulties of reality” are features of the world that are apprehended *by* being incomprehensible: incomprehension is the proper way to grasp them, where straightforward comprehension would precisely miss the phenomenon and the value at issue. What I want to do is clarify the value of such incomprehension, and also demonstrate

that difficult realities are essentially *reflexive* and *self-implicating*, such that the opacity of the world non-accidentally involves the opacity of the self.

§3. The Difficulty of Reality and the Difficulty of the Self

In “The Difficulty of Reality and the Difficulty of Philosophy,” Cora Diamond works through a series of examples to make a case for the fact and importance of “experiences in which we take something in reality to be resistant to our thinking it, or possibly painful in its inexplicability, difficult in that way, or perhaps awesome and astonishing in its inexplicability” (2003, 2-3). What is it to take something to be inexplicable or resistant to thought, and what might be the value of such an experience?

Diamond’s examples include the following: a photograph of young men who were killed at war, eating animals, an act of goodness in conditions of extreme risk, a child’s childlike response to extreme deprivation, death, the fact of human separateness. While Diamond does not specify any broader categories under which these experiences fall, I will argue that difficulties of reality in fact concern exactly the dimensions of human being that I have discussed in the previous two chapters: our animality, and our profound dependency. As I will put it (though it will take more argument to make this clear): the realities we experience as difficult in this specific sense are the realities *we also are*.

It is important for Diamond that these realities do not always present difficulties to thinking: “it is plainly possible to describe [these phenomena] so it does not seem boggling at all, [such that] there is no problem about our concepts being adequate” (2003, 2). But we *can* take such things to be resistant to our thinking, and for Diamond, this is not only an intelligible but is—and this is the more challenging claim—an *intelligent* response, a way of getting something right and

bringing our world into view, something that would be missed with greater comprehension and self-comprehension.

Diamond does not present a traditional argument for the intelligence of this response or for the value of incomprehension. Indeed, the very nature of what she is trying to convey prohibits such an approach as exactly inappropriate. A central example in her essay is J.M. Coetzee's Tanner Lectures, which take the form of a story of a novelist who gives lectures about our treatment of animals. The lectures themselves do not take the form of an argument or treatise and do not include any actionable recommendations of what we should do differently. We are rather presented with a story of a woman who is undone, almost debilitated, by the thought of our mass consumption of animals. Instead of presenting an argument on the topic, Coetzee/Costello's unusual lectures present "a kind of woundedness or hauntedness, a terrible rawness of nerves. What wounds this woman, what haunts her mind, is what we do to animals" (2003, 4).

How should we understand the idea of raw nerves and hauntedness as a response to a fact of the world? How else might one respond? A major part of Diamond's discussion is spent describing the responses of the official commentators of Coetzee's lectures. As she notes, these thinkers mine the lectures for recognizable philosophical arguments championing better treatment of animals. The commentators read Coetzee as offering, through the character of Costello, an implicit treatise on a moral issue, a kind of submerged lecture on applied ethics, which it is their job to reconstruct and make explicit. In this way, the commentators make Coetzee/Costello's "topic" readily comprehensible and readily practicable, ensuring not only that we straightforwardly understand the topic but can also do something straightforward about it. The commentators thus bring what they take to be the lectures' obscured but wholly intelligible ethical content within the scope of the reflective deliberative stance.

Diamond calls this a move of *deflection*. While philosopher-deliberator takes himself to be helpfully clarifying an inchoate but in principle lucid experience, Diamond sees him instead as deflecting from the kind of disorientation manifested in Costello and presented by Coetzee, and re-orienting by framing the experience as part of a familiar discussion about an ethical-political issue. The commentators turn the story into a deliberative problem that can be encompassed by their ordinary deliberative orientation. While a difficulty remains in the latter case (what ought we do about those animals and are we sufficiently resolved to do it?), it is a difficulty different in kind precisely because it is a difficulty that can be accommodated within our ordinary point of view: we know what it means to deliberate and act. (Thinking back to Herman's case, the rush to "manage" or "outlaw" rather than "listen and learn" from or "be moved" by her anxieties could be understood as just such a case of deflection). But a difficulty of reality is essentially *not* at home in the deliberative perspective, precisely because it resists comprehension from that point of view.

So a difficulty of reality is either apprehended as inconceivable *or* this mode of understanding is deflected and the experience is domesticated and made manageable. This indicates that the philosopher-deliberator does not merely elect a different perspective or mode of knowing; nor is it a matter of the philosopher-deliberator innocently missing or being ignorant of some fact. It is rather, on Diamond's analysis, that the philosopher-deliberator actively avoids an experience of the world, deflects away from what is properly unsettling. Deflection is motivated by a wish not to "feel one's thinking come unhinged" (2010, 12), not only because such an experience is uncomfortable, but because, from the reflective deliberative perspective, such an experience would seem to hold no possible value or promise, would seem mere confusion, at best a way-station en route to the desired clarity.

Because deflection is an avoidance of a certain form of appreciation, it can be tempting to describe it as a straightforward kind of failure.¹⁸ Alice Crary glosses deflection as the unwillingness or inability to rise to the emotional challenge of grasping what is at issue (2011, 348). And this is intuitive: if grasping a difficulty of reality is a mode of awareness and appreciation, then deflection would seem to be a failure to achieve such awareness or appreciation. There is a sense in which this is true: deflection fails at whatever it is that acknowledging a difficulty of reality succeeds at. But in another sense we cannot parse deflection in terms of failing to grasp such difficulties, since the essential point is that these difficulties are, *by definition*, something that we cannot grasp in any straightforward way. Faced with these difficulties, thought and reality *fail* to meet (2010, 25). Or as Simone Weil notes, “human thought is *unable* to acknowledge the reality of affliction” (quoted in Diamond 2003, 23). In this way, inability is constitutive of the experience; failure is how we register these realities. So whatever it means to deflect, it is not just a matter of failing to grasp what is at issue. In deflecting, the philosopher-deliberator doesn’t simply fail to grasp something, rather he avoids the failure of his grasp. Deflection is a failure to fail in the right way.

Now, let us imagine that we accept Diamond’s diagnosis that the commentators deflect and that in so doing they miss something important. One way we might understand deflection is as involving a move from an aesthetic appreciation to a practical stance. On this interpretation, the commentators deflect from a difficult and important *aesthetic* experience, by re-framing it as a *practical*

¹⁸ It also might be tempting to define it as a paradigm case of self-deception: avoiding knowledge thanks to a wish that it not be true (because, for example, acknowledging the truth would be painful). The problem is that it would not fit the phenomenon to refer to a difficulty of reality as evidence that gives one reasons to believe something. Even though deflection avoids encountering something, there is no straightforward, positive alternative. As we shall see presently, failure is internal to a difficulty of reality: it is precisely incomprehensible. By contrast self-deception concerns the avoidance of more ordinary forms of knowledge and evidence that one *could*, both in principle and in fact, straightforwardly apprehend.

issue.¹⁹ So we might think that the mistake is to hear Coetzee's aesthetically arresting lectures as some kind of moral-political propaganda or call to action, where this mistake pushes them out of an experience of aesthetic appreciation and into the practical.

