THE UNIVERSITY OF CHICAGO

THE CONTINGENCY OF MORAL PERSONHOOD

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

EMILY DUPREE

CHICAGO, ILLINOIS

AUGUST 2021

TABLE OF CONTENTS

ACKNOWLEDGMENTS

My experience working with my committee has been so wonderful – their feedback has been deeply challenging, their support so inspiring, their co-curiosity so appreciated. Agnes Callard, my chair, has the habit of asking lengthy series of questions that, from multiple angles, challenge the very core assumptions of my work. Feedback conveyed to me at breakneck speed in meetings often takes me weeks piece together, and developing satisfying responses to these worries has been one of the most pleasurable aspects of the later stages of this dissertation. Matthias Haase has the unique ability to turn my work upside down and convince me that what I'm really trying to say is much better understood from this angle. His philosophical vision is bird's eye and expansive; with him on my team, all ideas can be integrated into an elegant and systemic whole. Sally Haslanger has been a champion of my work and my commitment to the close alliance between the personal and the philosophical, between style and substance. Conversations with them feel like the hard conversations one has with someone navigating the same difficult terrain – in this sense I have felt her presence on my committee to be both uplifting and grounding. I would like to thank each of these three philosophers for the care and attention shown to me and to my work – it has truly buoyed me in these last years, especially during the COVID-19 pandemic, and given me a sense that I have a team of people on my side.

My time at the University of Chicago has been made better and more rewarding through participation in a number of student groups, including Women in Philosophy, Minorities and Philosophy, the Diversity and Philosophy Working Group, and Graduate Students United. These are the groups working to counteract the embedded hierarchies

of academia that too many are eager to reproduce, and I owe them my thanks. My work on this dissertation specifically has been made easier by Ali Erani's invaluable research assistance on Chapter 3 and Kristy Lueshen's expert assistance with the intricacies of the Chicago Manual of Style. More generally, I am grateful to my students, my like-minded peers, the supportive faculty, the Philosophy Department administrators, and all of the support-staff on campus who have kept the infrastructure of academia up and running so that my own professional development could progress unhindered.

I would also like to thank many of my dear friends and co-thinkers in this world whose presence gives meaning and joy to my life. To Karina Bull, part of the fabric of my life, who exemplifies what it means to be there for someone in a time of need. To Shana Crandell, there from the beginning, whose relentless questions and ability to talk about Hegel until it's way past bedtime inspire me to ask harder questions, think more deeply, and connect this depth of thought with the urgency of solving our broadest philosophical issues. Conversations with you are an antidote to the dull pallor of professionalism that academia can cast on us all. To Melina Garibovic, resident of the Hyde Park philosophy clubhouse and a reliable interlocutor when I am craving sweeping conversations that seamlessly traverse aesthetics, ethics, and the limits of interpersonal knowledge. To Claudia Hogg-Blake, who became my friend in a time of great suffering, and whose growth fueled my own. Our unspooling conversations in the garden weave the therapeutic with the philosophical and the absurd, and my work contains meditations on the texture of life in part because yours does so beautifully. To Francey Russell, with whom I have expansive and moving and world-building conversations – ones that get underneath us and make me feel vivid and urgent. Your Chicago companionship, long-

distance tenderness, and absolutely singular view on the hidden contours of our shared, human lives are cherished beyond measure. To Bella Week, my chosen sibling, my best friend. And to Dustin Currier, my companion in life with whom I have grown in ways that reveal the absolute beauty of what's possible while we're here – a beauty that I am privileged to create and to witness alongside you. Your unyielding inquiries into my work, complemented so perfectly with epistemic humility and a courageously open mind, have pushed my thinking further and required me to make my work more honest, more rigorous, and more grounded in the nuance of what is. To my family, wellsprings of support and familiarity when I've needed it most. To Brian, who did not make it long enough to read this. To my care team, Abby, Amy, and Heather. And to me, for persevering and resting with equal measure.

INTRODUCTION

GETTING THE CONTINGENCY OF MORAL PERSONHOOD IN VIEW

My fantasies about killing the people who have wronged me involve my entire body – I would never be satisfied by a remote assassination, a fortuitous heart attack, or a slip off of the California cliffs. No, part of the fantasy is that *I* am the one doing it. My hands wrapped around their necks, my fists repeatedly bashing in their face. I imagine this to require some level of exertion, breathing heavily, being overcome with delicious rage, finally directed at the right person. In this fantasy, violence flows through me uncontrolled, and it is precisely the unbridled nature of it, my finally allowing it to be unbridled, that intoxicates me. I imagine the fervor reaching its crescendo at the precise moment that my transgressor loses his life. And then I imagine myself sitting there, disgusted by the patheticness of death, and swelling with an until-then dormant sense of my own freedom.

*

Squaring these fantasies with my research as a moral and political philosopher has never been an easy task. Revenge is taken to be so obviously morally wrong that it serves as an argumentative backstop for theories gone awry – if your view ends up condoning revenge, you have reached the *reductio ad absurdum*. It is the jumping off point for the imperative of civil society, the brutish behavior that any rational beings can and indeed must conspire to eradicate if they are to live in a state worthy of their kind. It is thought of as the worst excess of an already-suspicious emotion, a behavior that can single-handedly mark one as acting heteronomously. The desire for revenge, then, is taken for granted to be pathological. The possibility that this might also be accompanied by pleasure is almost too much for the moral philosophers to bear.

And yet the fantasies persist. Not only that, but also the sense that these fantasies are not altogether misguided. I am not alone in thinking that there is something beyond catharsis that I might get from carrying out this plan, that *freedom* might be the reward for a well-chosen revenge plot. But how could a robust sense of freedom (and not merely a *freedom from* some haunting externality) be the outcome of such abhorrent and violent interpersonal conduct? Isn't bodily autonomy, after all, the bedrock of this freedom?

One of my mother's earliest childhood memories is of her sitting on her grandmother Ester's sofa in 1960's Brooklyn, her small legs sticking to the plastic sofa-covering that apparently all Jews of that era had. The family photo album had been brought out, and Grandma Ester was narrating the turning of the pages, putting names to faces, and fates, one by one:

"Alter – killed by the Nazis.

_____ – killed by the Nazis.

_____ – killed by the Nazis.

_____ – killed by the Nazis."[1]

They went through the entire photo album like this, a procession of the trauma that my mother was deemed old enough to learn was her inheritance. Not all of the relatives were killed by Nazis, as some lucky few escaped Poland before the borders closed. I myself owe my life to my second great-aunt Becky's escape from Poland at the turn of the century, when she came through Ellis Island as so many other refugees of her era did. But she was in my family's minority, and the vast majority of her siblings, cousins, aunts and uncles, parents and grandparents were killed in concentration camps. Whomever Becky could not send money to in time, was murdered. No wonder there was rage in Ester's voice as she flipped through the photo album, a sepia procession of loved ones whose fate was so different from hers only because her older sister Becky ran out of time and could not choose everyone. One by one, and then all at once, my family suffered the same fate as all of the other Jews across Europe, what Hannah Arendt described as "monstrosities no one believed possible at the beginning."[2]

Arendt also described the Holocaust as consisting of "the total collapse of normal moral standards."[3] In addition to the totalitarian inversion of what is typically considered "legal" and "criminal," citizens across Europe were faced with the task of judging (or failing to judge) this regime in a moral landscape that defied application of the ordinarily

[1] Neither my mother nor my grandfather Bill, Ester's son, can remember the names of these murdered relatives beyond Alter's. As Bill remarked, upon realizing he had forgotten: "They're lost in the mist."
[2] Hannah Arendt, "Personal Responsibility Under Dictatorship" in *Responsibility and Judgment*, ed. Jerome Kohn (New York: Schocken Books, 2003), 23.
[3] Arendt, "Personal Responsibility," 34.

reliable moral standards on which we depend. From this dizzying and unhinged location, they had to choose: do I comply, or do I resist? Is there a way to do nothing? They were, as Arendt put it, "between the devil and the deep sea."[4] And yet they somehow made their choices. We know how these choices turned out – most chose to comply, a vast apparatus of death and destruction was activated, and terror swept the continent not only via armies and military weaponry but also through the ordinary decisions of neighbors and colleagues who were once considered friends.

Making sense of the ethics of what happened within this "total collapse" is no easy feat. Even the seemingly straightforward task of judging those *responsible* for the Holocaust was rife with moments of extraordinary difficulty. The complications of this historic task are well-known: we could not have put every German citizen on trial, especially when they were making choices from within a legal and moral fog that might have led us to make the same grisly choices, reprehensible as they were. There were difficult questions of responsibility within bureaucratic organization, collective guilt, and the imperative of having *some* citizenry to continue on the German project. Indeed the very idea of responsibility strained under the weight of the world's attempt to confine it to the bounds of legal accountability.

More difficult still is judging those who actively fought back – how does one judge those who fought for an eminently worthy cause, but with methods that are forbidden by those "ordinary moral standards" that collapsed during the Holocaust? This is another very difficult ethical question. But the complications of answering this question are of a different sort than the one above. They do not stem from procedural

[4] Arendt, "Personal Responsibility," 25.

limits or the difficulty of an accountability project with the sheer magnitude of for example the Nuremberg Trials. Rather, complications surrounding the resistance have to do with the propriety of judgment at all. It is not: *who are we to judge?* – but rather: *what good is judgment at all?*

Nowhere is this more acute than when considering the Jewish resistance movements focused on revenge. Unlike many of the other resistance movements for which an exculpatory case for self-defense is plausible (even in those seemingly "preemptive" strikes which became more common near the war's end), the revenge movements continued on even after the war was over. They seemed to be the opposite of the Nuremberg Trials, with its atemporal legality, its orderliness, its publicity. Instead, Jewish revenge plots were carried out in secret, outside of the law, and according to a moral code forged in the fires of war.

When asked why he embarked on a revenge plot to kill as many Nazis as possible after the war, Jewish resistance fighter Abba Kovner said the following:

> The destruction was not around us. It was within us. We did not imagine we could return to life, that we had a right to have families, to get up and go to work as if accounts with the Germans had been settled.[5]

What did he mean by this? A natural interpretation of Kovner's sentiment is that the revenge plots were meant to settle these very scores – to avenge the deaths of those who did not survive. After all, the Jews who made it through the Holocaust were often the only surviving members of their families, the walking dead dragging unbearable shadows behind them. They could not ethically rebuild their lives, this interpretation goes, until

[5] Rich Cohen, *The Avengers: A Jewish War Story* (New York: A.A. Knopf), 191.

they had gotten some sort of justice on behalf of those from whom this chance was forever stolen. The vengeance was in some sense how the survivors earned from those who were killed the permission to continue on with their own lives.

Surely there is some truth to this interpretation. But there is also truth to another interpretation of Kovner's statement, one that has to do not with permission but with possibility. "The destruction…was within us." On this alternate reading, revenge was a way of responding to their own destruction – it was a way of settling a score not on behalf of others but on behalf of oneself. The Jews who took revenge did so as an act of self-reconstitution. Killing Nazis became the activity that mended their internal destruction; it was a necessary step towards the possibility of rebuilding the fractured personhood that barely made it out of the Holocaust alive. Revenge pulled them out of the abyss of nothingness and made it possible to embark on a life that had the chance of being fully human again.

The ease or difficulty of judging Jewish revenge depends, in part, on which interpretation you believe. If it is about settling scores, about taking matters into one's own hands in order to seek permission to go on, then it is easy to see what a moralistic perspective would yield: they should have left it up to the state. Or, perhaps, one would conclude that their vengeance was justified prior to the war's end insofar as it constituted self-defense, but that anything after was beyond the morally permissible. Whatever the conclusion, following this interpretation gets us into the territory of the ethics of the oppressed, with all of its paternalism and insistence on moral clarity. Worse, this insistence on moral clarity obscures the much more interesting moral phenomenology of

revenge itself – by coming to a condemnatory moral conclusion, it is easy to forget that there might be something gained in the act nonetheless.

If on the other hand the revenge plots were not about settling scores but instead about moral emancipation, about making oneself out of nothing, it is harder to maintain a fixed moral perspective on the activities. After all, the Jews themselves claimed that they were destroyed; surely the moral law does not apply to them in the same way (or at least applying it straightforwardly here is the least interesting thing one can do with it). They also claimed that they were seeking freedom, at last; surely we have some right to freedom, even where the path that leads to it is not immune from criticism. The point is that this latter interpretation of what it means to say what Kovner said opens up a space to consider the moral phenomenology of revenge, and perhaps the moral good of it, without the encumbrance of moral judgment.

This moral curiosity is what leads me to pause on revenge itself as a robust moral phenomenon worthy of philosophical attention beyond the condemnatory. What it has revealed is what this dissertation is about: not only what revenge takers take themselves to be doing, but also an entire ethical world in which that moral vision is made coherent. Indeed, the focus of the project has shifted to foreground this ethical world, offering a conception of the person that adequately captures the interpersonal nature of moral actualization; the true extent of oppression's harms; the requirements of justice for people whose moral personhood is contingent on their interpersonal environment; and finally the moral good of revenge and its relationship to moral personhood. Considering revenge from the perspective of those who had the greatest reason to pursue it reveals that it is at the very heart of an alternate moral framework, one which when viewed as a

whole is quite different from the moral frameworks with which we are familiar in moral philosophy.

The dissertation, then, begins with the moral framework and ends in revenge. Though its explanatory order is the reverse of its discovery as presented in this introduction, the concepts established early on are essential to understanding the full moral phenomenology of revenge. Each chapter builds on the last.

Chapter 1 is concerned with two main questions: first, what conceptions of moral personhood are implicit in first-personal accounts of various forms of oppression? Second, which more formal articulations of moral personhood can adequately account for and cohere with these accounts? I argue that first-personal accounts of oppression reveal that it has the capacity to turn its victims into *objects*, and that this is a moral metaphysical transformation and not merely a distortion in its victims' moral self-conception. I call this possibility *the contingency of moral personhood*, and argue that it is a necessary feature of the human condition revealed by considering oppression's effects on its victims. Oppression reveals this possibility through the operation of its "five faces" in the term coined by Iris Marion Young, and in particular through the role of traumatic violence that weaves throughout them all. Considering the moral psychology of trauma, then, is essential to understanding the experience of oppression's harms. From within this moral psychology, we find accounts of moral transformation that can only be explained by an account of moral personhood that regards the interpersonal as an essential element. Further, we have independent reasons for trying to make sense of these accounts on their own terms (rather than attempting to explain them away with error theory) – their origins in the necessity of survival give them an epistemic advantage

over alternate conceptions of moral personhood. In particular, being targeted by the forces of oppression requires one to see through oppression's ideology toward a veridical understanding of one's moral situation if one is to survive. An interpersonal account of moral personhood, then, is a stronger conception than alternatives, because it can account for the moral metaphysical insights that we have independent reasons to take philosophically seriously. Finally, I explore one way that oppressed individuals can overcome their status as an object without having to rely on interpersonal recognition for that moral emancipation. In this way, we can maintain our claim about the contingency of moral personhood without being committed to the idea that the oppressed are doomed to moral metaphysical insignificance if their social and political conditions fail to change. In other words, we can have a conception of moral agency under oppression without giving up on a conception of moral personhood that is consistent with the moral psychology of having that agency diminished under conditions of oppression.