However I think this interpretation involves taking for granted the narrow picture of the practical as reducible to the deliberative. I argued in the last section if we accept that a person is actively engaged in her orientation or moral vision or stance towards the world, this requires that we acknowledge that the practical perspective is broader than the deliberative and choice oriented. The concept of orientation is relevant here because it is essential that a difficulty of reality constitutes a form of practical *dis*orientation. It shoulders us from our familiar sense of moral life, "from our sense of being able to take in and think a moral world" (2010, 15). Such an experience destabilizes our ability to orient, to coherently think and move in a moral world from a coherent, stable position. This means that while a difficulty of reality may be anchored in an aesthetic object, it cannot be understood as a narrowly aesthetic response to that object, precisely because, as a form of disorientation, the response cannot be confined to that object alone but rather implicates one's basic practical hold on the world. This is not to suggest that the object is a mere occasion, like a fungible springboard to some further, more important experience. It is rather that the object needs to be understood, not as an isolatable thing, but as an inextricable part of a broader moral reality: the experience of incomprehension and incapacity with respect to *this object* implicates the broader orientation through which experiences ordinarily make sense. If an orientation is the frame of ordinary experience, then as we saw with respect to Herman's case study, this is not merely an encounter with a difficult object but an experience that unsettles the frame itself. Because of the

¹⁹ I actually think that the very idea of moral vision makes drawing any clean distinction between aesthetic and practical distinctions hard, if not possible. I lean on the idea that we can draw a contrast here in order to bring out the idea that difficulties of reality *are* practical even if not deliberative, rather than to insist that they are *not* aesthetic.

way in which one's orientation is implicated and challenged, a difficulty of reality must be understood as an experience that is only possible from the practical perspective, construed more broadly than the deliberative.

Drawing this connection between difficulties of reality and orientation helps make sense of Crary's remark that "we only bring [these realities] adequately into focus insofar as we treat them as occasions for modes of thought that...essentially call on us to work on ourselves" (2011, 348).²⁰ While the philosopher-deliberator deflects in the direction of action, Crary is proposing that not deflecting involves standing in an ethically significant, self-relation. Because of this self-implicating element, we could say that these experiences are precisely *not* transparent to their object: one cannot, as it were, forget oneself and attend only to the object. Yet there is a way in which the transparency condition does hold here: it is precisely *because* I cannot comprehend or otherwise make sense of this object that I am to that extent unable to make sense of myself; such that the very unclarity of the object entails an opacity of the self. So these disorienting encounters prompt a pronounced self-relation that involves being exposed to a kind of self-blindness: insofar as *this* bit of reality, heretofore unproblematic, can now throw me, I am also thrown out of my ordinary orientation, from my sense of world and my sense of self.

I want to focus now on this issue of reflexivity: why should difficulties of reality necessarily involve difficulties with the self; why must this be a reflexive, self-implicating relationship whereby the *self* is experienced as opaque?

Let us begin with a thought experiment: try to imagine a non-reflexive difficulty of reality, where *only* the object were incomprehensible, but not me. Imagine that I am staggered by the

²⁰ I think we should be cautious as to how to interpret this idea that difficulties of reality call on us to "work on ourselves." I would simply want to qualify Crary's helpful elaboration and insist that the point is *not* to funnel such an experience into a concerted effort of self-transformation; to do so would be another form of deflection. So whatever "work" we do will not be an example of deliberation and decision making.

thought of death. I have not always found this to be an impossible thought, but now I do: the reality of death resists my comprehension. The question is: what would it mean to say that while I find death incomprehensible, I myself remain perfectly coherent and comprehensible? In this imagined situation, it is as though I am perplexed by a strange fact—that life ends—but do not grasp that this is true of *me*: that I will end. To imagine someone registering a difficulty of reality that was wholly world-directed is to imagine someone struggling with a kind of puzzle. In this way, I might wonder and marvel at the fact that life ends, but what I have not done—and this is essential to difficulties of reality—is acknowledge the internal connection between this staggering fact about life and death, and myself. Difficulties of reality are difficulties of which the self is a part.

As should be clear, a non-reflexive, non-self-implicating registration of difficulty is in fact a form of deflection. It involves someone struggling with what Diamond calls an “issue,” something about which philosophers might debate, a topic about which we might deliberate. These are forms of deflection because they deflect the difficulty out and away, so that it doesn’t directly bear on me in any way that could shake me: while something in the world has become puzzling, I remain oriented.

But the kinds of things that prove incomprehensibly difficult in Diamond’s sense precisely cannot be registered as difficult while being kept at arm’s length. I am suggesting that something can only be difficult in Diamond’s specific sense when the difficulty involves *being* the thing that can’t be thought. As she puts this challenging idea, “I am inviting you to think of what it would be not to be deflected as *an inhabiting of a body*...in the appreciating of a difficult of reality” (2010, 13-my emphasis)). Registering a difficulty of reality involves inhabiting a body. I propose we understand this to mean that appreciating a difficulty of reality means that one also *is* that very difficulty, where experiencing reality as incomprehensible just *is* to experience oneself as incomprehensible. In these

experiences, reality and self are two faces of an inextricable whole, precisely because the self is part of that very reality. It is only because of the reflexive, self-implicating nature of such difficulties that it make sense that they can unhinge a person in the way Diamond thinks they do.

Can anything in the world be difficult in this way? While Diamond does not make this explicit, we can discern two broad categories into which her examples of difficulty fall: our animality, and our dependency on the world and others. The preceding discussion clarifies why it should be just these realities that throw us: these are the realities that we also are. We are embodied, creaturely beings, strangely like and unlike other animals; and we are separate from and wholly dependent on the world and other persons, who are like and unlike us.²¹ These are the realities that shoulder us out of our ordinary orientation when we try to think ourselves into them or inhabit them in a bodily way. If experiencing these realities as difficult is a way of appreciating them, then a failure to appreciate these aspects of reality also amounts to a failure to appreciate aspects of oneself.

Of course, our animality and our deep dependency were the topics of Chapters Three and Four of this dissertation, exactly those aspects of our agency that cannot be adequately encompassed by the reflective, deliberative perspective, and that make self-opacity essential to our condition. Of course from certain perspectives there is no problem of our concepts being adequate: if we figure them as considerations in our deliberations or turn “our own bodies into mere facts” (Diamond 2010, 13)—which is to say, view them from a theoretical point of view—then these aspects of our life are not perplexing or resistant at all. Yet they *can* resist. Such resistance marks not simply a failure to understand some thing but a failure to understand ourselves, yet where, again, *this* kind of failure is *productive*, a failure of ordinary understanding through which we register in a living way our dependent, animal being in and of the world.

²¹ These two categories of animality and dependence might be brought under the still broader category of finitude, though I think Diamond does not use this word because it is abstract, does not land with our form of life the way that bodiliness, vulnerability, and separateness do.

Recall Nagel's claim that "we are *unable* to view ourselves simply as portions of the world," that "something in the idea of agency is *incompatible* with...people being things" (2012, 37). We can now see that there is something right in Nagel's sense of inability, and that actually Diamond is grappling with this very same sense: something in the idea of our being finite dependent animal bodies in the world cannot be straightforwardly reconciled with our sense of ourselves as self-conscious deliberating agents.