Chapter 2 explores the political philosophical consequences of these moral insights. The forms of oppression discussed in Chapter 1 are paradigm examples of injustice; as a result, theories of justice will fall short if they cannot adequately explain oppression and prevent it from occurring in the sociopolitical organization meant to be regulated by the theories' principles. Each theory, moreover, will be connected to a certain conception of the person from which these principles are derived – it is what ensures that the principles of justice are *for* the relevant types of beings meant to be regulated by them. So the extent to which a theory of justice can account for injustice, including oppression, will be related to the extent to which its conception of the person adheres to the features of people that make them vulnerable to all sorts of *in*justice. What

I argue in Chapter 2 is that the contingency of moral personhood, defended in Chapter 1, is an essential vulnerability that political theories must incorporate into their conceptions of the person; to the extent that they do not, then, the theories fall short of the ideal of justice. I take a close look at two theories of justice in particular, i.e. John Rawls's "justice as fairness" and Martha Nussbaum's capabilities approach, in order to reveal both that neither view incorporates the contingency of moral personhood into their conceptions of the person, *and* that this omission is especially egregious in Nussbaum's account – hers is presented as an oppression-sensitive correction to the Rawlsian view, and yet it contains the same theoretical omission that prevents a theory of justice from fully accounting for the injustice of oppression. Taking the Rawls-Nussbaum debate to be indicative of a larger debate within political philosophy, we can see that the inviolability conception of moral personhood haunts political philosophy more generally, even where it is especially concerned with getting it right about oppression's deepest impacts on the person.

Finally, Chapter 3 confronts the following question: what are victims of oppression to do when the unjust conditions continue? When they cannot count on the fully just state to safeguard or reinstate their moral personhood, but they also cannot in some sense continue on with a diminished moral personhood? Revenge is one avenue that victims of oppression choose; Chapter 3 is about making sense of this choice. Contrary to views that regard revenge as necessarily irrational, as something involving only costs and no benefits, I argue that revenge's successful pursuit under failed political conditions is rational insofar as its targeted, interpersonal nature can restore the moral personhood of the revenge-seeker. To argue for this, I take a close look at a number of

theories on anger and revenge as well as historical examples of revenge plots carried out during and in the aftermath of the Holocaust. What these reveal is that revenge does indeed have the capacity for moral emancipation: it requires that the revenge-seeker regard one's own nascent moral personhood as a reason for action, and when that action is carried out it forges an interpersonal moral connection that was severed by the prior revenge-worthy conduct. Where Chapter 1 explored ways the interpersonal nature of moral personhood makes us liable to the deepest of harms, Chapter 3 shows that it also offers those harmed a way out toward freedom.

Though the dissertation as a whole does not make much use of the concept of freedom, there is a sense in which concern with it runs throughout, just off the page. There is an unfurling of the person that is possible under the right conditions, a transformation into and confirmation of the types of beings we are, by nature, able to become. The dissertation, though, is less about articulating a robust account of this moral existence than it is about articulating our movements towards and away from it – our transit in a moral-ontological landscape. There is a lot to say about becoming or failing to become a moral being; about the role others play in this transitory moral existence; about how we can step in and be moral *causa sui* when all else fails. This is the liminal space of moral personhood, its rich and jagged edges, and it is from here I believe we can understand the nature of freedom and of ourselves most clearly.

CHAPTER 1
THE CONTINGENCY OF MORAL PERSONHOOD

> "…I transported myself on that particular day far, very far, from my self, and gave myself up as an object. What did this mean to me? Peeling, stripping my skin, causing a hemorrhage that left congealed black blood all over my body. Yet this reconsideration of myself, this thematization, was not my idea. I wanted quite simply to be a man among men. I would have liked to enter our world young and sleek, a world we could build together."
>
> – Frantz Fanon, *Black Skin White Masks*

> "I know what I'm talking about because I'm talking about myself."
>
> – Ruth Shay

1. Framing

When oppression does not altogether silence its victims' stories, we find accounts of living through it that portray this experience as a sort of living death. People becoming objects; bodies becoming things – these are the details that emerge. Fanon's description of living under white colonialism is no different, and his work is filled with visceral descriptions of bodies being flayed for the sake of others, agential consciousnesses drifting away and witnessing the shriveling of once-vibrant personalities. Or, more often, he describes the inhibition of the development of vibrant personalities from the start, and in their places the development of selves and self-understandings that fully integrate the power dynamics driving the engine of oppression. White colonialism,

on Fanon's account, turns Black life into a merely instrumental existence. Indeed part of the beauty of his work, in addition to its significant theoretical insight, is its expression of the pain of this moral transformation.

First-personal accounts of living under white colonialism, like Fanon's, are not the only documentations of oppression that foreground the experience of this moral transformation. Many Jews writing after the Holocaust refer to themselves as having died in the concentration camps; memoirs of women navigating the aftermath of patriarchal violence recount shattered selves and an intense dislocation from their alleged moral community. Susan Brison describes the time after she was assaulted with the following language:

> For the first several months after my attack, I led a spectral existence, not quite sure whether I had died and the world went on without me, or whether I was alive but in a totally alien world….I felt as though I'd somehow outlived myself.[1]

Put simply, a study of what it is like to bear the brunt of oppression, whichever axis of oppression it is based on, reveals that it can be an experience of becoming something other than a full moral being. The metaphorical language of death is prevalent, suggesting the end of a certain moral status, though even those who use it point out its conceptual insufficiency. Nevertheless its use captures a liminal space of experience between full moral personhood and nothingness. In that space, where one's consciousness persists but no longer seems to ground oneself in a space of moral belonging, is a rich terrain of moral insight.

[1] Susan J. Brison, *Aftermath* (Princeton, NJ: Princeton University Press, 2011), 9.

The moral observations coming out of these traumatic liminal spaces, however, are not often reflected in our philosophical theories of the person. To be sure, theories of personal identity are catching up to the phenomenon of "shattering" described above, putting pressure on views that rely on continuity and/or coherence of identity for a distinct self.[2] There is a rich literature on the psychological aspect of living under oppression, the "double consciousness" and internalized prejudice that so often accompany such a life.[3] And philosophers have (relatively speaking) recently begun to merge the practical with the epistemic, pointing out how oppressive conditions can stymie the development of concepts that would facilitate the full flourishing of someone's identity and personality.[4] But little has been said, so far, putting these moral observations in conversation with the conceptions of moral *personhood*, rather than personality, driving our ethical theories. This is my task here.

In this chapter, I will discuss two competing accounts of moral personhood and their adequacy in capturing the moral insights that come out of oppressive conditions. The fundamental difference between these two accounts is whether they take a person's social and political life to be relevant to their moral status – whether interpersonal treatment is related to moral ontology. To explore this question, I will inquire into the role of trauma in people's experience of oppression as told from their perspective. And I will argue that that these narratives both rely on and reveal a conception of moral personhood that requires us to take the interpersonal seriously. This argument will

---

[2] See, e.g., Susan J. Brison, "Outliving Oneself: Trauma, Memory, and Personal Identity," in *Gender Struggles: Practical Approaches to Contemporary Feminism*, eds. Constance L. Mui and Julien S. Murphy (Lanham, Maryland: Rowman & Littlefield Publishers, Inc., 2002).

[3] See W.E.B. Du Bois, *The Souls of Black Folk: Essays and Sketches* (Chicago, A.G. McClurg, 1903. New York: Johnson Reprint Corp, 1968).

[4] See Miranda Fricker, "Hermeneutical Injustice," in *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford: Oxford University Press, 2007).

receive support from considerations about the epistemic authority of the perspective from which these narratives emerge. Together, these arguments support what I am calling *the contingency of moral personhood*: reflecting on the role of the interpersonal in moral ontology reveals that there is a contingency to the actualization and/or endurance of our moral personhood, and this contingency is a necessary feature of the human condition. Note the claim is not that our *goodness* is subject to circumstances not wholly in our control – all of the philosophy of tragedy and moral luck has, to some degree, this worry in mind.[5] The claim is one step prior: that contingency touches not only the content of one's moral character but also, prior, the actualization of one's status as a moral being at all. That this deeper contingency does not have to do with the content of one's moral character does not mean that it has nothing to do with moral critique. In fact, a central component of the view is that the stifling of the development or full expression of one's moral personhood is one of, if not *the*, most objectionable forms of moral harm possible. Thus these arguments will simultaneously present a novel account of moral personhood and a novel understanding of the tremendous harms that can result from political conditions that fail to live up to its requirements. This account, I argue, is the only conception of moral personhood that can adequately make sense of the insights that arise from within the immense political failure that is oppression.

[5] The contemporary view closest to my own is in Lisa Tessman's *Burdened Virtues: Virtue Ethics for Liberatory Struggles* (Oxford: Oxford University Press, 2005), in which she argues that oppression burdens moral goodness. But even there, she focuses on the virtues themselves rather than the moral status that grounds them.

2. The Dialectic

Let us begin with the two competing accounts of moral personhood. In the course of this chapter, I argue that first personal accounts of living under oppression both reveal and are best understood through a two-part conception of moral personhood: one on which both an individual and an interpersonal feature must be met for someone to count as a full moral person.[6] Let's call this the *interpersonal* conception of moral personhood. It can be understood as follows:

Some feature (of the **individual**	+	Some feature of that individual's **interpersonal** context)	⟷	**Full moral personhood**

What this means is that, on this view, neither element alone is sufficient for "counting" as a full moral person, though each is necessary. This is not a radical claim as it relates to the interpersonal element: we all agree that it is not sufficient that we treat e.g. a stapler as a moral being for its actually *being* one. Something else has to be the case, however much we want our stapler to be a member of the moral community.

The radical element of the view is rather that it is not sufficient that an individual has the morally relevant feature, whatever it may be, for them to count as a full moral being. Why is this a radical position? The individual element of moral personhood should be familiar to each of us: this is the property that, for example, philosophers of equality are searching for in trying to delimit the bounds of moral consideration. It is what separates us from the stapler – what makes it the case that throwing me across the

[6] I mean "feature" in the most theoretically-neutral way possible. If the reader is concerned with this phrasing, they are welcome to substitute "aspect," "element," or any other preferred neutral term.

room is morally different than throwing the stapler. Kant famously argues that the morally relevant individual feature is our practical rationality; without it, we are not members of the kingdom of ends.[7] But he also argues that there is no *other* morally relevant consideration, at least for purposes of determining what kind of thing one is. Let us call this alternate conception of moral personhood the *individual* view.

Some feature of the **individual**	⟷	**Full moral personhood**

It is a view on which answering the question, "Is this a moral person?" only requires one to look at the entity in question and find out whether it has a certain feature. This could be practical rationality; it could be the capacity for pleasure or pain; it could be an endless array of more and less plausible candidates for moral personhood. The point, though, is that the question of the social does not enter into the picture. Moral beings are the kinds of things they are in virtue of facts about only *them*, understood apart from their social context. Moreover, this focus on only the individual for grounding moral personhood is what grounds the individual view's ability to critique interpersonal contexts on moral grounds. It is *because* full moral personhood resides only in the individual that we can say, on this view, that social and political arrangements are unjust where they treat the individual as if they were not a full moral person. This conceptual architecture maintains a gap between the treatment someone deserves and the treatment they receive, and within this gap is born the concept of right. Thomas Nagel captures this well when he says: "If [someone] has [the status of inviolability], he does not lose it when his rights are

[7] Immanuel Kant, *Groundwork of the Metaphysics of Morals* (London, England: Oxford University Press, 2012).

violated – rather, such treatment counts as a violation of his rights precisely because he has it."[8] This is a substantial upshot of the individual view – it allows us to easily explain the moral impermissibility of certain actions by reference to something stable and unchanging in the individual who is harmed.

The interpersonal view is radical insofar as it rejects these claims. Specifically, it rejects the claim that moral status is determined only by features of the individual. It therefore also rejects the claim that the injustice of social and political conditions like oppression must be explained by their dissonance with what its victims are in fact entitled to. Instead, the interpersonal view holds that having the relevant individual feature is a necessary precondition for full moral personhood that nonetheless requires it to be taken up in interpersonal life in a certain way for that moral personhood to become or remain actualized. Having one, but not the other, is not enough to count as a full member of the moral community. Furthermore, the injustice of oppression will be explained in reference to a stymied *potential* rather than (or in certain cases, in addition to) a violated actuality.[9] That is, oppression's harms consist not (only) in violating its victims' actual moral entitlements but rather in undermining the development of moral selves that would have those entitlements were they allowed to flourish in the right way. Where the individual view can only explain the *moral* harm of oppression, the interpersonal view can also explain the *moral metaphysical* harm. Individuals fail to become the kind of being

[8] Thomas Nagel, "The Value of Inviolability," in *Morality and Self-Interest*, ed. Paul Bloomfield (Oxford: Oxford University Press, 2011), 107.

[9] The parenthetical remarks in this sentence and the next are meant to account for those circumstances in which oppression emerges after the formation of full moral selves, and therefore the interpersonal element that once secured the actualization of full moral personhood vanishes. In these cases, there is a violation of actual entitlements that can eventually transform oppression's victims into beings with only potential entitlements.

that they have the capacity to become, losing out on all of the moral entitlements that would have emerged along with that actualization.

The interpersonal view is thus radical in a second way. To put a finer point on it, introducing an interpersonal element to the actualization of moral personhood means that our moral personhood is *contingent* on how it is taken up in moral community. This also means that our moral personhood can fail to actualize despite us having the requisite morally relevant individual feature (e.g. rationality) where the social conditions are not right. The above description of the moral metaphysical harm of oppression is an instance of this phenomenon.

Thus the interpersonal view, and the contingency of moral personhood to which it is committed, rejects a core feature of the individual conception of moral personhood: that however our social conditions turn out, it will never touch our core moral personhood. Let us call this *moral inviolability*.[10] Moral contingency and moral inviolability are negations of one another. I argue for the former. I also argue that the latter compounds the harms of oppression insofar as it obscures oppression's deepest effects. Above, I said that the individual element of moral personhood is what distinguishes us from the stapler. The interpersonal element can be thought of as what distinguishes us from the angels – it is the source of our deepest moral vulnerabilities, vulnerabilities related to not just our personalities, characters, our life outcomes and struggles, but also to the very type of beings we are. Unlike the angels, we have interpersonal lives that can

[10] This notion of inviolability can be found, as I am arguing, across a wide array of philosophical views. In some cases, though, it goes by a different name, as in Elizabeth Anscombe's idea of "impregnability," which could stand in for my notion of inviolability. "What do I mean, then, by saying that the value and dignity of a human being is impregnable? I mean that it cannot be taken away." Elizabeth Anscombe, "The Dignity of the Human Being," in eds. Mary Geach and Luke Gormally, *Human Life, Action, and Ethics: Essays by G.E.M. Anscombe* (Exeter, England: Imprint Academic, 2005), 65.

turn out in such a way that we cease to be the kinds of beings we are. This kind of social failure can inhibit or erode our status as moral beings at all.

3. Oppression and Trauma

Oppression is one example of the kind of social failure that can have this moral metaphysical effect. A close consideration of the first-personal perspective of living under oppression reveals this to be the case. What do I mean by the term 'oppression'? Here, I make use of Iris Marion Young's multifaceted account of oppression, while also calling attention to the phenomenon of trauma that weaves throughout the various aspects of oppression that she carefully distinguishes. Bringing in an analysis of trauma is a way of bringing the first-personal perspective in to a discussion that already captures a more impersonal, third-personal perspective quite well. In this way, I am borrowing from her account but also adding to it insofar as I take the possibility of traumatic violence to be constitutive of oppression and integral to its full understanding. Furthermore, where trauma is linked to social and political arrangements (as it is in oppression), it is of special relevance to moral philosophy in ways that more isolated instances of trauma are not.[11] Thus this discussion will allow us to both get a better grip on the harms of oppression *and* see the injustice of it at the same time.

On Young's view, oppression is a social phenomenon that can consist of a cluster of related phenomena: exploitation, marginalization, powerlessness, cultural imperialism, and violence. Though each different axis of oppression manifests differently, making it

[11] By "isolated instances of trauma," I am referring to cases where an individual experiences a traumatic event and the lingering aftermath of that trauma, but the event itself is not linked in any relevant way to an injustice. These could include witnessing a car crash, losing a parent to old age, or experiencing intense and unexpected pain.

the case that there is no singular definition that can encompass all instances, it can generally be described as the state in which "oppressed people suffer some inhibition of their ability to develop and exercise their capacities and express their needs, thoughts, and feelings."[12] Moreover, this inhibition occurs as a result of one's membership in what Young refers to as "social groups," social objects of affiliation and social relations, combined with those groups' standing in relation to other groups. Oppression, on this view, is a structural phenomenon that makes one liable to the external inhibition of one's capacities due to membership in various social groups.

Her taxonomy of oppression provides us with a finer-grained understanding of the different ways that social interaction, mediated by social groups, can impact the person in oppressive ways; for our purposes, it is a taxonomy of the different arenas in which the actualization of moral personhood can be eroded. For example, it is not some diffuse idea of patriarchy writ large that erodes women's moral status, but more specifically the daily violence, exploitation, and powerlessness that together constitute patriarchal oppression. These distinct phenomena work together to create the unjust totality that oppression amounts to. A full understanding of oppression, then, includes attention to the varied interpersonal phenomena that make it up.

Moreover, this full understanding also requires (I am arguing) that we analyze the first-personal experiences of the diverse phenomena making it up. The following is such an analysis, and it reveals that the potential for *trauma* is a through-line of oppression from within which we learn more about oppression's devastating effects on the moral status of its victims. Of course, not all instances of trauma are a result of oppression; it is

[12] Iris Marion Young, *Justice and the Politics of Difference* (Princeton, NJ: Princeton University Press, 1990), 40.

a varied phenomenon with varied causes, many of which are not relevant to the moral and political philosopher (for example, trauma in the wake of a car crash). But some are, and for this reason the moral and political philosopher should be concerned with it – it is a unique sort of moral vulnerability, and if it is caused by *other people* in interpersonal settings that are properly regulated by political principles (e.g. oppression), then the specific contours of the harm are worth exploring. Part of this exploration is, as Young observes, a kind of arm's-length understanding of the various social and political phenomena that can constitute oppression; the remainder of this picture is the more intimate understanding of what such an experience can do to you. Let us analyze the phenomenology of oppression-related trauma, then, to see the full scope of oppression's effects. I believe this perspective, at the jagged edges of moral personhood, is one rich with meaning.

3.1 The Trauma of Oppression

Susan J. Brison's philosophical memoir on the trauma of sexual violence begins with a statement on the immense difficulty of such a project. She writes:

> The challenge of finding language that is true to traumatic experience is, however, a daunting one. How can we speak about the unspeakable without attempting to render it intelligible and sayable?...For months after my assault, I had to stop myself before saying (what seemed accurate at the time), "I was murdered in France last summer."[13]

Of course, Brison was not murdered in France last summer – she was very much alive and recovering from her injuries. The difficulty though, as she puts it, resides in its

[13] Brison, *Aftermath*, xi.

paradoxes: how can the living feel that they've died? How can the "ordinary concepts of time and identity cease to apply"?[14]

> For the first several months after my attack, I led a spectral existence, not sure whether I had died and the world went on without me, or whether I was alive but in a totally alien world…I felt as though I'd somehow outlived myself.[15]

This complete shattering of her sense of self and of her sense of the world around her was characteristic of post-traumatic stress disorder, one of the real possibilities we face in the aftermath of trauma. Insofar as the trauma she endured was of a *political* nature, i.e. was linked to patriarchy and therefore to a system of oppression, it reveals one of the many possible harms that can occur within the framework of oppression. For this reason, it will be helpful to explore the phenomenology of her experience (and of similar oppression-linked trauma narratives). My aim is to reveal one of the potential harms of oppression *from the inside*, before moving on to an analysis of moral conceptions that might be adequate to the task of making sense of it.

Not all traumatic incidents result in post-traumatic stress. But some do, and the experience of it is an immense struggle in its own right, as Brison methodically describes. Its symptoms include intrusive memories and feelings, negative affect, hypervigilance and an exaggerated startle response, a skewed sense of time, and in some cases dissociation, depersonalization, and derealization, or the feeling that oneself or the world is not quite real. Brison recalls that in her case her sense of time and ability to focus were so damaged that she "could no longer imagine how [she] would get through each day, let

[14] Brison, *Aftermath*, xi.
[15] Brison, 9.

alone what [she] might be doing in a year's time."[16] She did not believe that she would ever write or teach philosophy again. This, coupled with all of the other symptoms of post-traumatic stress she exhibited, had a profound effect on her sense of self. And in this respect, Brison's experience with post-traumatic stress was typical – it can add up to an experience of a deeply fractured and disorganized self.

Dissociation, in particular, can contribute to this sense of personal fracturing that accompanies traumatic harm, including the traumatic harm of oppression. There is a sense of unreality to it – a sense that one is merely observing oneself from afar. As it is described in polyvagal theory, dissociation originates in our nervous system's most-distant phylogenetic past, when freezing in the face of insurmountable danger, and somehow separating ourselves from the being in danger, was our strategy of last resort.[17] It persists into the present day as an involuntary response that results when we have, consciously or unconsciously, made an assessment of our agential possibilities and determined that there is no escape. So we play dead, though of course the "play" in this phrase suggests an intentionality that is absent in the phenomenon. All of us have seen animals do just this, especially animals whom we regard as most "unlike" us – lizards, bugs, possums, and the like. Scientists refer to this behavior as "apparent death;" apparent, that is, from a third-personal perspective. From within, though, this death enters into the first-personal phenomenology; it is the feeling of death washing over you.

Indeed, this is exactly what so many victims of oppression-linked violence report: think of Brison's repeated mistake in recounting what happened to her as a murder, of

---

[16] Brison, *Aftermath*, 15.

[17] For an in-depth study of groundbreaking research on the impact of stress and trauma on nervous system regulation, see Stephen Porges, *The Polyvagal Theory: Neurophysiological Foundations of Emotions, Attachment, Communication, and Self-regulation* (New York: W.W. Norton & Company, 2011).

her having a sense that the assault impacted her so deeply that she finally understood "the appropriateness of what a friend described…as a Jewish custom of giving those who have outlived a brush with death new names."[18] Or, from a different context and different time, think of Fanon's statement at the beginning of this chapter: "I transported myself on that particular day far, very far, from my self, and gave myself up as an object."[19] Again, we find here a report of self-alienation, of a departure from the active, agential being that one had the potential to be.

Looking more directly at the Fanon material, we also find him describing colonialist oppression as generating a "a zone of nonbeing, an extraordinarily sterile and arid region."[20] The Black man, according to Fanon, observes upon self-reflection: "Here are the fragments put together by another me."[21] When the colonized Black man looks inward, he sees only reflections of and aspirations to whiteness – there is nothing substantial in him except for either his distance from whiteness (manifested as the "savage") or his always-precarious embodiment of it (always at risk of, say, slipping back into his original accent). Fanon refers to this (self)-alienation as the distinct mark of living under, of *becoming* under, white colonial rule.

Already, we can see common ground between Fanon's description of shattering under colonialism and the shattering that occurs under patriarchy. Consider Brison's description of the paradox faced by assault victims: "How does one go on with a shattered self, with no guarantee of recovery, believing that one will always 'stay tortured'

[18] Brison, *Aftermath*, 21.
[19] Frantz Fanon, *Black Skin, White Masks*, trans. Richard Philcox (New York: Grove Press, 2008), 92.
[20] Fanon, *Black Skin, White Masks*, xii.
[21] Fanon, 89.

and never 'feel at home in the world'?"[22] Both are accounts of the trauma of oppression, and both detail the ways that it can fracture the continuity and coherence of identity with which we are normally accustomed, leaving a feeling of emptiness in its wake. But there is a difference between a feeling of emptiness and the moral status of being a *thing*.

Fanon addressed this distinction head-on in *Black Skin, White Masks*. He writes about Black consciousness under colonial rule: "A feeling of inferiority? No, a feeling of not existing."[23] When we take this observation seriously, we see that it merges the first personal with the ontological – it is not a mere experience of perceived inferiority, but of becoming *nothingness*. Elsewhere, he writes: "I came into this world anxious to uncover the meaning of things, my soul desirous to be at the origin of the world, and here I am an *object* among other objects."[24] These are strong philosophical claims. Fanon does not, for example, limit himself to describing a phenomenology, by e.g. saying "here I am *treated as* an object." No – he makes an ontological claim about the type of thing he could have become and the type of thing that he became instead. His account of the harms of oppression seem to reach beyond self-perception, toward the very type of being he became.

3.2 Making Sense of Oppression

How are we to make sense of these claims? One common strategy is to chalk these claims up to the distorting effects of pathology. Post-traumatic stress, this story goes, ruptures one's sense of self and the world so substantially that it can generate

[22] Brison, *Aftermath*, 65-66.
[23] Fanon, *Black Skin, White Masks*, 118.
[24] Fanon, 89-90 (emphasis added).

deeply held but *mistaken* beliefs about the contours of the world. Dissociation, in particular, and the freeze response that it accompanies, seems to be a prime candidate for a pathology-based "explaining away." After all, the world *is* real, these victims of traumatic oppression are *alive*, and our task is not to support them in upholding their shattered worldview but rather to aid them in picking up the pieces and recovering the unity of self-understanding that is characteristic of a healthy and functioning psychology. On this explanation, there is no need to take seriously the e.g. moral ontological claims that come out of post-traumatic stress; to the contrary, it is the determination that these claims are *mistaken* that allows us to view what is happening *as* a deviation from normal functioning.

The shortcoming of the pathology-explanation, though, is that it disregards both the strength of conviction with which these observations are made, and the orienting sense that the claims give to victims of oppression. In other words, the pathological designation renders victims of oppression epistemically unreliable on the very subject about which they have first-personal knowledge. In particular, it ignores the extent to which the "symptoms" of traumatic oppression can in fact be manifestations of the sense-making that is going on behind the scenes; that perhaps an entirely different understanding of the moral landscape has been revealed to them and is now informing the types of claims they make about themselves and their social world.

On this alternate reading, we can see dissociation and the beliefs that accompany it as reasons-responsive, whether or not we determine it is also pathological. Let us consider in more detail this phenomenon, i.e. the phenomenon of *becoming nothing*, that can occur in the midst of patriarchal violence to bring out this point. For this, I'll use

observations from my own upbringing in our patriarchal culture, with the hope that its specificity will yield something more general with which the reader can identify. In the absence of identification, though, these observations can be regarded as testimony from within oppressive conditions, testimony that reveals both the agency-centered beliefs of the victims of oppression and the moral-ontological claims that come out of these circumstances. They are one example among many of the sense-making that goes on in the life of any rational agent faced with the task of navigating a hostile world.

To recall, an important element of dissociation in the midst of oppression-related trauma is that it is triggered by one's beliefs about the world – more specifically, it is triggered by beliefs about one's own agential efficacy. Dissociation and the freeze response only come about when one has determined that there is no way to exercise one's agency in order to save oneself; it is, in a sense, a self-protective measure sensitive to the way one assesses the world to be. How might these beliefs be formed in oppressive conditions?

Patriarchal oppression, as one example among many, can be the site in which these beliefs form. And it is no wonder: patriarchy is both pervasive and devastating. There is no need to catalogue all the statistics here. All that is needed is for us to reflect on the kinds of people we have been groomed to become. In my case, as a young girl surrounded by young girls, we were groomed for passivity and the desire to please the boys around us. We were told that the sexual harassment and overt objectification were just the cost of going out with these bodies. We developed twisted strategies for escaping the sexual violence that we felt lurked around every corner – strategies that involved mimicking consent until it was safe to flee. And yet this violence did not lurk only in our

fearful imaginations. Everywhere we looked, we received messages about our own violability, about our diminished sovereignty. We knew that our entry into adolescence was at the same time entry into collective property. Where we rebelled against it, in our confused ways, we encountered a silence so pervasive that it silenced us anew. And as we collected traumatic experiences, witnessing ourselves and our friends become *things*, we concluded, rightly, that our survival was not going to depend on our environment changing.

Instead, we unknowingly turned to dissociation in order to survive. This was triggered by an assessment of the modal rigidity of the dangers of our environment – an assessment about our agential capacities. And on this front, two things must be noted: first, it was rational for us to believe, within the conditions I have just described, that individual action would have done almost nothing to change the circumstances. I do not believe it is cynical or overly-pessimistic to simply note that the forces of patriarchy are strong, and circumstances have to be world-historically right for the actions of a single person to make a substantial dent in the overarching conditions of oppression. So it is *true* when we unconsciously determine that no amount of individual action is going to get us out – there is nowhere to go, and speaking up will get us nowhere.

Second, it is not irrational for those living under patriarchal conditions to conclude that growth into girlhood and womanhood is growth into thinghood. From this perspective, we concluded that entry into adolescence really was the erosion of the moral personhood of girls. The nothingness that we felt wash over us, like the "living death" described by Brison, was not perceived as a metaphysical aberration. Rather, we felt that this is what we really were in a patriarchal world – we were objects for the use of

others and had to make the best of this fate. Of note is that this conclusion was a philosophical claim: it is not merely the claim that this is how others perceive us, for example, but rather that this is what we *are*. Of course we did not explicitly characterize our beliefs about the social landscape in this way; but the beliefs were nevertheless moral-ontological ones. And in a sense, they had to be. Those were the only ones that made sense of the world and our place within it. There is comfort in coming to understand the world, and it is no different for those who discover that one's place in it is somewhere between nothingness and full moral personhood. The experience of traumatic dissociation is a perspective which generates moral-ontological claims about the world.

One way of making sense of these claims, then, is simply to refuse the pathological explanation. The pathological explanation regards such beliefs as only symptoms, and in a sense moral progress is measured by reference to individuals' ability to distance themselves from these beliefs. Healthy functioning is, on this account, the ability to come to the belief: "I am not a mere thing." But when we refuse the pathological account, we not only refuse this metric of moral progress but also refuse to regard these beliefs as symptoms. Instead, the beliefs themselves are taken seriously *as* philosophical claims, and the position from which they are held is not cause for skepticism but rather moral curiosity. This moral curiosity, I argue, leads us to the following question: when we take these claims seriously, what larger ethical views do they reveal? What conception of the person is being mobilized such that these claims are part of the sense-making that occurs from within oppressive conditions? The next section explores these questions further, bringing in the conceptions of the person discussed in

Section 2 and analyzing their varying adequacy in regarding the people making the claims as sense-making agents rather than mistaken victims in need of correction.