Yet there are important differences in Nagel's and Diamond's approaches. Nagel frames this as a straightforward inability or as a brute incompatibility between two views we can take of ourselves: we can see ourselves *either* from the perspective of agency *or* we can see ourselves theoretically, as things, but *as* agents we *cannot* view ourselves as of the world. By contrast, Diamond recognizes that we cannot get our minds around the complex fact that we are *at once* agents and things in the world, and yet for her this is not a brute contradiction or a mere failure. Instead Diamond regards the unthinkability as a productive inability that constitutes a valuable, practical ethical experience. The sense of incomprehension and inability makes uniquely available the aspects of our lived reality that are simply not available on Nagel's picture.

In both Herman's case and Diamond's cases, the agent encounters something that cannot be made sense of through her ordinary orientation, rendering both the object and the self incomprehensible, opaque. In Herman's case, this is a more personal, local problem: I cannot understand myself (in relation to my son), given these insistent anxieties and whatever desires they express: there is something of *myself* (not an impinging force or outlaw desire) that I can't get at. In Diamond's case, the problem takes more global proportions: I cannot comprehend myself in relation to the world, given the sheer fact of my creaturely, dependent being. But notice that the more narrowly individual and the broadly human levels intersect: part of what makes Herman's

personal experience unsettling is her emphatic registration of the fact of human dependence (she on her mother, her son on her), and Diamond is explicit that difficulties of reality are “inextricably intertwined with relations of power between people” (2010, 11), where, for example, any difficult registration of human dependence in general will be shot through with personal and interpersonal inflections.

At the close of her essay, Diamond writes that the coming apart of thought and reality “belongs to flesh and blood.” There are two dimensions to this thought: first, Diamond is claiming that being thrown out of the ordinary in fact belongs to our ordinary moral life. That is, part of what it is to be part of a moral life, to be a moral animal, is to be open to experiencing this fact as staggering and unthinkable. In such moments, the expectation that our ordinary modes of thought should be wholly adequate is misplaced. That we cannot live well in a constant state of difficulty does not mean that the experience of difficulty is not part of living well.²²

Second, that the coming apart of thought and reality “belongs to flesh and blood” brings attention to the fact that difficulties of reality concern our animality and dependency. So incomprehension and opacity are part of our moral life, not only in the sense that they are internal to the ordinary, but in that they are part of *life*, i.e., part of being alive. These are experiences proper to rational animals who cannot everywhere bring that life into a thinkable unity. Difficulties of reality are expressive of the productive unrest of our rational animality.

²² Living in a *constant* state of difficulty might be one way to characterize the wrong of systemic oppression. For example, we might understand W.E.B Du Bois’ concept of “double consciousness” as giving voice to an experience of reality as *consistently* difficult in virtue of the fact that, as Black, he had to navigate the world through two visions: through his own eyes and through the eyes of the White world around him. Everyday life involved being thrown from any straightforward relationship with reality. Insofar as he had consistently to see himself as an object in the eyes of White people, Du Bois had consistently to engage in the task that Nagel thought simply impossible. So Du Bois had to live in and through a difficult reality, rather than have his ordinary orientation upended in productive ways.

In the next and final section I will clarify the structure common to the kinds of encounters described by Herman and Diamond, which will in turn help clarify the ethical value of these encounters.

§4. Being Real

Following Herman and Diamond, I have tried to show ethically significant ways in which our self-opacity shows up, namely by disrupting the ordinary first-personal practical perspective. I argued that making sense of these familiar, ethically significant experiences depends upon the idea of self-opacity, where the opacity that blinkers us in these experiences is essential to their value, such that greater self-knowledge and self-possession would diminish it.²³ These experiences all involve a kind of learning through and with self-opacity. Now this might seem like a trivial point: of course, if we are to learn anything about ourselves, we must have first *not* known whatever it was. But the point is not that opacity is like the ignorance one leaves behind in achieving more knowledge. In the case of difficulties of reality, one does not learn some new information that one did not know before, either about oneself or about reality; instead, the experience just is the vivid experience of *not* knowing. In the case of encountering one's orientation or complex, one may learn something new about oneself and one's relations with others, but it is crucial that this *mode* of experience involves registering oneself *as* opaque, as *less* knowable or familiar.

Herman's and Diamond's scenes share a common structure: they involve an unwilled and unexpected encounter with something that resists ordinary comprehension, where such resistance is not *mere* incomprehension (this is how it is seen under the aspect of deflection) but somehow—in ways still to be clarified—ethically significant. Further, the “something” that resists comprehension

²³ See Korsgaard (2011) for an effort to re-conceive the ethical and practical significance of disruptions as moments in which to enact even greater self-possession and integrity.

is essentially reflexive: in Herman's case, it involved an encounter with herself; in Diamond's case, it was an encounter with something that implicated the self *qua* dependent animal: the idea that we *are* the very thing we cannot comprehend was how I analyzed Diamond's proposal that such encounter involve "the inhabiting of a body." In both cases, the encounter involves registering oneself in a mode that resists and challenges the ordinary, practical perspective of agency because involves the first-person practical experience of oneself as a non-self-sufficient dependent rational animal that is implicated the world, something Nagel thought to be wholly unthinkable.

If these experiences practically disclose one's non-self-sufficiency, it is no accident that they take the form of disorientation and disruption, of something necessarily unwilled and irrecoverable within the deliberative standpoint. In this final section I want to make perspicuous the ethical *content* of these kinds of experience and how this form is internally related to their *form*. I will do so by looking at one final exemplar, the experience of being an *object* for another *subject*. This particular experience is important to emphasize, because the encounter described thus far might not seem to involve other people *as* subjects. While others are, as it were, on the scene—Herman's son (and mother), Diamond's experience of the "separateness of persons"—they seem only to occasion my experience and do not really figure as subjects in their own right. So my aim here is to show how the experience or practical apprehension of an other *as* a subject can be both ethically essential and essentially disruptive in the ways we have been seeing.

In an essay on respect for persons, Sarah Buss argues that "too little [philosophical] attention is spent on the relationship that obtains between respecting people and apprehending them in a special way" (1999, 519). She analyzes the special phenomenology of experiencing another person *as* a proper subject, and she argues that recognizing another person as a proper subject—with a

point of view on the world and ends of her own—necessarily involves *experiencing oneself as an object*, where this phenomenology of such an experience involves a kind of existential dislocation.

Buss writes,

The only way to experience another subject *as a subject* is to experience oneself as an object that *he* contemplates, makes something of, evaluates from *his* perspective. But to experience oneself *as* such an object is incompatible with regarding the Other's appraisals as just so many characteristics of an object in one's world...In becoming conscious of oneself as an object in the world, one experiences the abrupt demotion of one's self from its exalted position. 1999, 525.