4. Mapping our Ethical Theories onto the Claims of Oppression

To recall, Section 2 described two opposing views of moral personhood: the individual view and the interpersonal. The *individual* view holds that one's status as a full moral person is determined only by some feature of the individual in question:

The Individual view:

Some feature of the **individual** ⟷ **Full moral personhood**

And the interpersonal view holds that individual features of the person are relevant to one's moral personhood, but not the whole picture. Instead, full moral personhood is a product of this individual feature of the person combined with how that individual feature is taken up interpersonally.

The Interpersonal view:

(Some feature of the **individual** + Some feature of that individual's **interpersonal** context) ⟷ **Full moral personhood**

Let us apply each in turn to the claims coming out of oppression. First: How might someone holding the individual view make sense of e.g. Fanon's ontological claim that colonialism has in fact turned him into an object? It seems they have two options:

either agree that Fanon is correct, and attribute that loss of moral personhood to a loss in the relevant individual capacity; or disagree with Fanon, argue that he never lost the individual capacity relevant for moral personhood, and instead present a diagnosis explaining where Fanon went wrong in his own moral self-understanding (i.e. a deflationary account of his claims). Neither of these, in the end, are adequate ways of making sense of his claims.

On the first option, what individual capacity can the individual theory point to that would both count as a reasonable criteria for membership in the moral community and also be clearly lost by Fanon and indeed all others living under colonialism who have that experience of nothingness? The traditional candidates cannot fulfill both of these requirements: Fanon clearly has not lost his practical rationality, evidenced by his astute analysis of the experience of becoming an object. He has not lost his status as a member of the human species. He has not lost his consciousness; his self-consciousness; his capacity to suffer. What else is there? Those criteria cover the highlights of the history of moral philosophical reflections on the question of membership in the moral community. On the individual conception of moral personhood, Fanon cannot be correct about the moral-ontological impact that oppression has on its victims.

So the individual theorist will have to pivot to offering a diagnosis of where he went wrong. In fact this strategy is already familiar in the common understanding of Fanon's work. He has been typically read as providing us a "merely" psychological account of oppression, despite the ontological flavor of his claims. Fanon was, after all, a psychoanalyst. And (the deflationary account goes) this psychological interpretation of oppression is still a tremendous contribution to our understanding of the phenomenon.

We could read Fanon as describing the phenomenology of oppression, a first-personal account of the deformations of character we undergo under systems of domination. We would have become acquainted with the specific pathologies, neuroses, and beliefs that form under such conditions (much like J.S. Mill did in *The Subjection of Women*), the ways such a psychology restricts the possibilities in one's life, and the impoverished sense of self that this gives rise to.

Note that this deflationary psychological account is consistent with the claim that Black men under colonial rule truly were full moral *persons*, though they did not experience their lives as such nor were they treated by white colonizers as such. It therefore is not only a rejection of Fanon's specific ontological claim, but also a rejection of the *kind* of claim that could possibly come out of such reflections. That is, the individual view, in offering a deflationary account of the insights coming out of oppression, not only denies the particular conclusions about becoming an object, but also insists that we reinterpret those conclusions as of a different philosophical kind. We are forced into reading e.g. Fanon's work under a philosophical description different than the one under which it is presented.

But we have good reasons not to do this. First, it is a general prudential principle in philosophical analysis to take philosophical claims as the *kinds* of claims they are at face value, unless we have reason to believe that those claims are the product of pathology, deception, ideology, or other misleading influences. I have already argued above that the lens of pathology is not the most philosophically useful (or accurate) one through which to understand the experience of oppression from the inside. The conclusions one comes to in order to survive are ones that are reason-responsive; the

experience of oppression-linked trauma is one that is both responsive to *and* generative of ontological assessments about the world and oneself. And we have no reason to think that the claims themselves are the products of deception or other misleading influences, either, without begging the question – holding that Fanon is e.g. deceived about his moral status requires one to believe that his moral personhood is not susceptible to change in the circumstances he describes. But this is the very question at issue.

Second, there is a sense in which offering a theoretical diagnosis as a response to someone's first personal claims about what oppression has done to them is a way of perpetuating that very oppression. On this reading, the individual view (and the inviolability of moral personhood to which it is committed) is in service of systems of oppression insofar as it misdescribes the depth of its injustice. According to those who have lived through oppression, we are not moral persons no matter what. The contingency of moral personhood is in other words just the possibility of a moral harm so extreme that it is in the interest of dominant ideological systems to deny its possibility. This denial establishes a logical floor to the possible effects of interpersonal behavior below which any victim's descent is rendered (allegedly) impossible. It therefore guarantees that testimony from this below-ground space, and especially philosophical testimony (insofar as it makes claims about what is really going on), will not be taken seriously.

Finally, there is good reason to think that this below-ground space in fact has an epistemic advantage over other perspectives within oppression, at least insofar as it relates to the moral-ontological realities of that system. This is not a new idea. In Marx's analysis of class relations under capitalism, the proletariat's lived experience has the

potential to reveal the exploitative nature of capitalist production. In Black feminist epistemology, the situated perspective of Black women most accurately identifies the workings of patriarchal white supremacy in part because of Black women's urgent need to survive those conditions. As James Baldwin says,

> People who cling to their delusions find it difficult, if not impossible, to learn anything worth learning: a people under the necessity of creating themselves must examine everything, and soak up learning the way the roots of a tree soak up water.[25]

The common thread here is one in which being a member of the oppressed class gives you a unique opportunity to understand the workings of that oppression more generally, an opportunity that is precluded for the oppressors (see for example Charles Mills' work on white ignorance). Whether this is because of the necessity of survival, or proximity to the absolute core of the political mechanism, does not change the larger point. To be oppressed is to be simultaneously acquainted with both the truth of oppression's staggering impacts and the illusory nature of the stories told to support it, perpetuate it, and hide it. Denying the claims that come out of this space, then, is not only epistemically irresponsible but also normatively suspect – it is a failure to discern when one is hearing from an "expert" in some sense solely on account of their social standing or identity, as the case may be. This itself constitutes a harm, one that both fails to treat others as having a reliable subjectivity, which I take to be one of the bare minimums of interactions among equals, *and* furthers the epistemic and normative disenfranchisement that is characteristic of oppression.[26]

---

[25] James Baldwin, *No Name in the Street* (New York: Vintage Books, 2007), 129.
[26] For more on this, see Miranda Fricker's *Epistemic Injustice*.

For the above reasons, we have good reason to reject the deflationary account that the individual conception of moral personhood can at best offer to explain Fanon's claims. Does the interpersonal conception of moral personhood fare any better? That is, does the interpersonal conception of moral personhood make sense of Fanon's claims (and others like them) in a way that does not distort them beyond recognition? Yes – in fact it avoids all of the shortcomings of the individual conception just described.

First, the view's two-part structure makes it easy to see the mechanism by which one could become a moral object and yet retain the morally significant features of oneself. Fanon became an object because his *humanity* was not taken up interpersonally in the right way. As he described it, his full moral personhood was stymied not because of any lack of his but rather because of the pervasive treatment he encountered in colonial Martinique. He retained the features of himself that made him a potential full moral agent, but those features were not regarded by others as reasons to include him in moral community. Instead, he was subject to the myriad manifestations of oppression. So his full moral personhood never actualized. There is no mystery to the structure of this phenomenon.

Second, the interpersonal account avoids the three problems of the individual account's deflationary approach. Most obviously, it is able to take Fanon's claims as the *kinds* of claims they purport to be. There is no explaining them away; no attempt to fit an ontological claim into a psychological framework (though it is consistent with holding that the claim itself originates in the moral psychology of oppression). The interpersonal account, because of its two-part structure, makes sense of the possibility that there could be an ontological shift that happens as a result of oppression. In this respect the

interpersonal account has a methodological advantage over the individual account. It is able to preserve the kind of the claim it is meant to explain.

Furthermore, because it preserves the ontological status of claims such as Fanon's, the individual view is not in service of oppressive structures in the way the deflationary account is. In particular, preserving the ontological status of Fanon's claim is a way of accurately describing the true extent of oppression's harms on its victims – it refuses to implement a floor on assessments of how bad things can get. In this respect, the interpersonal account is part of a *critical* reflection on oppression rather than a *complicit* one.

Lastly, the interpersonal account properly acknowledges the epistemic advantage of the perspective of oppression's victims. This should count in its favor. The term 'epistemic advantage' itself has many interpretations. For my purposes here, I take it to refer to a vantage point that is uniquely situated to facilitate veridical observations and understandings of something about the world in ways that are obscured from other vantage points. It is like being perched up in a lighthouse and seeing the rough shoreline, invisible to those in the midst of the waves.[27] Of course there is a distinction between having an epistemic advantage and being infallible – the person in the lighthouse could lack the requisite knowledge for operating the search light, or for estimating nautical distance, etc., and therefore have worse knowledge of the ocean's conditions than those at sea. Being in the epistemically advantaged position is no guarantee of getting it right. But being in the epistemically *dis*advantaged position is a guarantee of getting it wrong

[27] This analogy is, in the end, imperfect for many reasons. The ideal analogy would be interactive and structurally organized such that the same feature both epistemically advantages one party and *dis*advantages the other.

(or at least of not having *justified* true beliefs), and it is this contrast that the point about epistemic advantage is meant to highlight. The perspective of oppression's victims is one that has the potential to reveal oppression's workings, and those who benefit from that system are in effect blinded from this same revelation without external intervention (usually from the testimony and theory that emerges from the oppressed classes).

This point about epistemic advantage is central enough to my argument that it deserves a full section (which follows this). To review, though, I have argued that a cluster of first-personal accounts of oppression share a common theme: that one of oppression's most devastating effects on the person is indeed an erosion of that very status. I have also argued that more familiar accounts of moral personhood, on which it is dependent on only a feature of the individual, can only explain these claims away – in order to make sense of them, the individual account must show how these claims are the product of oppression's distortions of its victims self-conceptions. It thereby alters the *kind* of claims found in these first-personal accounts, transforming them from ontological claims to moral psychological ones. I then showed how an interpersonal account of moral personhood can not only easily make sense of these claims because of its two-part structure, but also avoid the shortcomings of reverting to deflationary accounts and changes in philosophical kind.

A reader may ask, though: why do we want to make sense of these claims at all? There are many perspectives in social and political organization – it is at least plausible that the most accurate one is not one that is developed in extreme marginalization and under extreme duress, but rather one that is formed apart from the pressures of survival. Or perhaps the correct one is a hybrid account, taking the partial and limited

perspectives of each social location and aggregating them into one comprehensive and therefore more reliable conception of the person. And so on. The point is that the theories underlying first-personal perspectives of oppressive victimization do not *prima facie* demand explanation any more than any others do.

This is where the discussion of epistemic advantage can intervene. The following section is an exploration of the type of epistemic advantage that may be developed in conditions of oppression; what exactly that advantage's object is and isn't; and ultimately why it is the case that, contrary to the imagined worry above, moral philosophy has a special obligation to make sense of *these* claims *on their own terms* (and not others). As I have done in previous sections, I would like to return to first-personal accounts, this time however shifting our attention to the world of film and analyzing Barbara Loden's 1971 film *Wanda*.[28] It is, in my opinion, one of the best cinematic depictions of what it is like to be a thing, and how relentless interpersonal treatment can cause such a status. Unlike Brison's memoir, which focuses on the aftermath of acute traumatic patriarchal violence, Loden's film instead is a meditation on the slow accumulation of patriarchal abuse which contributes, over a lifetime, to women's status as *things* in a patriarchal moral metaphysics. It is a masterpiece in the quiet ways that patriarchal objectification can seep into one's self-conception and one's life. And it beautifully explores the painful reality of navigating this world when you know that this is what you are.

[28] *Wanda*, directed by Barbara Loden (1971; New York: Foundation for Filmmakers)

5. *Wanda* and Knowing What You Are

Wanda depicts the life of Wanda Goransky, a beautiful and adrift woman in a 1970's Pennsylvania coal town.[29] We observe Wanda's unmooring from her last shreds of respectability as she divorces, loses her children, and slides into aimlessness. Eventually the aims of strange men take over, and Wanda becomes unwittingly entangled in a doomed plot to rob a bank. These traditional plot points are more dramatic on paper, though, and in the film they merely punctuate an otherwise slow character study, one in which Wanda's passivity in the face of men's desires is the main event. It is a study in what it means for a life to happen *to* someone, as opportunistic men see in Wanda a fulfillment of their own desires; and a study in what it means for someone to view this instrumental role as the most that they can offer. As Bérénice Reynaud describes it, Wanda is a character who "drift[s] in the sea of her own insignificance, clinging to unworthy men as a way to avoid drowning."[30]

This "clinging," a sad coping strategy in a world that confines Wanda to a space of moral servitude, permeates the film. It also shows us the first-personal perspective of patriarchal objectification and the specific types of knowledge that this perspective generates. One sequence in particular, after her summary divorce and being denied re-employment at a dressmaking factory on account of being "too slow" to be "used" (according to her former boss), shows us this perspective in full daylight. Wanda has stopped at an anonymous and nearly-empty diner, her blond hair in rollers. Broke and without any clear plan, she orders a Rolling Rock. But before she can hand over

[29] I am grateful to Bellamy Mitchell for numerous conversations on this film.
[30] Bérénice Reynaud, "For Wanda," in *The Last Great American Picture Show: New Hollywood Cinema in the 1970s*, eds. Thomas Elsaesser, Alexander Horwath, and Noel King (Amsterdam: Amsterdam University Press, 2004): 225.

payment, a middle-aged man at the bar interjects, addressing the waiter: "I'll take care of that." What happens next is a quick succession of subtle and devastating movements from Wanda: she glances at the man and puts away her coin purse, as if resigned; she collapses her grimaced face into her hand in a moment of defeat and apparent pain; and, seemingly gathering some unknown fortitude, as if willing herself to do what must be done, she pours the beer, takes a sip, and gazes out the diner window while pressing the cool glass to her cheek. Not once do Wanda and this man exchange words.

The scene abruptly cuts to Wanda, naked and asleep in a disheveled motel bed, her white purse perched on the radiator beside her. So this is what she felt had to be done – a repayment of sorts. The man from the diner is busy trying to sneak out undetected, uninterested in Wanda once she has fulfilled her sexual duty. What is so revealing about these two scenes, and the cinematic movement between them, is that what is included is what Wanda's perspective deems relevant. And this is the perspective of being an object: what is relevant here is a strange man's assertion, his financial usurpation, his making-it-so. Both a debt and an ownership emerged in that diner, a gendered property relation on account of an unpayable debt. Wanda knows this is how it goes. She knows that her role, what it is to be what she is, is to sleep with this man who has announced, in his way, that this is what he wants.

A different cinematic perspective would have deemed other narrative moments relevant and essential to the story: to start with, *any* of the words they must have exchanged would have been included – a backstory for this man, who remains nameless, a flirtation perhaps, a seduction. Instead, *Wanda* defies cinematic convention and omits them. The absence of these conventional elements is an expert way of showing us the

story from Wanda's perspective, steeped as it is in her own objectification. It is a way of taking the first-personal perspective seriously. And hers is one of knowing her place as a sexual object: it is one that is highly attuned to assertions of ownership, of knowing when one's status as a mere means has been called upon. Where others might see a flirtation in the buying of a drink, Wanda knows better: her number has been called.

This type of knowledge – i.e. acute awareness of one's moral status as well as when it is being called upon – is one of Wanda's main survival strategies throughout the bulk of the film. The reality is that Wanda is in need: she has no home, no job, no apparent friends to call upon, and only a few dollars left in her purse. Her options are severely constrained by the sexism and sexual objectification she encounters at every turn; these draw a border around how she might get "out" of this precarious position. Her tacit knowledge is what Patricia Hill Collins describes as "wisdom" insofar as it is "essential to the survival of the subordinate."[31] And according to Collins, there is good reason to "give such wisdom high credence in assessing knowledge."[32] It is a sort of epistemic survival of the fittest – the urgency of making it another day puts enormous pressure on getting it right about the world and one's place in it. Wanda herself could have uttered the same sentiment that Collins refers to in her own epistemic work: "I know what I'm talking about because I'm talking about myself."[33] This is an epistemic authority grounded in the first-personal perspective of the oppressed, because the pressures of survival generate epistemic reliability.

---

[31] Patricia Hill Collins, "Black Feminist Epistemology," in *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment* (New York: Routledge, 2000), 257.
[32] Collins, "Black Feminist Epistemology," 257.
[33] Collins, "Black Feminist Epistemology," 258.

In addition to this knowledge-as-survival, Wanda navigates her world using her (white, female) passivity and her stubborn kindness, perhaps the only other currency she has available to her. She comes to meet Mr. Dennis, another opportunistic man, in circumstances similar to the diner scene: After wandering through town and getting her money stolen while asleep at a movie theater, she enters a late-night dive bar all but empty save for a man who barks at her that they're closed. No matter – she just needs to use the restroom. Unbeknownst to her, Wanda has entered the scene of a crime: the man at the bar, whom we later come to know as Mr. Dennis, has knocked out the real bartender (who lays supine behind the bar, just out of sight with a rag stuffed in his mouth). Mr. Dennis is busy getting cash out of the register between furtive glances out the closed blinds. And Wanda sits down at the bar, disheveled and munching on some stale chips, unaware of what exactly is going on. She asks for a comb, a drink, a mirror, which Mr. Dennis tosses to her with disdain. But soon he needs to make his getaway. He barks at Wanda "let's go," and the pair is solidified. Where else can she go? Her aimlessness fits well with patriarchal domination. She knows her place, and so does Mr. Dennis.

When the film was released in the United States in 1971, it garnered intense backlash from feminist film critics. None was more explicit than Pauline Kael's, who wrote in the New Yorker that Wanda was "such a sad, ignorant slut that there's nowhere for her and the picture to go but down…."[34] Critics at the time wanted instead to see the empowered woman protagonist, the woman who knows her own worth, who perhaps has some feminist awakening during the film. But of course the film is a slow meditation

[34] Kael, Pauline, "The Current Cinema," *The New Yorker*, March 20, 1971, 136.

on what it looks like when this never happens. Indeed, Barbara Loden selected her editor, Nicholas T. Proferes, precisely because she felt "he knows how to show ugly things without it appearing ugly."[35] This is the ugliness of merely surviving patriarchy – of having no narrative redemption, of leveraging one's status as a sexual object because one has nothing else, of never becoming the good wife, the good mother, the good woman. In a sense, Kael was right: Wanda *was* a sad, ignorant "slut," insofar as that picks out a particular place in our social landscape. This very accuracy, while an object of scorn from feminist critics at the time, is its striking feminist insight. *Wanda* depicts the experience of patriarchy from the perspective of its victim; this is an ugly, misunderstood place. It is also a revealing one.

As the film continues, Wanda and Mr. Dennis are on the run, hopping from motel to motel, stealing cars outside of churches, making quick clothing changes at roadside rest stops. Wanda endures the motel sex, the hand on her thigh, the slaps across the face as the price of admission. Mr. Dennis is particularly preoccupied with her appearance, insisting that she wear only the most obvious trappings of puritanical femininity (yes to sweet dresses, no to trousers and red lipstick, which Mr. Dennis tosses out the car window on the highway). And he can't stand her disheveled blond hair – at this point she has lost most of her rollers, her scarf, even her comb. Wanda eventually finds a solution that pleases him: a ridiculous white hat with white fabric flowers creating a halo above her head. This hat is one of her main capitulations to Mr. Dennis's gruff complaints, his strident opinions about what she must wear; it represents her passivity, in all its glory, and she wears it like a crown.

[35] 00:46:57, *I Am Wanda*, directed by Katja Raganelli (1980; Munich, West Germany: Diorama Film Munich).

At one of their roadside stops, Wanda makes a stunning announcement, stunning for its insight and for the nonchalance with which she says it. Mr. Dennis, in his usual gruff manner, is interrogating her about her family after having seen their picture in her wallet. The conversation is as follows:

Mr. Dennis:	Where's your husband?
Wanda:	What husband?
Mr. Dennis:	*Your* husband.
Wanda:	I guess he's got himself a real good wife by now. Got a *real* good wife.
Mr. Dennis:	What about the kids?
Wanda:	Kids…
Mr. Dennis:	Yeah – I saw their picture in your wallet.
Wanda:	Well they're with him. Better off with him. [long pause] I'm just no good… [laughter] I'm just no good!

Wanda makes this final declaration with a mixture of resignation, protest, stubbornness, and a hint of recognition of the absurdity of it all. "I'm just no good!" It is as if she is exasperated that she even has to say it – after all, doesn't the fact that she's on the run with Mr. Dennis speak for itself? That she is divorced, estranged from her children, with no job, and sleeping around? Her frustration with stating the obvious is apparent.

Her choice of words is also revealing. Wanda doesn't contrast herself with her ex-husband's new wife by using the polar opposite – that is, she doesn't say she's a *bad* wife, a *bad* mother. Instead, she pivots to the language of "no good." What is the difference between bad and no good? Part of what it means to take the first-personal perspective of oppression seriously is to assume that statements like these are made from a place of knowledge. Our task, then, is not to scrutinize whether it's true but rather to extrapolate

a coherent worldview in which this statement makes sense. In this case, Wanda's statement reveals that in the gendered landscape she inhabits, the opposite of a good wife is not a bad wife but rather someone who is worth nothing. Being a "good" wife is to have proximity to gender success; and under patriarchy, gender success is a necessary condition for moral worthiness. Now of course, even the good wife is not on an equal moral footing with e.g. the "good" husband. The point, rather, is that Wanda's failure to adhere to gender norms turns out to be a moral failing and not a mere social failing.

This scene, and the worldview that her brief comments reveal, is one in which gender is a very high stakes endeavor. Not only is her safety up for grabs, which is bad enough. But so too is her moral worthiness which would ground any such critique of that moral precarity. Wanda does not have a secure grasp on the ability to justifiably protest when her safety is compromised by Mr. Dennis, or any of the other men in the film for that matter, because she does not have a firm grasp on being a good woman. This is the logic of patriarchy. Wanda is a mere means, and even she knows it.

What else could she have possibly concluded? Every interaction of hers confirms this – she is denigrated by her ex-husband for failing to take care of the kids, for failing to have meals ready when he returned home from work; everywhere she goes, men are vying amongst each other for who will take her home. Mr. Dennis himself refuses to even speak with her after those nights where they have sex, instead demanding that she run out in the middle of the night to fetch him burgers. And, of course, the robbery: Wanda makes it clear throughout her time with Mr. Dennis that she doesn't want any part in it. She sobs, "I can't do it, I can't do it" and vomits in the bathroom before they embark. But she also knows that this is the only way he will keep her around – if she

carries out *his* will, *his* projects, *his* aims. So she carries on, and even comes to blame herself when the robbery doesn't go according to plan. She berates herself for failing to be a good means rather than blaming Mr. Dennis for expecting her to be one. Not only does Wanda lack the self-conception on which this appropriate blame is possible, she also, because of the way she has clearly been treated her whole life, lacks the full development of her moral personhood which would ground that critique. Patriarchy, in its daily interpersonal manifestations, has stymied her full development as the type of person she has a right to become.

I take Wanda's experience to show how one can become a thing under pervasive conditions of oppression. I also take it to show the shortcomings of the individual account of moral personhood, one on which Wanda was in fact a moral being from *some* eternal perspective. Where is this perspective? All evidence in Wanda's life suggests that she is a thing – she takes herself to be one, she is treated as one, and she doesn't expect to ever be anything else in her moral community. Indeed, the film ends where it began: Wanda is wandering around the perimeter of a boisterous bar; she is taken in and sandwiched between various revelers singing and drinking, her small frame wedged in a sea of familiarity, highlighting her alienation and lack of belonging. And on her face we can see that she has nowhere to go. In all likelihood, she will go home with one of the men at her table, yet again spending the meager currency of her objectification in exchange for food and shelter. To insist that Wanda is an end in herself in *these* circumstances is both a disrespect to her experience and an impediment to her survival if she were to take up this perspective. She knows what she is.

It's likely, though, that this knowledge has precluded her from seeing what she might be. This is the outer limit of her epistemic authority, because it is also the outer limit of her immediate survival. That is, Wanda's survival depends on her knowing her moral status as a means. But it also requires that she be in the dark with respect to the full picture – using the language of the interpersonal account of moral personhood, we can say that Wanda's survival requires her to *not know* that she is the type of being who deserves conditions in which her full moral personhood can flourish. So her lack of any apparent understanding of the injustice of her situation – her wearing her passivity like a crown – makes sense. This ignorance reduces the amount of interpersonal friction she faces. It also reduces the amount of friction she might experience internally. Many of us have had the distressing experience of the cognitive dissonance that emerges when one has their "consciousness raised" in the parlance of second-wave feminism. The world becomes more painful when you begin to realize that you deserve more but do not have the tools to secure it; this further knowledge adds another layer of distress to an already-distressing experience under oppression. Talk of Wanda's epistemic authority, then, is not in tension with this "mistake" about herself – indeed, when we understand epistemic authority from the perspective of survival, we see that it is fully compatible with certain omissions in one's moral self-conception where those omissions help you carry on another day. Patriarchy can sharpen our tacit moral knowledge but dull our moral curiosity. Wanda is an exemplar of this dual impact – she is an authority on the social ontology of her situation, but not on the normative implications of that social ontology.

I prefer this explanation of the distortions in self-conception that happen under traumatic conditions of oppression. Instead of insisting that she is a full moral being but

simply doesn't know it, as the individual conception of moral personhood does, we can take a more respectful and subtle approach with the interpersonal conception of moral personhood. Respectful, in the sense that we accept her as an authority on her own moral (non)status given these conditions. And subtle, in the sense that there is still room to describe the ways that this moral (non)status precludes her from inquiring into the whole picture, from seeing that she in fact has a right to become something other than what she has in fact become, without giving in to a condescending form of moral paternalism. I mentioned this sort of paternalism in my earlier discussion of Fanon, who is typically read as providing us a moral-psychological deflationary theory and not a moral metaphysics. As in that case, we must avoid this sort of paternalism, because it is a denial of the epistemic authority of the position of the oppressed. And this denial prevents us from developing an account of moral personhood that has a grip on lived experience.

6. *Wanda*, Barbara Loden, and Individual Strategies for Moral Emancipation

Before concluding this chapter, it is worth mentioning one more thought about consciousness-raising. In second-wave feminist terms, this is the experience of joining together with others in similar social locations as oneself, sharing one's stories, and coming to see that what once seemed to be an isolated, "personal" problem is in fact not only shared but *political*. This shift takes place both in one's understanding and in one's moral being – one comes to see that the status quo need not be the case, and insofar as there is a "problem" it is located not in the individual but in the political conditions in which she lives. This de-problematization of the self makes way for new political analyses

and self-understandings. Consciousness-raising, therefore, can be understood as the process of entering into the interpersonal in order to begin the journey towards full personhood. Indeed the interpersonal element was crucial: it was the leaderless, informal groups of women simply sharing their stories, and finally being heard, that contributed to the dawning feminist movement of the 1960's and 70's.

It is clear from my analysis of Wanda that there is a missing element in her self-understanding that could have been provided by the interpersonal affirmation of consciousness-raising groups. Imagine if she had a group of women with whom she could share ordinary stories and come to see then as unpleasant, painful, and demoralizing in similar (sexist) ways. But she lacked access to such groups – everywhere she went, she was surrounded by men who were heavily invested in the patriarchal status quo. Wanda was therefore systemically precluded from making the conceptual leap from her knowledge of the social ontology of her situation to a developed sense of its moral implications. On the interpersonal view of moral personhood, does this mean she's out of luck? That is, if it is indeed the case that moral personhood is contingent on the social, does this mean we are doomed if we find ourselves in oppressed positions without alternative interpersonal environments in which to cultivate our moral personhood?

No – there are other strategies for climbing one's way out of nothingness. Indeed Barbara Loden herself, the writer, director, and actress in *Wanda*, shows us this is the case: considering her relationship to both the film and its titular character reveals the first steps of moral becoming from within a position of a stifled moral personhood, where one cannot rely on the interpersonal to generate that moral becoming. Loden demonstrates how one can use the *intra*personal in order to mimic the interpersonal

experience of e.g. consciousness-raising and develop a full moral self. The following is a discussion of that intrapersonal process.

Prior to making *Wanda*, the only feature film she ever wrote and directed, Loden was already a highly successful actress. She starred in numerous films directed by the behemoth Elia Kazan, who later became her husband, and rose to even greater acclaim for her Broadway role as Maggie in Arthur Miller's *After the Fall*. But her ascent took place in roles written by men – self-destructive women; women scorned for their promiscuity; women who were passive in their own lives.[36] And throughout it all, she harbored deep insecurities about herself. She longed to be accepted as a human being and an artist, describing herself as a person, like Maggie in *After the Fall*, "who was very insecure, and felt that she was not educated, that she didn't know very much, and [who] had a very strong need to be accepted."[37] She married the men who wanted to marry her; she danced for ogling men at The Copacabana in New York City; her acting talent, which included an exquisite capacity for rendering the subtleties of female suffering, was for decades a springboard for other men's fame rather than her own (beyond being known as a blond bombshell actress).

At some point, though, Loden became tired of this arrangement. She had also come across an intriguing story in the newspaper about a woman who had been convicted of being an accomplice in a failed bank robbery. At sentencing, when this woman learned that she was facing twenty years in prison with no appeal, she thanked the judge. In an interview many years after the production of *Wanda*, Loden reflected on

[36] For an excellent account of Barbara Loden's acting career and the role that gender played in it, see Nathalie Léger, *Suite for Barbara Loden*, trans. Natasha Lehrer and Cécile Menon (St. Louis, MO: Dorothy, 2016).
[37] 00:34:00, *I am Wanda*, Raganelli.

finding this story in the paper: "I started wondering, why would she thank him? Why would she be glad to be put in an institution, into a prison?"[38] Loden held on to the clipping for many years, quietly reflecting on this woman's story. Something had stuck – something about the way this woman had left her children, wandered aimlessly in her own life, and attached herself to a "petty criminal." Something about her gratitude for her own incarceration.