Buss' point is that experiencing another person as a subject, rather than as an object for me, is not only to experience the other as irreducible to however I comprehend him; it also involves the realization that *I* am irreducible to however I understand myself, that I am more than what I can know from the first-person perspective. For Buss, essential to this experience is the awareness that one's being in the world is "not conditional upon one's own point of view" (*ibid.*), where such awareness "effects a dramatic shift in one's conception of oneself" (1999, 525) from one in which the other is straightforwardly available *for me*, to one in which I experience myself as something for the other. In Diamond's terms, we can say that this experience involves the "inhabiting of a body," a kind of "exposure" (a word she considered adding to the title of her essay). Crucial to this experience of the other as subject and self as object is the emphatic realization of one's lack of total self-sufficiency (1999, 526), the realization that one cannot "encompass oneself completely" to use Nagel's words.

Like stumbling over one's own orientation or a difficulty of reality, the experience of oneself as an object for another subject involves an unwilled and unexpected encounter with something that resists comprehension from the first person point of view. For Buss, what makes this encounter ethically crucial is that it discloses the other person *as* a subject; it is the vivid experience or apprehension of the other as respect-worthy. So for Buss, the upshot of this disruptive, dislocating

encounter is that it makes the other available to me in an ethically significant way. But given the exposure or reversal effect by this experience, we can also ask if there is any ethical significance to experiencing oneself an *object*.

I propose that such an experience can be ethically significant by being, as I want to put it, “reality conferring.” Experiencing my deeds as mediated by or *for* an other—which is to say significant, meaningful, and real *for* another—provides them a kind of reality that I could not give them as their author. That is, being an object for another gives *me* a kind of reality that I cannot provide for myself, a kind of reality that cannot be wholly comprehended from the first person deliberative point of view.²⁴ Here, again, the unwilling and unexpected encounter with something that resists ordinary comprehension has an essentially reflexive dimension: the difficulty of experiencing the other as a subject involves the difficulty of experiencing oneself as an object. This is what I have been calling the emphatic realization of one’s lack of self-sufficiency. Part of the value of this experience is thus that it lends one a kind of reality that one is structurally unable to realize for or by oneself.

This means that the gain in reality that comes from being an object for another is inseparable from a kind of loss. This helps explain why this is exactly the kind of experience that would motivate deflection, in two senses. First, experientially or practically, because experiencing oneself as an object involves a loss of (the fantasy of) self-sufficiency, this is the kind of experience that we agents might be inclined to avoid or try recuperate into the deliberative perspective. But second, at the level of philosophical theorizing, because the idea that seeing oneself as an object puts pressure

²⁴ I think this idea is echoed in Bernard Williams, when he writes: “if one attaches importance to the sense of what one is in terms of what one has done...one must accept much that makes its claim in that sense solely in virtue of its being actual” (2002, 29-30). That the *actuality* of my life makes a claim on who I am, is the idea I am trying to make sense of in this section.

on certain ideals of self-sufficient rational agency, this is the kind of insight that we philosophers might be inclined to discount, as either trivial, or impossible, or even morally suspect.

This is not to say that being seen as an object is *always* good or reality conferring in an ethically productive way. To appreciate this we need only look to philosophical theories of oppression, so many of which are articulated in terms of the moral and existential damage of being regarded as a kind of object (Frantz Fanon, for instance, talks about the effects of racism in terms of being “sealed into that crushing objecthood” [2008, 109]). In insisting that the ethically significant experience of reality involves the experience of objecthood and non-self-sufficiency, I am not rejecting views that insist that agency is realized in successful deliberative activity but rather suggesting that these views are incomplete or one-sided. As Hannah Arendt points out, there are necessarily two aspects of the acting subject:

Because the actor always moves among and in relation to other acting beings, he is never merely a ‘doer’ but always and at the same time a sufferer. To do and to suffer are like opposite sides of the same coin, and the story that an act starts is composed of its consequent deeds and sufferings. 1998, 190.

In a philosophical climate that focuses almost exclusively on agency as self-directed self-realization, and that either recoils from or positively degrades that idea that we are also objects, it is important to remind ourselves that our practical reality *as* agents essentially involves passivity and the experience of being comprehended by another, as well as activity and self-comprehension.

If there is a value proper to being an object for another, how then can agents pursue this good? In fact, we can’t. Part of the unique value of this experience depends on its being exactly the kind of thing one cannot choose or control or deliberately try for. This is how the *form* of the experience—unwilled disruption—is essential to the *content*—the ethically significant experience of non-self-sufficiency. That this experience is both unwilled (and often unwanted) and resists ordinary comprehension is essential to the value of these encounters and what they disclose. So Korsgaard is

right that such experiences of non-self-sufficiency fundamentally challenge one's sense of oneself as an agent. What I have been trying to argue is that such challenges need not for that reason be any less essential to that very sense of self.

*

In this chapter I have explored the ethical and practical importance of *not* being able to encompass oneself completely or fully comprehend oneself. Self-opacity, the failure of ordinary practical self-understanding, is crucial to these experiences, not only in the sense that they reveal one's self-knowledge as fundamentally limited, as shot through with shadow, but because their basic structure involves challenging or otherwise "shouldering one out" from the perspective proper to such self-understanding. In the scenarios I have analyzed, suspending the work of self-comprehension and yielding the deliberative stance makes available certain difficult encounters. And though one cannot, by definition, wholeheartedly pursue them, one can try to remain available to these experiences, to not deflect, despite the inevitability of the tendency to do so. So it can be important that there are times and contexts in which we should *not* try to "be as we want to be, to as deep a level as possible" (Nagel) or be wholly self-satisfied (Frankfurt) or to endorse as much of ourselves as we can (Moran). This means seeing that there can be value in restlessness and resistance, that is, in *not* realizing these deliberative ideals of agency. To unreservedly try do so would foreclose the possibility of the kinds of experience described in this chapter.²⁵

²⁵ See also Arpaly (2006) for a discussion of cases when people seem to want the absence of self-control or self-authorship.

Conclusion: “Something You Can’t Get At.”

In a discussion of Freud in *Contesting Tears*, Cavell makes the following remarkable claim:

the advent of psychoanalysis [is] the place, perhaps the last, in which the human psyche as such, the idea that there is a life of the mind, hence a death, receives its proof. It receives its proof of existence in the only form in which that psyche can (any longer) believe it, namely, as essentially unknown to itself. 1994, 94.

What could it mean to say that the psyche’s existence is proved, *to itself, as unknown, to itself*? Why does the mind know itself to be real only as opaque? What is this connection between self-opacity and proof of reality?

As I have discussed in the previous chapters, the idea that there is an important relationship between opacity and reality cuts against a prominent guiding commitment in moral philosophy, namely that the person realizes herself through self-conscious free activity. Whether the activity we imagine is stepping back and reflecting, or deliberation culminating in decision, or the unfettered execution of action, it can seem as though we human beings prove our unique reality precisely in self-conscious, self-directed activity. On this picture, the self proves itself by proving its difference from other self-moving creatures and its independence from the world. Korsgaard, for instance, writes that whereas animals are moved by instincts that they do not choose,

we human beings on the other do choose the principles of our own causality ... in one sense, to be autonomous or self-determined is to be governed by the principles of your own causality, principles that are definitive of your will. In another, deeper, sense to be autonomous or self-determined is to *choose* the principles that are definitive of your will. 2009, 108.