Eventually Loden wrote the script that would later become *Wanda*. She shopped it around to various male directors, assuming that the same pattern that launched her career would manifest in this creative turn, too: Loden as passive actress, a powerful man as director. And yet, each time she spoke with prospective directors, she "didn't think they really understood what [the script] was really about. They were all men, which wouldn't necessarily make a difference, but they didn't seem to understand what this woman was about."[39] Loden was emphatic when she said this. So what was "this woman" all about? Who, to Loden, is Wanda?

Loden describes her in the following way:

> [She was] a woman who was ill-equipped to deal with life…She had hardly any education, she was not particularly bright, she came from very poor working class people, and she could hardly function even as a working person. She married very young and had several children, there again she was hardly what would be called a good housekeeper or even a good mother.[40]

Later, she comments that Wanda "drifted from one situation to another, from one man to another." In a sense, Wanda was just like the other women Loden had played in her

[38] 00:43:50*, I Am Wanda*, Raganelli.
[39] 00:45:05, *I Am Wanda,* Raganelli.
[40] 00:47:37, *I Am Wanda*, Raganelli.

career – a sexual object, a drifter, a recipient of men's abuses. Wanda is absent from her own life. In an interview after the film was released, Loden said the following:

> [Wanda is] trying the best thing she can, which is just really to drop out…a lot of people do this, and they become very passive. This is one type of person we have in our society. A person that completely resigns, and lets everything walk over them."[41]

Wanda represented a certain sort of self-imposed passivity with which Loden was very familiar. And yet the prospect of another man directing Loden in *this* role, among all others, increasingly dissatisfied her. What didn't these male directors understand?

For starters, they didn't have first-personal access to such experiences, steeped in American gender norms as they are. A male director just isn't going to know, first-personally, what it's like to walk into a bar with nothing (as Wanda does) and realize that, in fact, what you have is a status as a sexual object that, when called upon, further plunges you into nothingness. They aren't going to understand and emphasize the narrative irrelevance of the words exchanged between Wanda and her "suitors." In short, they aren't going to know the intimate contours of life as an American white woman *from that perspective*. And *this* is what stuck with Loden about the newspaper article: to thank a judge for one's own incarceration is to be fully divorced from the possibility of agency in one's own life.

So it was Loden herself, familiar as she was with this perspective, who had to direct the film. She writes of this as if she discovered it through a process of elimination: after rejecting a series of male directors, and beginning to articulate her reasons why, she

[41] From Loden's appearance on "The Mike Douglas Show – Loden, Lennon, Ono," May 7, 1971 (the relevant clip of which can be accessed at https://www.youtube.com/watch?v=PtBuOTWoRpw).

became more closely acquainted with her unique vantage point as a filmmaker – and more importantly, with the value of this vantage point. In describing Wanda's pathetic situation, Loden often sounds like she is describing her own upbringing in rural North Carolina. And her marriage to Elia Kazan was well-known to be one that vacillated between patriarchal paternalism and misogynistic contempt. In an especially poignant moment, Loden remarked, "I get like that sometimes, like Wanda was, where I sort of – I don't have myself. I'm sort of floating, detached from any sort of security."[42] Only someone who knows what it's like to not have oneself could render Wanda's character with such exquisite humanity. Loden, in the end, had to direct herself.

In so doing, Loden gives us an insight into what it looks like to begin to emerge from one's own thinghood through an intrapersonal mimicry of the ordinarily interpersonal conditions for full moral personhood. It is a dawning realization, a gradual separation, where one occupies the object position but also, newly, the subject position. Loden was able to occupy the identity of Wanda while also viewing it from a critical, directorial distance. In this way, she began to have a critical distance on her *own* life and the ways that Wanda's trajectory could have been her own. In describing how she was able to direct herself in the film, she says, "I think one reason that I could be the actress and the director in that particular film was because of the nature of the role, it was a very passive role."[43] There is power in taking ownership of one's passivity, even if one doesn't immediately eliminate it. To become one's own director is to try on the guise of a subject without losing the identity as object that one has made one's home. From this perspective, one can begin to see the interpersonal context that generates this form of

[42] 00:55:49, *I am Wanda*, Raganelli.
[43] 00:54:43, *I Am Wanda*, Raganelli.

passivity. And yet, without having to give either up, one can find the resources for a dawning personhood. Though Wanda's passivity is what is captured on film, it is, as a result of Loden's self-direction, including (especially) her brilliant directorial choices and restraint in selecting what to film and what to leave out, an exemplar of a feminist awakening. Considering the authorship of this film in addition to its content shows us how one can emerge from a place in which these two positions are collapsed together. Months before her untimely death from cancer at the age of forty-eight, Loden explained: "I tried to be independent and to create my own way, otherwise I would have become like a Wanda, all my life just floating around."[44] Filmmaking itself, in allowing Loden to *critically witness* this aimlessness without fully inhabiting it, actualized the moral personhood that Loden was able to glimpse through the intrapersonal consideration of subject and object.

Reflecting on the film's place in feminist film history, Loden said the following:

> "When I made *Wanda*, I didn't know anything about consciousness raising or women's liberation. That had just started when the film was finished. The picture was not about women's liberation. It was really about the oppression of women, of people."[45]

Wanda's characteristic feature is her propensity to drift along the currents that the men around her control for their own benefit. Loden's ability to observe this passivity, rather than reenact it as a director, was the seed to its feminist insight. She created just enough distance for the viewer to see in the full light of day what patriarchal control of women has wrought. The film neither sentimentalizes nor plunges into nihilism. Loden made a

[44] 01:00:35, *I Am Wanda*, Raganelli.
[45] Barbara Loden, "Barbara Loden Revisited," interview with Madison Women's Media Collective, *Women and Film*, nos. 5-6 (1974): 67.

film about the quiet evisceration of the person that happens under oppression. In so doing – that is, in creating a subject position where there was once only an object – she took the first step towards her own moral personhood.

Her creative work, then, is an example of how one might emerge from stifling oppressive conditions even in the absence of the interpersonal conditions that typically facilitate that emergence. In a sense, she created those conditions for herself, *in* herself, in order to occupy both of the positions that are required for moral becoming. This is a genuine moral accomplishment, and should give us hope in thinking about the deepest harms of oppression. In the right conditions, with the right skills and resources, one might be able to actually pull oneself up by the bootstraps, moral-ontologically speaking.

But hope is different from justice. Loden's creative self-transformation (in addition to the transformations facilitated by consciousness-raising groups) shows us that there are morally-permissible responses to oppression for individuals who take it upon themselves to generate their own moral emancipation. We are not doomed in oppressive circumstances, even when those circumstances have harmed us in a deep moral sense. For this reason, we can be hopeful. But this is not the end of the story as far as our moral personhood goes, because this solution leaves the interpersonal conditions of oppression unanalyzed. When we turn a critical eye towards *that*, we see that there have to be *political* solutions to what is at its core a political problem. The degradation of full moral personhood happens in our sociality; what we should do with our sociality is a question of justice. The following chapter, then, looks at theories of justice with this type of deep moral harm in mind, and asks what theoretical commitments must be in place for political philosophy to adequately respond to this second question.

7. Concluding Remarks

In this chapter, I have argued that there is an ontological vulnerability at the farthest edges of our moral personhood: the contingency of moral personhood, as I have called it, refers to the way that our status as full moral beings is susceptible to the interpersonal context in which we live. I have also argued that this contingency becomes apparent through consideration of the first-personal accounts of living under oppression, and the very first-personal nature of these accounts is reason to take them philosophically seriously. When we do, we are able to see the advantages and disadvantages, respectively, of interpersonal versus individual accounts of moral personhood. These more fully-developed conceptions of the person are standalone elements of various ethical theories – and only the interpersonal ones can adequately describe the full moral harm of oppression that the contingency of moral personhood captures. Because of this, we have good reason to favor interpersonal accounts over individual ones.

Furthermore, this discussion allows us to make progress on the practical questions that arise from considerations of oppression. On an individual account of moral personhood, the contingency of moral personhood is a mere appearance, and therefore the problem to solve is an *epistemic* one – Fanon, Wanda, and anyone else dealing with the erosion of their moral personhood have made an error of moral self-understanding. But because moral personhood is an individual matter on the individual account, it is not obvious that it would require others to intervene. More likely, intervention would be a supererogatory act – and the recommendation would be to issue

an epistemic correction. We might be encouraged to remind people like Fanon and Wanda that they are moral persons or to correct them when they say otherwise.

But on the interpersonal account of moral personhood, the contingency of moral personhood is real, it signals a failure of the interpersonal, and therefore it requires that *we* intervene in ways that honor our shared responsibility for one another's moral personhood. Furthermore, this intervention will not be an epistemic one, since the interpersonal account posits no core epistemic error; rather, it will require that we facilitate the moral actualization of those who have lost their moral personhood (or never developed it) on account of the social. Part of this facilitation will simply be entering into interpersonal relations of equality and respect; taking up their individuality in the right way. Another part will include facilitating various individual attempts at moral becoming, as in e.g. Barbara Loden's *intra*personal transformation. The point is that part of thinking about the interpersonal requires that we see ourselves as implicated in other people's moral ontology, and what we do in response to this will itself have ethical implications. Part of the obligations we are under, then, as moral agents, include facilitating *others'* entry into this obligation, too – both creating and sustaining the interpersonal conditions that are required for the full moral personhood of all beings capable of that status.

CHAPTER 2
OPPRESSION, MORAL VULNERABILITY, AND JUSTICE

"Thus to respect persons is to recognize that they possess
an inviolability founded on justice that even the welfare
of society as a whole cannot override."

– John Rawls, *A Theory of Justice*

1. Introduction

Ideal political theorists and those that focus their political theories on oppression have been poised in tension for some time: ideal theory is, if anything, meant to conceive of the principles that the fully just state is required to have in order to achieve justice; to the extent that oppression is a paradigm example of *in*justice, then, these principles are assumed to rule it out.[1] But little is actually said by ideal theorists about the inner workings of oppression itself. Critics have seized on this omission, arguing that adequate political theorizing requires us to think more explicitly about oppression itself in its many forms, prioritizing this "nonideal" focus in order to conceive of a theory of justice that

[1] For purposes of this dissertation, I am assuming that oppression is one kind of injustice (among many), though I don't deny that oppression can also manifest in those social spaces that are not the proper purview of the state. Others have situated oppression in different relation to the idea of (in)justice. For more on this, *see*, e.g., Iris Marion Young, *Justice and the Politics of Difference*.

might actually address the real issues we face today. Oppression on the basis of race, gender, and disability, to name a few, is after all what motivates so many of us to begin imagining something better. And yet the ideal theorists engaged in constructing what Rawls calls a "realistic utopia" do not seem to give us much to go on for these specific injustices – they seem neither realistic for us *now*, nor utopian for us in the future given their lack of attention to the injustices that we have good reason to think will emerge if left unchecked.

This tension, however, can mask fundamental agreement between the two sides: not only that oppression is a paradigm example of injustice, but that a theory of justice fails if it cannot adequately explain oppression and prevent it from occurring in the social organization meant to be regulated by that theory's principles. In the previous chapter, I argued that one of the deepest moral harms of oppression was the erosion of moral *personhood* – not just the erosion of virtue, or one's moral capacities, but rather the erosion of the very moral status that we take to ground our various entitlements in social life. This phenomenon occurs as a result of an interpersonal failure to take up the morally relevant features of the individual in the right way. Contrary to the *inviolability* conception of moral personhood, which denies the possibility of this form of moral transformation, the interpersonal plays a large part in determining what kind of things we are. I argued further that this harm has an ontological dimension insofar as it makes us a different kind of being than what we have a right to become. In this sense, oppression reveals that one might have the potential to be a full moral person, an end in oneself, and yet become a mere means because of how one's life goes. This is *the contingency of moral*

personhood – something revealed by a close analysis of oppression but not confined to it insofar as the contingency is an enduring feature of the human condition.

In this chapter I will extend this moral-ontological insight into political philosophy, entering one argument (among many) between ideal theorists in the liberal tradition and those who wish to focus more explicitly on the questions of justice that considerations of oppression raise. For rhetorical purposes, I will select John Rawls's ideal theory of justice as our main ideal theory to consider, and I will select Martha Nussbaum's discussion of justice for the disabled as our foil. My main purpose, however, is not confined to remarks about Rawls or Nussbaum. Instead, my aim is for this discussion to reveal that the inviolability conception of moral personhood, which fails to account for the true moral harms of oppression, haunts political philosophy even in those moments that are explicitly concerned with getting it right about oppression. Insofar as Rawls and Nussbaum represent two prominent attempts to carry out this aim, I believe we can draw the larger philosophical conclusions I want from a more local analysis. I will argue further that Rawls's and Nussbaum's shared aim, i.e. to properly account for the injustice of oppression, is stymied in both cases by their failure to acknowledge the contingency of moral personhood. In other words, their shared implicit commitment to an inviolability view of moral personhood undermines their shared vision of the point of their theorizing: to properly account for and prevent one of the deepest injustices we face as humans.

The chapter will proceed as follows: I will begin with a discussion of two problems any political theory must address when drawing the boundaries of moral personhood. I refer to these as the "classification problem" and the "transit problem,"

both of which relate to the forms of moral status and moral transformation that are possible in a human life. These two problems are most pressing when considering oppression's effects on moral personhood, and the extent to which each of the above theories addresses them, if at all, will be the subject of the chapter.

I will analyze Rawls's social contract theory as presented in *A Theory of Justice*, and argue that his own Kantian conception of the person, which is a version of the inviolability conception of moral personhood discussed in Chapter 1, prevents him from adequately addressing either the classification problem or the transit problem. The result is that Rawls's theory of justice cannot fully account for the injustice of oppression, thus undermining the realistic utopian aim of his theory (captured so well in Rousseau's famous statement that we must take "men as they are and laws as they might be"[2]).

I will then shift to a discussion of Martha Nussbaum's critique of Rawls in *Frontiers of Justice*, and argue for three claims: first, that she correctly identifies Rawls's failure to solve the classification problem; second, that her capabilities approach solves the classification problem as it relates to the oppression faced by disabled people; and third, that Nussbaum nonetheless fails to recognize and account for the transit problem in her own theory. The transit problem is, I argue, essential to understanding the contingency of moral personhood and oppression's possible effects on it; as a result, Nussbaum's failure to address it undermines the realistic utopian aims of her theory, too. I trace this failure back to a reproduction of the same inviolability conception of moral personhood as found in Rawls, though Nussbaum's is more expansive and therefore more difficult to see. Finally, I will conclude with a discussion of the shared inviolability

[2] Jean-Jacques Rousseau, *The Social Contract and Discourses*, ed. G.D.H. Cole (New York: E.P. Dutton and Company, Inc., 1950), 3.

conceptions of moral personhood in Rawls and Nussbaum, and argue that any political theory that makes use of it will fail to solve the transit problem and therefore fail to properly account for one of the deepest injustices the theory is meant to solve.

2. Two Problems with the Boundaries of Moral Personhood

An important starting point for any theory of justice is fixing the class of beings with political *entitlements* in the right way – we need to be able to distinguish between those who have them, and those who don't. And this task needs to align with our considered intuitions about the question; it needs to be able to put e.g. pencils on the side of *no entitlements*, and Paris Hilton on the side of *entitlements*. Making matters more complicated, it also needs to align with our intuitions about the possible fluctuations in this status that happen over a human lifetime. There is a difference between children and adults, between cognitively capable adults and those with severe cognitive disabilities. As I argued in Chapter 1, there is also a difference between everyone mentioned so far and the victims of severe oppression – *these* are cases of beings with shifting entitlements during their lifetime, but it seems importantly different than the shifts that occur as a result of natural development or accident. Out of this morass, two problems related to moral personhood emerge:

1. **Classification Problem**. A theory of justice must be able to sort between those that have a moral status and those that don't, in a way that aligns with our considered judgments. It must also be able to make sense of the changes in moral status that occur as a result of a change in the individual feature that grounds a

being's moral status or non-status.

2. **Transit Problem**. A theory of justice must be able to make sense of the moral transformations that can occur as a result of oppressive interpersonal conditions, even where there is no change in the individual feature that grounds a being's moral status or non-status.

Each of these can be understood as related to one of the parts of the interpersonal conception of moral personhood argued for in Chapter 1. There, full moral personhood was a two-part matter: it requires both that the individual in question have some relevant feature (like rationality, or a capacity for suffering, etc.), *and* that this individual feature is taken up interpersonally in the right way. Since the actualization of moral personhood consists of these two-parts, so does its possible failure. The Classification Problem picks out failures related to the first necessary feature of moral personhood; the Transit Problem picks out failures related to the second. An adequate theory of justice must be able to solve both.

The Classification Problem should be familiar to the reader – it is an enduring question in political philosophy, and divergences in solutions often mark substantial divergences in the resulting theories. Kant famously picks out rationality as the relevant criteria for moral status; others pick out the less-stringent requirements of species-membership, subjectivity, or self-consciousness. And debates within contemporary philosophy (such as those mentioned in the introduction) are often debates around the Classification Problem. Rawls, most notably, has been criticized for drawing the

boundary of moral personhood too narrowly, and his myriad critics have suggested expansions on the basis of various features including disability, the capacity to care, and the capacity for a flourishing life. All are in agreement, though, that part of solving the Classification Problem will involve being able to explain fluctuations in moral status that go along with fluctuations in whatever individual feature is chosen as the relevant one. So the change in moral status that happens from an embryo to a baby, for example, is made sense of by reference to the shifts in individual capacity that happen along the way. An infant simply does not have the same e.g. rationality as an adult, and so it does not have the same moral status.[3] So too for the shifts from cognitive ability to severe cognitive *dis*ability. Our moral statuses change, but these changes are always a result of changes in *us*.

The Transit Problem is less familiar. In part, this is because of its basis in a conception of moral personhood that takes the interpersonal seriously as a metaphysical precondition. Such an account holds that our moral statuses can change not only due to changes in *us*, but also due to the interpersonal conditions in which we live. Oppression is one possible instance of these types of damaging interpersonal conditions. So the Transit Problem is really a problem about being able to make sense of the moral-ontological harms of oppression. Such conditions can erode our moral status, but intuitively speaking there is a difference between this type of change and the change of e.g. a severe brain injury. In each case, we start with a full moral person and end in something short of that; but from the point of view of justice we feel there is more to say. In part, that "more" comes from the fact that moral erosion due to oppression pries

[3] This is not to say that an infant has *no* moral relevance for theories of justice; only that certain forms of paternalism that would be appropriate for the infant would be inappropriate for an abled-adult.

moral status apart from features of the individual; the interpersonal intervenes in the moral rigidity of the Classification Problem and introduces a new way that we might undergo moral change.

3. Rawls and the Two Problems

Does Rawls's social contract theory as presented in *A Theory of Justice* adequately solve each of the above problems, especially as they relate to oppression? In what follows I will argue that, due to his own inviolability conception of the person, he does not. As a result, Rawls's theory cannot fully account for the injustice of oppression, thereby undermining the realistic utopian aims of his theory.

3.1. Rawls's Social Contract Theory and its Conception of the Person

What, if anything, does Rawls say about oppression? To the extent that oppression is at the opposite extreme of justice, Rawls does not take *A Theory of Justice* to be addressing it directly. He is concerned with ideal theory, and many features of oppression involve phenomena that he takes to be ruled out by his principles of justice. Interpersonal violence, for example, would be ruled out by the protection of the basic liberties, which include bodily autonomy. Insofar as exploitation is a transfer of value from one individual or group to another, with the former left worse off than they were to begin with, it should be ruled out by Rawls' difference principle – such inequalities of value are not to the benefit of the least well off position, and so they are unjust. The "faces" of oppression are matters of noncompliance in a Rawlsian view; and

noncompliance is to be worked out only *after* principles of justice for the regulation of the well-ordered society are determined.

Still, Rawls describes oppression as one of "the great evils of human history,"[4] holding that it will be eradicated by "following just (or at least decent) social policies and establishing just (or at least decent) basic institutions."[5] The two principles of justice, though they do not mention oppression explicitly, are meant to prevent all forms of social organization that would amount to it; oppression's injustice is one of the considered judgments that ground the justificatory process of reflective equilibrium. The protection of the basic liberties, in combination with fair equality of opportunity and the difference principle, are meant to regulate the basic structure in such a way that one's life chances are not determined by morally arbitrary features of the person. Oppression, insofar as it targets people based on morally arbitrary group membership, should be one of the paradigm injustices blocked in the well-ordered society.

These principles of justice are to be worked out through a project that he describes as "a theory of justice that generalizes and carries to a higher level of abstraction the traditional conception of the social contract."[6] He does this primarily through the construction of his "original position," taken to be a corollary of the state of nature in traditional social contract theories. This original position is a device of representation that is meant to justify a particular set of principles of justice to regulate the basic structure of society. Within it, rational agents are situated as free and equal, denied specific knowledge about morally arbitrary facts about themselves via the "veil of

[4] John Rawls, *The Law of Peoples* (Cambridge, MA: Harvard University Press, 1999), 6.
[5] Rawls, *The Law of Peoples*, 6.
[6] John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), 3.

ignorance," and then tasked with the selection of principles of justice in order to maximize their self-interest once the "veil" is lifted. On the question of how such an artificial and hypothetical choice situation could justify principles of justice to regulate the conduct of actual moral persons, Rawls says the following:

> The constraints imposed on the parties in the original position, and the manner in which the parties are described, are to represent the freedom and equality of moral persons as understood in such a society. If certain principles of justice would indeed be agreed to…then the aim of Kantian constructivism to connect definite principles with a particular conception of the person is achieved.[7]

The justificatory force of the original position, then, is rooted in its relationship to a particular conception of the person which Rawls himself takes to be a Kantian one.[8] What is the Kantian conception of the person, for Rawls? Rawls seems to draw upon the one found in the *Groundwork*:

> Now I say that the human being and in general every rational being *exists* as an end in itself, *not merely as a means* to be used by this or that will at its discretion; instead he must in all his actions, whether directed to himself or also to other rational beings, always be regarded *at the same time as an end*.[9]

It is a conception of the person as endowed with reason, the possession of which grounds their status as an end in themselves and requires that they be situated, in any political choice situation, as free and equal relative to other beings so-constituted. Without the faculty of rationality, one is a mere *thing* and therefore may be used for others' rational ends in accordance with the moral law. The distinction between an end

[7] John Rawls, "Kantian Constructivism in Moral Theory," *The Journal of Philosophy* 77, no. 9 (1980): 520.
[8] I am not taking a position on whether Rawls's conception of the person is in fact Kantian. Though I will go into Kant's conception of the person to some extent, this is in order to extrapolate on what Rawls takes his own view to be committed to rather than to commit myself to a particular Kantian interpretation.
[9] Immanuel Kant, *Groundwork of the Metaphysics of Morals* (London: Oxford University Press, 2012), 4:428.

and a mere means, then, draws a border in the moral landscape defined by the nature of the being in question and determining what sort of moral consideration they are due.

Each of the features of the original position, either with respect to the individual parties, the objects of their choice, or the epistemic constraints they are under, is meant to reflect one crucial aspect of the Kantian conception. Together, they make up the full Kantian conception of the person, one that "regards persons as both free and equal, as capable of acting both reasonably and rationally, and therefore as capable of taking part in social cooperation among persons so conceived."[10]

Indeed Rawls spends a great deal of time going into what he calls the "two moral powers" of such persons. These powers, our rationality and our reasonableness, are at the center of his own stated conception of the person. The former is modeled in the original position by the parties' knowledge that they have a conception of the good and their desire to advance it; the latter is modeled by the epistemic constraint that the veil of ignorance represents, in turn representing our "normally effective desire to apply and to act upon the principles of justice."[11] These are the capacities that ground Rawlsian moral personhood. And these are precisely what Rawls uses to draw the border between personhood and non-personhood. The location of this border should be familiar, as it is the Kantian one. He writes:

> The Reasonable presupposes the Rational, because, without conceptions of the good that move members of the group, there is no point to social cooperation nor to notions of right and justice, even though such cooperation realizes values that go beyond what conceptions of the good specify taken alone. The

[10] Kant, *Groundwork*, 518.
[11] Rawls, *Theory of Justice*, 505.

> Reasonable subordinates the Rational because its principles limit, and in a Kantian doctrine limit absolutely, the final ends that can be pursued.[12]

Reasonableness, for Rawls, is the dividing line of moral personhood, and it presupposes the rationality at the heart of the Kantian project.

It is important to note that Rawls specifies these two capacities as a sufficient condition for moral personhood and therefore for "being entitled to equal justice."[13] And he distinguishes between the mere possession of these capacities and their actualization. In Rawls's account, the capacity for the two moral powers, *not* its actualization, is the feature of persons that determines their moral status. He writes, "a being that has this capacity, whether or not it is yet developed, is to receive the full protection of the principles of justice."[14] Here in this distinction we find the inviolability of a Kantian moral personhood: however the world may go, it will not affect one's status as a moral person, because *that* status rests on a pre-social characteristic of the individual. If one is rational in the relevant sense, then one counts as a moral person. However that rationality is exercised in the world, or taken up in social spaces, is irrelevant from a moral metaphysical point of view.

Of note is that beings do not ordinarily cross this moral border (outside of the outlier cases of children and accidents), and in this sense the Kantian conception of the person on which Rawls relies contains an element of moral inviolability – if a being has the relevant rationality, a being has the relevant moral status. These two things do not come apart, and either they are both present or both absent (as in, again, the case of

[12] Rawls, "Kantian Constructivism," 530.
[13] Rawls, *Theory of Justice*, 505.
[14] Rawls, 509.

children and accidents). For example, a rational being cannot sacrifice his moral status as an end in himself through his own conduct. In discussing the suicide example, Kant presupposes that the violation of the duty to oneself that suicide represents does not, however, alter the fact that suicide targets an end in itself. Indeed, the very coherence of this act as a violation of duty requires the preservation of this gap: it is the gap opened up by the fact that our subjective wills are not objectively determined. The permanence of a rational being's location relative to the "border" I am describing, then, is a necessary condition for the possibility of morality. The coherence of the distinction between an end in itself and a mere means rests on the idea that bad action does not disenfranchise us from our inherent respectworthiness.

It is also the case in Kant that bad *treatment* does not disenfranchise us in this way. That is, being treated as a mere means does not make it the case that one *is* a mere means. Again, Kant is clear on this: "…he who transgresses the rights of human beings intends to make use of the person of others merely as means, without taking into consideration that, as rational beings, they are always to be valued at the same time as ends…."[15] Assaults on freedom and property are wrong precisely because they involve the types of beings who deserve otherwise. As in the above case, the inviolability of the victim's moral status preserves our ability to explain the wrongness of the conduct in question. It serves as a backstop against which human choice takes on a moral character.

[15] Kant, *Groundwork*, 4:430.

3.2. Rawls on the Two Problems

Does Rawls answer the Classification Problem in a way that aligns with our considered judgments about the proper place to draw these moral boundaries? Rawls uses the Kantian conception of the person to address this problem: rationality is the determining metric. I will argue in Section 4 that Nussbaum's critique of Rawls, which can be understood as an indictment of his solution to the Classification Problem, is successful. So I'll hold off discussion of this problem for now.

What I would like to highlight here though is that the Kantian conception of the person is in effect a denial of the Transit Problem, which must be addressed by any theory that adequately explains and prevents oppression. The Transit Problem states that a theory of justice must be able to make sense of the moral transformations that occur because of oppressive interpersonal conditions; Rawls's Kantian conception of the person rests on the presumption that this is not possible. So the theoretical commitment at the core of Rawls's political theory prevents him from acknowledging the Transit Problem. To this extent, then, Rawls cannot be thought of as answering the Transit Problem in any meaningful way. This itself is enough to undermine realistic utopianism. The following section explores this point at length.

3.3 Realistic Utopianism and Rawls

My claim is that the inviolability conception of moral personhood undermines the realistic utopian aims of Rawls's project, thereby undermining the justificatory force of his theory. Any theory of justice will of course have to make various preliminary assumptions in order to begin with a conceptually distilled account of personhood, one

which picks out certain morally relevant features and ignores a multitude of others. This sort of selective abstraction is a necessary feature of the practice. But this abstraction must be balanced against a competing aim of political philosophy: what Rawls calls the aim of describing a "realistic utopia." Such a project is utopian, in that we are thinking about justice; but it is also realistic, in that we are thinking about justice for beings like *us*. It is what is captured in the Rousseauian maxim for political philosophy, "taking men as they are and laws as they might be."[16]

Let us first consider whether the inviolability conception undermines the "realism" of Rawls's realistic utopianism. Realistic utopianism is, after all, a practical endeavor, one which aims to develop an account of the conditions and principles required for people to be treated as free and equal persons, *and* for this account to be action-guiding. Rawls writes, "We view political philosophy as realistically utopian: that is, as probing the limits of practical political possibility."[17] The conception of the person motivating any political philosophy defines, in part, that limit. Thus it is necessary that the conception of the person balance the need to make certain simplifying assumptions with the need to nonetheless preserve the features of the person which are relevant from a moral point of view. Theories of justice for humans do not, for the sake of simplicity, assume we are immortal, because our vulnerability to death is a crucial part of understanding the concept of wrongdoing that motivates the philosophical enterprise. Practical possibility, then, requires that we not be required to become something we are not in order to achieve it. The "realism" of realistic utopianism ensures that the theorized utopia is one that can serve as a practical end for real people.

---

[16] Rousseau, *The Social Contract*, 3.
[17] John Rawls, *Justice as Fairness,* ed. Erin Kelly (Cambridge, MA: Harvard University Press, 2001), 4.