On this picture, the self’s proper activity is free undetermined activity, organized by reflectively chosen principles. If the attitudes and principles that guide my actions are *not* reflectively chosen and endorsed but are simply the result of who I happen to be (my natural propensities, how I was

raised, the society in which I find myself, etc.), then I am, to use Korsgaard's term, "a mere heap." This is a demanding conception of agency and selfhood, for as Bernard Williams points out, it requires "that my whole outlook should in principle be exposed to a critique, as a result of which every value that I hold can become a consideration for me, critically accepted, and should not remain merely something that happens to be part of me" (2008, 158). And we see versions of this in Frankfurt's acts of self-ordering in the pursuit of self-satisfaction and Moran's unreserved endorsement of the project of self-avowal. On all of these pictures, the psyche or self is most real when it realizes itself.

So what could Cavell mean with this idea that the mind knows itself to be real *only as unknown*? Why should self-opacity confer reality?

One way to understand reality is, as Iris Murdoch suggests, as other and more than the mind that knows it. This is the real as "something which my consciousness cannot take over, swallow up" (1999, 373). It is essential to something's being real in this sense that it exceeds straightforward comprehension ("swallowing up"), not in the sense of containing some hidden essence or being essentially out of reach or of another metaphysical order, but in the sense of being something which the mind must work to understand, and may not. Different concepts bring this reality into different lights; different stances—practical, aesthetic, theoretical—afford different understandings of it. I can work with and through reality, my orientation might bring it into view or obscure it, but most basically, it is not up to me. The world is real in this sense; as are our animal bodies; as are other persons.

I believe that Cavell's proposal is that the mind is also real in this dense sense and that it experiences itself as real, in that sense, by being unknown to itself (though not simply unknown or totally unknowable). In *The Claim of Reason* Cavell suggests that self-knowledge is both something I

do and something I *suffer*. The idea that self-knowledge is something I do is, I take it, the aspect that the deliberative approach wants to emphasize as the proper form of self-knowledge. But a self-knowledge that I *suffer* is closer to a kind of self-opacity—a relationship to one’s mind that is not “up to me.” And Cavell’s point seems to be that this is not a regrettable failure of the project of self-avowal but is an essential part of what it is to have a mind that knows itself as real.

As minded creatures, one of our greatest activities and achievements is that we can maintain attitudes and engage in actions for reasons we endorse, and to take up different positions as reason dictates: in response to reasons we can change our minds. Moreover, we know our own minds *by* knowing those reasons: if you ask me what I think about the painting or person in front of me, I usually can tell you what I think—so tell you about myself—by telling you about *it*. Oftentimes I know what I want by *deciding* what I want, sometimes by deciding what would be best. When I know myself in this way, my self-knowledge is transparent to my knowledge of the world.

One way that Moran argues for this deliberative, agential rather than epistemic conception of self-knowledge is by way of a thought experiment: imagine an epistemically idealized individual with total perceptual or epistemic access to his own mind. He has unimpeachable epistemic authority, he knows everything about his mind. “And yet,” writes Moran, “for all this immediacy, accuracy, and reliability, there is a sense in which such a person would remain opaque to himself” (2001, 90-1). This is a picture of self-knowledge as a kind of mind-reading as applied to oneself, where no matter how epistemically successful it is, it would never amount to “ordinary first-person knowledge” (*ibid.*, 91). And as McGeer adds, “there is something deeply wrong with a person thus related to [his] own mind. Whatever drum [he] is marching to, it is not of [his] beating” (2007, 85).

Now let us imagine the opposite figure: imagine an ideally rationally deliberative agent, whose attitudes spontaneously conform to the dictates of reason. Such an agent enjoys

unimpeachable rational agency; appreciating the reasons there are for some belief or desire is enough to make up her mind about what *she* in fact believes and desires. There is no space between knowing what is desirable and knowing what *she* desires (see Lawlor 2009). And yet for all this rationality and authority, there is a sense in which such a person would be too transparent to herself, or rather, too transparent to reason. Whatever drum this person is marching to, it is “up to reason” to an unnerving degree.

The problem with the epistemically idealized individual is that his mind was never *his*, he knows it only as a kind of object that remains unresponsive to his own rational deliberations. The problem with the ideally deliberative agent is that her mind is too completely reason’s,⁹⁷ she knows it only by making it up for good reasons. Of course, these cases are not symmetrical in terms of what they prove: in the former case, we lose the very idea of the person’s mind as *that* person’s mind or, even stronger, as a mind at all; in the latter case we lose the idea of the person’s mind being a *person’s* mind (in the sense of a finite, imperfect, not exclusively rational human being). The former is a conceptual impossibility (the idea that there could simply be thoughts yet no active thinker of them), the latter is a human impossibility (the idea that there could be a human mind wholly and only responsive to reason). Still, I think it suggests that a conception of a mind that was *wholly* reason’s own leaves us with a deeply unfamiliar picture. Part of what is unfamiliar about it is that it seems to lack exactly a kind of density and recalcitrance, which is to say reality.

We can often make up our minds in light of reasons, and yet we, our minds, are also conditioned and shaped by our creaturely natures and our drives, by our *Bildung*, by our habits and orientations, by the larger social world, by the ways our actions have significance and reality for others in ways we could never give them ourselves. This is not to say we are “mere heaps” shaped

⁹⁷ In his critical response to Moran, David Finkelstein makes a similar kind of observation, writing: “when [on Moran’s picture] I avow an intention to do something, it’s as if the reasons for doing it speak *through* me; I’ve bound myself to them, and now I intend whatever course of action they dictate” (2003, 161).

from without, or that these aspects of ourselves are brute facts to which we relate from a purely theoretical perspective. These contingent—and as I have emphasized, bodily and social—aspects of our agency have seemed almost regrettable to some philosophers,⁹⁸ something to be transformed as much as possible by our own self-conscious, rational activity. As Nagel confesses: “I wish to subject my motives, principles, and habits to critical examination, so that nothing moves me to action without my agreeing to it” (1986, 119). And Korsgaard insists, “making the contingent necessary is one of the tasks of human life and the ability to do it is arguably the mark of a good human being” (2009, 23). Or again, “your identity is never just a given...and how you construct it is really up to you” (2011, 79). To make the contingent necessary is to make the given chosen, make it *mine*.

Cavell is offering an alternative vision of the mind and self. For Cavell, it is not merely a matter of fact that we cannot subject every part of ourselves to reflection and endorsement; again, this might leave it open that it is still the ideal to strive for. His point is that the relentless effort to achieve such an ideal would constitute a loss: the mind would lose its grip on itself as real. On this picture, it is the mind’s self-opacity, its very resistance to reflection, comprehension, and endorsement, that grounds our sense of ourselves as real, as worldly beings that matter. Such a mind may satisfy all the claims of reason, but having such a mind would not be, as Cavell would put it, “humanly satisfying” (1999, 455).

A philosophical account of agency and moral psychology that emphasizes self-opacity and yet is committed to understanding agents and minds “internally” can thus be understood as “realistic,” not only in the sense of “not idealistic” or “not metaphysically extravagant,” but in the

⁹⁸ There is a nice passage in Hegel where he explicitly states that philosophers (and persons) need not be embarrassed by embodiment and the way this shapes our practical lives: “We may ask at this point whether the human being has a right to set himself ends which are not based on freedom, but solely on the fact that the subject is a living being. *The fact that he is a living being is not contingent*, however, but in accordance with reason, and to that extent he has a right to make his needs his end. *There is nothing degrading about being alive*, and we do not have the alternative of existing in a higher spirituality.” (RPh §123Z)

sense that it seeks to understand the mind of the real world, where its being densely real is *part of* the mind's own (opaque) self-understanding. In this way, I am offering a qualified contestation to Nagel's conviction that "we are unable to view ourselves simply as portions of the world." We are perhaps unable to view ourselves *simply* as portions of the world, to get outside the perspective of agency and see ourselves as mere things. But this does not mean that the mind's reality simply escaped the first-person perspective, or that there is no way to see ourselves as *of* the world and real. In the last chapter I proposed that experiencing ourselves as more and other than rational agents, as deeply non-self-sufficient, involves modes of thinking and appreciation that have resistance, failure, and opacity built into their forms.

*

Murdoch claims that "in order to tell the truth, especially about anything complicated, we need a conceptual apparatus which partly has the effect of concealing what it attempts to reveal" (1991, 221). She is speaking about art, and literature in particular (though it could also just as well be about psychoanalytic discourse⁹⁹). I find that Murdoch's observation testifies to the difficulty of telling the truth about anything as complicated as a self-opaque mind, and suggests that philosophical analysis alone may not be sufficient for this kind of truth. In fact, it suggests an internal relationship between self-opacity and the methodological and conceptual resources proper to art (a connection Freud recognized with his concept of sublimation). As I take this project forward, it will be essential to explore the relationships between self-opacity and artistic expression, including the idea that art is a sufficiently "concealing" way of telling truth, and the idea that our experience of art *as* so concealing allows us an encounter with our own self-opacity, an encounter made significant precisely

⁹⁹ Gardner (2007) characterizes the aim of psychoanalytic treatment as the acknowledgment of psychic reality *and* the acknowledgment of the impossibility of assimilating it. In this way, the allowance for concealment is not only part of the method but also part of the aim.

thanks to the non-transparency or non-obviousness of its significance. For now, though, let me end by returning to *Scenes from a Marriage*.

As I presented this film in the Introduction, the film can be viewed as an exploration of the necessity *and* limits of avowal, and of the ways in which human beings can be opaque to themselves without for that matter being radically self-deceived. At the beginning, Johann especially is committed to taking his own self-conception very seriously, to endorsing and committing to who he is and what he believes. This is presented as a kind of ethical shortcoming. There is nothing he is exactly deceived or wrong about—for example, it seems theirs really *is* a happy marriage, we are not invited to be skeptical about this. It's just that he is a little too sure. Marianne was never so self-certain, and as the film develops, Marianne develops a capacity to relate to herself as self-opaque. This is presented as a kind of ethical achievement. This non-defensive capacity allows her to relate to both herself and Johann with compassion, care, and humor. It gives her space to learn and change without needing to wholeheartedly identify with the place she arrives at. She never achieves the kind of total self-satisfaction Frankfurt speaks of; rather she comes to experience her relation to self and world as what Cavell calls “humanly satisfying” (1999, 455).

The film's final section is entitled “In the Middle of the Night in a Dark House,” with Johann and Marianne away in a country cottage. By this time they are divorced and both remarried to other people, and now they are some kind of intimates, old friends who genuinely like each other, and who seem to understand each other and also to be less invested in understanding each other.

In the last scene, Marianne wakes up terrified from a dream, leaping from the bed, eyes wild. It is striking and significant that Bergman ends his film about human relationships and self-understanding with a dream, which is to say: with a disruptive display of the self as unknown to itself. Marianne marvels: “I can't think why it is I dream like that. What do you think causes it?”

Johann replies: “in your extremely well-ordered world, there is something you can’t get at.” “What would that be?” she asks. “How should I know?” he says.

Marianne is at first frightened and disoriented, grasping at the meaning of her dream and what it might tell her about life (“do you think we’re living in utter confusion?” she wonders), but she eventually calms down, and in a very particular way: she laughs at herself, and at Johann with his self-seriousness and lack of imagination (a trait which he cannot avow it, but now he can acknowledge), and she allows herself to be held. After so many years of knowing each other, moments of unknowing still break out, and while they can disrupt and unsettle and alienate, they can also be strangely grounding.

Our self-opacity can get in the way of our effective agency, but it is also what opens us to experiences that I would call stirring: disorienting and harrowing; also humbling, moving, and light. It is Marianne’s disruptive, opaque dream that allows she and Johann to share a moment of gentle humor, intimacy, and care. Indeed it seems that it is precisely and only because she cannot know or encompass herself completely—as the dream makes clear—that she can allow herself to be comforted and held by Johann. This is an instance of what Candace Vogler calls “depersonalizing intimacy” (1998, 329), a kind of intimacy “that has nothing to do with tracking or expanding the borders of selfhood” (ibid., 331). *Scenes from a Marriage* begins with a scene of exactly such tracking, and it ends with a moment of reprise from this task. Marianne and Johann give each other the space to forget who they are for a bit, or at least have a sense of lightness about it. In these intimacies, “something alien to the well-tempered self can act out, without [our] becoming overwrought about this” (ibid., 365). Or Johann puts it, smiling: “there is something you can’t get at.”

Our availability for these intimacies depend on our resisting the inevitable, all to human impulse to deflect: to either wholeheartedly identify with or endorse these aspects of who we are *or*

reject them as aliens and outlaws. These un-encompassable, un-excisable parts of the self are the self as real and unknown, real because unknown.

In this dissertation I have tried to show that self-opacity plays a more central role in moral life than has been acknowledged and that our moral philosophy and moral psychology should try to carefully and sympathetically reflect this. This is not only because philosophy would be thereby more realistic or true to experience. It is because it would allow us to appreciate the otherwise unrecognized kinds of value that both animate and complicate our human kind of life.

*

Bibliography

- Arendt, Hannah. *The Human Condition*. University of Chicago Press, 1998.
- Arpaly, Nomy. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press, 2003.
- _____. "Comments on 'Lack of Character' by John Doris." *Philosophy and Phenomenological Research*. Vol. 17, No. 3. 2005.
- _____. *Merit, Meaning, and Human Bondage*. Princeton University Press, 2006.
- _____. "Unprincipled Virtue: Synopsis (Of Sorts)." *Philosophical Studies: An International Journal of Philosophy in the Analytic Tradition*. Vol. 134, No. 3. 2007
- Arpaly, Nomy and Timothy Schroeder. "Alienation and Externality." *Canadian Journal of Philosophy*. Vol. 29, No. 3. 1999
- Bagnoli, Carla. "Self-Deception: A Constructivist Account." *Humana.Mente*. 20, 2012.
- Baron, Marcia. "What is Wrong with Self-Deception?" *Philosophical Perspectives on Self-Deception*. Eds. B. P. McLaughlin and A. O. Rorty. University of California Press, 1988.
- Bäckstrom, Stina. "What is it to De-Psychologize Psychology?" *European Journal of Philosophy*. Vol. 24, No. 4. 2017.
- Baz, Avner. "Motivational Indeterminacy." *European Journal of Philosophy*. Vol. 25, No. 1. 2017.
- Benson, Paul. "Taking ownership: Authority and Voice in Autonomy." In *Autonomy and the Challenges to Liberalism: New Essays*. Eds. J. Christman and J. Anderson. Cambridge University Press, 2005.
- Bernstein, J.M. "Re-Enchanting Nature." *Reading McDowell: On Mind and World*. Ed. N. H. Smith. Routledge, 2002.
- Boyle, Matthew. "Essentially Rational Animals." *Rethinking Epistemology*. Eds. G. Abel and J. Conant. De Gruyter, 2012.
- _____. "Additive Theories of Rationality: A Critique." *European Journal of Philosophy*. Vol. 23, No. 4. 2016.
- Brewer, Talbot. "Maxims and Virtues." *The Philosophical Review*. Vol. 111, No. 4. 2002.
- Brudney, Daniel. "Knowledge and Silence: *The Golden Bowl* and Moral Philosophy." *Critical*

- Inquiry*. Vol. 16, No. 2. 1990.
- _____. "Styles of Self-Absorption," *The Blackwell Companion to Philosophy and Literature*. Eds. G. Hagberg and W. Jost. Wiley Blackwell, 2010.
- Buss, Sarah. "The Superficial Unity of the Mind." *The Massachusetts Review*. Vol. 47, No. 2. 2006.
- _____. "Respect for Persons." *Canadian Journal of Philosophy*. Vol. 29, No. 4. 1999.
- Butler, Judith. *Giving an Account of Oneself*. Fordham University Press, 2005.
- Callard, Agnes. "Proleptic Reasons." *Oxford Studies in Metaethics 11*. Ed. R. Shafer-Landau. Oxford University Press, 2016.
- Calhoun, Cheshire. "Standing for Something." *The Journal of Philosophy*. Vol. 92, No. 5. 1995.
- _____. "What Good Is Commitment?" *Ethics*, 199. 2009
- Code, Lorraine. "Skepticism and the Lure of Ambiguity." *Hypatia*. Vol. 21, No. 3, 2006.
- Cavell, Stanley. *A Pitch of Philosophy*. Harvard University Press, 1996.
- _____. *The Claim of Reason*. Oxford University Press, 1999.
- _____. *Must We Mean What We Say*. Cambridge University Press, 2002.
- Crary, Alice. *Beyond Moral Judgment*. Harvard University Press, 2007.
- _____. "A Brilliant Perspective: Diamondian Ethics." *Philosophical Investigations*. Vol. 34, No. 4. 2011.
- Darwall, Stephen. "Self-Deception, Autonomy, and Moral Constitution" in *Philosophical Perspectives on Self-Deception*. Eds. B. P. McLaughlin and A. O. Rorty. University of California Press, 1988.
- _____. *The Second-Person Standpoint: Morality, Respect and Accountability*. Harvard University Press, 2006.
- Diamond, Cora. "The Difficulty of Reality and the Difficulty of Philosophy." *Journal of Literature and the History of Ideas*. Vol. 1, No. 2. 2003.
- _____. "Murdoch the Explorer." *Philosophical Topics*. Vol. 38, No. 1. 2010.
- Doris, John M. *Talking to Ourselves: Reflection, Ignorance, and Agency*. Oxford University Press, 2015.

- Dreyfus, Hubert L. "Overcoming The Myth of the Mental: How Philosophers Can Profit from the Phenomenology of Everyday Expertise." *APA Pacific Presidential Address*. 2005.
- Fanon, Frantz. *Black Skin White Masks*. Grove Press, 2008.
- Finkelstein, David. *Expression and the Inner*. Cambridge: Harvard University Press, 2003.
- Ford, Anton. "The Province of Human Agency." *Noûs*. Vol. 50, No. 4. 2016.
- Frankfurt, Harry. "Identification and Externality" in *The Importance of What We Care About*. Cambridge University Press, 2005.
- _____. "Identification and Wholeheartedness" in *The Importance of What We Care About*. Cambridge University Press, 2005.
- _____. *Necessity, Volition, and Love*. Cambridge University Press, 1998.
- Freud, Sigmund. *Civilization and its Discontents*. Trans. J. Strachey. W.W. Norton, 1962.
- _____. *An Autobiographical Study*. Trans. J. Strachey. W.W. Norton, 1963.
- _____. "Instincts and their Vicissitudes." Trans. J. Strachey. In *General Metapsychological Theory*. Ed. P. Rieff. Simon and Schuster, 1991.
- _____. *Three Essays on the Theory of Sexuality*. Trans. J. Strachey. Basic Books, 2000.
- _____. *Five Lectures on Psychoanalysis*. BN Publishing, 2008.
- Gaita, Raimond. *A Common Humanity: Thinking About Love and Truth and Justice*. Routledge, 2002.
- Gardner, Sebastian. "Critical Notice of Richard Moran, *Authority and Estrangement: An Essay on Self-Knowledge*." *Philosophical Review*, Vol. 113, No. 2. 2004.
- _____. *Irrationality and the Philosophy of Psychoanalysis*. Cambridge University Press, 2007.
- _____. "Nietzsche, the Self, and the Disunity of Philosophical Reason" in *Nietzsche on Freedom and Autonomy*. Eds. Ken Gemes and Simon May. Oxford University Press, 2011.
- Geuss, Raymond. *Politics and the Imagination*. Princeton University Press, 2009.
- Gould, Timothy. "The Names of Action," in *Stanley Cavell*. Ed. Richard Eldridge. Cambridge University Press, 2003.
- Herman, Barbara. *The Practice of Moral Judgment*. Harvard University Press, 1996.

Horowitz, Gregg. Review of Jonathan Lear's *Therapeutic Action*. Notre Dame Philosophical Reviews. <https://ndpr.nd.edu/news/23842-therapeutic-action-an-earnest-plea-for-irony/>

Jaeggi, Rahel. *Alienation*. Columbia University Press, 2014.

Kant, Immanuel. *Critique of Judgment*, trans. Werner S. Pluhar. Hackett, 1987.

_____. *Critique of Practical Reason*. in *Practical Philosophy*. Trans. Mary Gregor. Cambridge University Press, 1999.

_____. *Critique of Pure Reason*, trans. Norman Kemp Smith. Palgrave MacMillian, 2003.

Kekes, John. "Doubts About Autonomy." *Philosophy*. Vol. 86, No. 337. 2011.

Khurana, Thomas. "Three Varieties of Naturalism in Nietzsche's Moral Philosophy." *Forthcoming*.

Korsgaard, Christine M. *The Sources of Normativity*. Cambridge University Press, 1996.

_____. "Creating the Kingdom of Ends: Reciprocity and responsibility in personal relations" in *Creating the Kingdom of Ends*. Cambridge University Press, 1996b.

_____. *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press, 2009.

_____. "Self-Constitution and Irony" in J. Lear *A Case for Irony*. Harvard University Press, 2011.

Krause, Sharon R. *Freedom Beyond Sovereignty*. University of Chicago Press, 2015.

Laplanche, Jean. *Life and Death in Psychoanalysis*. Trans. J. Mehlman. Johns Hopkins University Press, 1985.

Le Carré. *Single and Single*. Pocket Books, 1999.

Lear, Jonathan. *Open Minded: Working Out the Logic of the Soul*. Harvard University Press, 1999.

_____. *Love and Its Place in Nature: A Philosophical Interpretation of Freudian Psychoanalysis*. Yale University Press, 1999.

_____. *Happiness, Death, and the Remainder of Life*. Harvard University Press, 2000.

_____. "Avowal and Unfreedom." Rev. of *Authority and Estrangement: An Essay on Self Knowledge*. *Philosophy and Phenomenological Research*. Vol. 69, No. 2. 2004.

_____. *A Case for Irony*. Harvard University Press, 2011.

- Lawlor, Krista. "Knowing What One Wants." *Philosophy and Phenomenological Research*. Vol. 79, No. 1. 2009.
- Leiter, Brian. *Nietzsche on Morality*. Routledge Press, 2002.
- Loewald Hans. "Ego and Reality." *International Journal of Psycho-Analysis* 32,10–18. 1951.
- _____. *The Essential Loewald*. University Publishing Group, 2000.
- Longuenesse, Béatrice. "Kant's 'I' in 'I Ought To' and Freud's Superego." *Proceedings of the Aristotelian Society Supplementary Volume LXXXVI*, 2012.
- Lovibond, Sabina. *Ethical Formation*. Harvard University Press, 2002.
- _____. "Practical Reason and its Animal Precursors" in *John McDowell: Experience, Norm, and Nature*. Ed. J. Lindgaard. Blackwell, 2008.
- Lugones, Maria. "Playfulness, 'World'-Travelling and Loving Perception." *Hypatia*. Vol. 2, No. 2. 1987
- Mason, Elinor. "Rationality and Morality: Thoughts on Unprincipled Virtue." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*. Vol. 134, No. 3. 2007.
- McDowell, John. *Mind and World*. Harvard University Press, 1996.
- _____. *Mind, Value, and Reality*. Harvard University Press, 1998.
- _____. *The Engaged Intellect*. Harvard University Press, 2009.
- McGeer, Victoria. "The Moral Development of First-Person Authority." *European Journal of Philosophy*. Vol. 16, No. 1. 2007.
- Menke, Christoph. *Force: A Fundamental Concept of Aesthetic Anthropology*. Trans. G. Jackson. Fordham University Press, 2013.
- Moran, Richard. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press, 2001.
- _____. "Cavell on Outsiders and Others." *Revue internationale de philosophie*. No. 256. 2011/2.
- Müller, Anselm. "Backward-Looking Rationality" in *Essays on Anscombe's Intention*. Eds. A. Ford, J. Hornsby, and F. Stoutland. Harvard University Press, 2011.
- Murdoch, Iris. *The Sovereignty of Good*. Routledge, 1991.
- _____. *Existentialists and Mystics*. Penguin Books, 1999.

- Nagel, Thomas. "Moral Luck." *Mortal Questions*, Cambridge University Press, 1991.
- _____. *The View From Nowhere*. Oxford University Press, 1989.
- Nietzsche, Friedrich. *The Gay Science*. Trans. J. Nauckhoff. Cambridge University Press, 2001.
- Okrent, Mark. "Heidegger and Korsgaard on Human Reflection." *Philosophical Topics*. Vol. 27, No. 2. 1999
- Paul, L.A. *Transformative Experience*. Oxford University Press, 2014.
- Pippin, Robert. "Can There Be 'Unprincipled Virtue'? Comments on Nomy Arpaly." Rev. of *Unprincipled Virtue* by Nomy Arpaly. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, Vol. 10, No. 3. 2007.
- _____. *Hegel's Practical Philosophy: Rational Agency as Ethical Life*. Cambridge University Press, 2008.
- _____. *Nietzsche, Psychology, and First Philosophy*. University of Chicago Press, 2010.
- _____. *Fatalism in American Film Noir: Some Cinematic Philosophy*. University of Virginia Press, 2012.
- _____. "Psychology Degree Zero? The Representation of Action in the Films of the Dardenne Brothers." *Critical Inquiry* 41. 2015.
- _____. "Participants and Spectators." Position Paper, On the Human: A Project of the National Humanities Center.
- Poltera, Jacqui. "Is ambivalence an agential vice?" *Philosophical Explorations*. Vol. 13, No. 3. 2010.
- Russell, Francey. "Unity and Synthesis in the Ego Ideal: Reading Freud's Concept through Kant's Philosophy." *American Imago*. Vol. 69, No. 3. 2012.
- Santner, Eric. *On the Psychotheology of Everyday Life*. University of Chicago Press, 2001.
- _____. *The Weight of All Flesh: On the Subject-Matter of Political Economy*. Oxford University Press, 2015.
- Schechtman, Marya. "Making Ourselves Whole: Wholeheartedness, Narrative, and Agency." *Ethical Perspectives*. Vol. 21, No. 2. 2014.
- Sher, George. *Who Knew? Responsibility Without Awareness*. Oxford University Press, 2009.

- Smith, Angela. "Conflicting Attitudes, Moral Agency, and Conceptions of the Self." *Philosophical Topics*. Vol. 32, No. 1/2. 2004.
- _____. Rev. of *Merit, Meaning, and Human Bondage: An Essay on Free Will* by Nomy Arpaly. *Notre Dame Philosophical Reviews*, January 5, 2007.
- Sommers, Tamler. "The Objective Attitude." *The Philosophical Quarterly*. Vol. 57, No. 228. 2007.
- Stroud, Sarah. "Moral Worth and Rationality as Acting on Good Reasons." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*. Vol. 134, No. 3. 2007.
- Taylor, Charles. *Philosophical Papers Volume 1: Human Agency and Language*. Cambridge University Press, 1985.
- _____. *Philosophical Papers Volume 2: Philosophy and the Human Sciences*. Cambridge University Press, 1985.
- Velleman, David J. *Self to Self*. Cambridge University Press, 2005.
- Vogler, Candace. "Sex and Talk." *Critical Inquiry*. Vol. 24, No. 2. 1998.
- _____. "Rev. of *Unprincipled Virtue: An Inquiry into Moral Agency*, by Nomy Arpaly." *The Journal of Philosophy*. Vol. 103, No. 9. 2006.
- Walker, Margaret Urban. "Moral Luck and Impure Agency." *Metaphilosophy*. Vol. 22, Nos. 1 and 2. 1991.
- Wallace, R. Jay. *Normativity and the Will*. Oxford University Press, 2006.
- Watson, Gary. "Two Faces of Responsibility", *Philosophical Topics*. Vol. 24, No. 2. 1996.
- Williams, Bernard. *Moral Luck*. Cambridge University Press, 1999.
- _____. *Shame and Necessity*. University of California Press, 2008.
- Wittgenstein, Ludwig. *Philosophical Investigations*. Trans. G.E.M. Anscombe. Blackwell Publishing, 2001.
- Woolf, Virginia. *Orlando: A Biography*. Mariner Books, 1973.
- Zupancic, Alenka. *Why Psychoanalysis?* NSU Books, 2008.