As I have argued throughout, a necessary feature of our moral personhood is that it is contingent on the social and political circumstances of our environment. Moreover, this is a deeply important point about what it is like to be us: understanding the human experience requires understanding the fragility of the moral status which grounds our moral claims. Because of this, it is not one of the features of persons that can be abstracted away from while nonetheless maintaining a grip on what it is to be like us. *Realistic* utopianism, then, must include this moral contingency in its conception of the person if it is to generate a theory *for us*. Rawls's social contract theory, insofar as it relies on an inviolability conception of personhood, fails to achieve this requirement of realistic utopianism, because it fails to include this type of moral vulnerability in its founding assumptions. Its recommendations, then, whatever they may be, are rooted in a conception of a person that is too far from a morally adequate conception; as a result, its justificatory force is undermined.

Consider how Rawls's social contract theory fails to incorporate this moral vulnerability and the ways in which this undermines its justificatory force: As I've described earlier in this chapter, the original choice scenario is constructed with the aim of modeling the morally relevant features of persons and therefore producing principles that are connected to us in a moral sense. This connection is the source of its justificatory force. On this, Rawls writes:

> [O]ne conception of justice is more reasonable than another, or justifiable with respect to it, if rational persons in the initial situation would choose its principles over those of the other for the role of justice…Understood in this way the question of justification is settled by working out a problem of deliberation: we

> have to ascertain which principles it would be rational to adopt given the contractual situation.[18]

Furthermore, the structure of this contractual situation reflects certain reasonable assumptions about what would constitute a morally adequate framing for an original choice situation. These assumptions include the fact that the parties in the original position are equals; that "no one should be advantaged or disadvantaged by natural fortune or social circumstances in the choice of principles"[19]; and that the parties can each "make proposals, submit reasons for their acceptance, and so on."[20] Together, these conditions are taken to be reasonable insofar as they "represent equality between human beings as moral persons, as creatures having a *conception of their good* and capable of a sense of justice. The basis of equality is taken to be similarity in these two respects."[21]

What is a "conception of the good" on Rawls's account? He takes this to be a natural extension of our rationality – it is "the most rational long-term plan of life given reasonably favorable circumstances."[22] It is what is meant when people have aspirations to becoming a skilled painter, an expert seamstress, a learned philosopher. It is, simply, what someone wants to do with what the poet Mary Oliver calls our "one wild and precious life."[23] A conception of the good is a rational conception of value, and this serves as a life-long axis around which all other judgments of value for oneself are made. In this respect, Rawls's "conception of the good" is a kindred spirit with the idea of the

---

[18] Rawls, *Theory of Justice*, 17.
[19] Rawls, 18.
[20] Rawls, 19.
[21] Rawls, 19 (emphasis added).
[22] Rawls, 92.
[23] Mary Oliver, "The Summer Day," in Mary Oliver, *New and Selected Poems* (Boston, MA: Beacon Press, 1992).

good and its role in a happy life as found in Aristotle, Kant, and Sidgwick.[24] It is not merely an end; it is also a well-considered and well-tailored plan constructed in the light of all possibilities open, one that "permit[s] the harmonious satisfaction of [one's] interests"[25] and which cannot be improved upon. When satisfied, it is an essential element to human happiness.[26]

Self-respect is one of the most important resources one can have with respect to this feature of a life – it is integral to carrying out one's life plans. Rawls describes two aspects of self-respect:

> First of all…it includes a person's sense of his own value, his secure conviction that his conception of his good, his plan of life, is worth carrying out. And second, self-respect implies a confidence in one's ability, so far as it is within one's power, to fulfill one's intentions.[27]

So self-respect has elements of both self-worth and self-efficacy. Indeed without these two psychological features, it is unclear how anyone could have a robust and action-guiding conception of the good; as Rawls writes, "[w]ithout it nothing may seem worth doing, or if some things have value for us, we lack the will to strive for them."[28] Any rational being, then, given the fact that their rationality grounds (for Rawls) their having a conception of the good, will take self-respect to be one of the most important goods to have for the successful pursuit of their larger conception of the good.

Now – what does this have to do with justification? As I'm arguing so far, the justification of Rawls's social contract theory stems from the original position's

---

[24] Rawls, *Theory of Justice*, 92.
[25] Rawls, 93.
[26] Rawls, 409.
[27] Rawls, 440.
[28] Rawls, 440.

connection to the morally relevant features of persons who are to be subject to that social contract theory's political principles. One of the assumptions determining the structure of this original position is the idea that it "represent[s] equality between human beings as moral persons, as creatures *having a conception of their good* and capable of a sense of justice."[29] This, in turn, has the effect that "parties in the original position would wish to avoid at almost any cost the social conditions that undermine self-respect."[30] That is where we are so far.

The remainder of my argument will focus in on these last two features of Rawls's account: (1) that the parties in the original position know that they have a conception of the good, and (2) because of what they know about such conceptions generally, they will select principles of justice that ensure as much as possible the social bases of self-respect. I will argue that claim (1) is true, but structurally excludes certain types of moral harm from rational consideration within the original position; and that (2) is false because it is overbroad – the parties, because of (1), will instead select principles of justice that ensure *certain* social bases of self-respect will be maintained. The result of both of these related points is that the justificatory force of the original position is undermined: it does not represent *us* as moral beings enough to count as generating a *realistic* utopian conception of justice.

Let's look back at the film *Wanda*, extensively discussed in Chapter 1, and ask how its protagonist might fit into this conception of the person. To recall, Wanda Goransky is someone whom one critic described as "drift[ing] in the sea of her own

[29] Rawls, *Theory of Justice*, 19 (emphasis added).
[30] Rawls, 440.

insignificance."[31] She is a woman with no aims; no self-worth; so thoroughly crushed and shaped by patriarchy that she seems to be a mere vessel for the aims of the men around her. As I argued in Chapter 1, Wanda is one paradigm example of someone who has been so harmed by patriarchal oppression that her moral personhood is impacted – someone who lived her life without the social conditions that would generate her full status as an end in herself. The result, I argued, was that in an important sense she became a mere means, and this moral transformation is one of the deepest moral harms we can experience.

In no way does Wanda have what Rawls refers to as a "conception of the good." Her sense of time is incompatible with a "long-term plan of life" – she seems to be concerned with only the next meal, where she might kill a few aimless hours in the afternoon, where she's staying that evening. She's late everywhere she goes, despite having nowhere in particular to be. She also lacks a definitive idea of what would count as the "favorable conditions" under which she might pursue some sort of life plan – she knows that making money at a job would be nice, but this knowledge seems divorced from her day-to-day choices (e.g. she gives up her job hunt after only one rejection). She certainly has not identified *patriarchy*, or sexism or misogyny, as the source of her unexpressed misery, such that she would be able to imagine liberation from them as more favorable to her good. And she lacks an awareness of what would count as her unique interests, skills, or values. Simply put, she lacks a conception of the good.

She also lacks the Rawlsian self-respect that would be required if she *were* to have or develop a clear conception of the good. To recall, this type of self-respect consists of

[31] Reynaud, "For Wanda," 225.

both a sense of self-worth and self-efficacy. But two core "truths" orient her worldview and the film: first, that she "is just no good," which she exclaims to Mr. Dennis in a moment of exasperation; and second, that she "is too slow to be used" (this coming from the factory boss who rejects her only job application we see in the film), a failed mother and wife, someone who's not good at nearly anything. Later in the film, when she realizes the bank robbery has gone awry, Wanda's lack of self-efficacy is portrayed in stark relief: she hears the gunshots from outside and believes that her getting lost on the way to the bank is the reason Mr. Dennis got killed. Of course, as viewers we know that the robbery plot failed because of Mr. Dennis's own miscalculations; but Wanda does not, and for her this is one more addition to what is for her a mountain of evidence that she lacks the ability to successfully carry out any plans she takes on. Wanda has no self-worth; she has no self-efficacy; in a strange sense, she clings to these beliefs about what she lacks in order to make sense of her world. She would be the first to affirm, in likely a very matter-of-fact way, that she doesn't take herself to be a proper candidate for self-respect.

Returning to Rawls's theory of justice, we know that the parties in the original position both have and *know that they have* a conception of the good. Indeed it is this knowledge which compels them to be acutely aware of ensuring the social bases of self-respect, given self-respect's integral role in the successful pursuit of one's conception of the good. Rawls's veil of ignorance is meant to obscure from the parties' view any morally relevant facts about them which would bias their selection of the principles of justice. But I take the knowledge that one has a conception of the good to be morally relevant knowledge – it is not, as Rawls takes it to be, sufficiently abstract so as to

prevent them from favoring certain social positions once the veil is lifted. They know that they are not Wanda. What do I mean by this? More pointedly, the parties in the original position know, in virtue of knowing they have *some* conception of the good that they are keen to advance, that they have not suffered the sort of moral-ontological harm that is the subject of Chapter 1. They know that they are fully-actualized moral beings.

Because of this knowledge, combined with the fact that Rawls stipulates the parties in the original position choose principles of justice only out of rational self-interest, their focus on ensuring the social bases of self-respect will be restricted to a more narrow set of conditions than what Rawls describes. Rawls holds that the parties in the original position would "wish to avoid at almost any cost the social conditions that undermine self-respect."[32] But this need not be the case, strictly speaking. If I know (simply because I am in the original position) that I have not endured the moral-ontological harm to my moral personhood that e.g. Wanda endures, then I need not select principles of justice that ensure the social bases of self-respect even in these extreme conditions (without some further argument about perhaps the connection between these conditions and the ones I would be subject to). Instead, I will select principles of justice that preserve self-respect in the range of circumstances I might find myself in as a full moral person, and will be willing to sacrifice it in other cases if there is a more favorable tradeoff.

At this point a natural objection arises: don't the parties engage in counterfactual reasoning, and wouldn't this activity lead them to see that they would need to guard against the possibility of the erosion of their personhood? After all, they know that they

[32] Rawls, *Theory of Justice*, 440.

might get sick, and so they make sure that the basic structure is organized such that this will not severely impact their life chances. This counterfactual reasoning about health is perfectly compatible with their ideal project; it just means that they are thinking about what kind of being they are and selecting principles accordingly. In the same way they would be eager to ensure that there is a major social institution that supports the health of all citizens, they would also want to institutionalize whatever protections would be associated with their full moral personhood.

Nussbaum discusses the possibility of this type of self-interested counterfactual reasoning in Rawls. Though her argument is about whether the parties in the original position would select a sort of insurance scheme to guard against disability, it is apt for the question of guarding against the erosion of moral personhood, too. Following Amartya Sen, she argues that this sort of preventative theorizing "complicates…the use of primary goods, in particular income and wealth, to rank social positions"[33] – the same allotment of income will be less useful to someone whose disability requires them to spend a portion of it on securing basic human functioning than it will be for someone who has that ability for free, so to speak. This suggests that the parties in the original position would be inclined to shift to a capabilities-like metric in order to account for these expected variations in ability to turn one's income into flourishing. But Rawls selected income and wealth as a proxy for well-being precisely because he wanted to avoid the "intuitionistic balancing"[34] of non-fungible goods that a capabilities metric would require. Introducing the idea of counterfactual reasoning about, and principles

[33] Martha C. Nussbaum, *Frontiers of Justice: Disability, Nationality, Species Membership* (Cambridge, MA: Harvard University Press, 2006), 124.
[34] Nussbaum, *Frontiers of Justice*, 124.

guarding against, the disadvantages that come along with disability will destabilize the clarity in social choice that the metric of income and wealth provides.

This argument is easily transposed onto the idea of moral personhood – it is simply not the case that income and wealth (the metrics of the second principle) are good proxies for whether someone's moral personhood has been eroded by oppressive conditions. In part, this is due to the diverse possible manifestations of oppression that Iris Marion Young points out: plenty of wealthy women, for example, are nonetheless mere means in the Kantian sense because of their place in the patriarchal social order. Although oppression often comes along with decreases in income and wealth, and the statistics bear this out, it is not a perfect proxy. Further, increases in income and wealth will not solve the issues of eroded personhood in the same ways that it might for disabilities –being able to afford an adequate wheelchair might restore a robust list of capabilities, including autonomy, freedom of movement, and association. But increases in income do not provide the interpersonal conditions that are required to restore moral personhood. This suggests that the parties in the original position would again seek a metric that takes this into account, even if that metric involved the sort of intuitionistic balancing that Rawls is keen to avoid. But if we do this, then the Rawlsian project lacks the definitive ranking of the various political schemes available. It turns out that when the parties are considering which principles to include in order to safeguard the social bases of self-respect, they cannot engage in the counterfactual reasoning required to imagine the possibility of their moral personhood being eroded without introducing an element into Rawls's account that destabilizes the original position's justificatory force.

Rawls's first principle is not enough to guard against this problem, either. This principle is the one that guarantees the basic rights and liberties, including the liberty of conscience, freedom of association and speech, and so on. But recall Rawls is engaged in ideal theory, which he regards as a "strict compliance" theory concerned with a society in which "everyone is presumed to act justly and to do his part in upholding just institutions."[35] This means that oppression is not to be considered in the selection of the first principles – oppression is, if anything, a systemic violation of the basic liberties. In this respect, then, there is a disanalogy with the hospital example. The parties do engage in counterfactual reasoning, but constrained by the conceptual limits of ideal theory itself; accidents leading to disability are not necessarily a matter of noncompliance, so the parties are free to consider this possibility and guard against it in the first principle if it's possible. But oppression is by definition noncompliance. The result is the parties in the original position will not consider oppression (or its possible erosion of moral personhood) until after the principles are selected; before then, they will be willing to sacrifice certain social bases of self-respect if there is a more favorable tradeoff for them as full moral persons.

Still it may be the case that one is not convinced by this discussion and believes there are ways that the parties in the original position could anticipate and guard against moral-ontological harm. This reveals yet another difficult problem: Rawls's own picture does not have them consider this type of harm; they surely do not construct principles that explicitly guard against it. So if one is inclined to solve the problem of realism by having the parties in the original position contemplate the contingency of moral

[35] Rawls, *Theory of Justice*, 9.

personhood and select principles that protect it, one will have tacitly admitted that the Rawlsian picture actually offered fails to be utopian. That is, it would be an admission that when the parties consider their vulnerability in the right way, they would select principles of justice that go beyond what Rawls's justice as fairness provides. Here, too, we find that Rawls's realistic utopianism is undermined by his denial of the contingency of moral personhood, and with it his theory's justificatory force.

4. Nussbaum's Capabilities Approach and the Two Problems

Various critiques have emerged since the publication of *A Theory of Justice* arguing that it fails in one way or another to account for the full phenomenon of oppression. Martha Nussbaum's, in particular, focuses on the oppression of disabled people, animals, and the global poor, arguing that considering justice from these three perspectives shows us the conceptual shortcomings of Rawls's social contract theory. She pays special attention to the original position and its founding assumptions, and argues that its rigid conception of the person both leaves out certain morally relevant beings who should be included and generates principles of justice that cannot actually answer the pressing questions at the frontiers of justice. The following is an analysis of her critique of Rawls, followed by an argument that this can be understood as a critique of Rawls's proposed solution to the Classification Problem. Though her work focuses on three different categories of oppression, I will focus only on the question of disability in order to draw out my main points. After this discussion, I will turn to her capabilities approach and ask whether she solves the two problems with this alternate theory of justice. I will argue first that her capabilities approach does solve the classification problem as it relates to

disability; but that, second, she cannot solve the Transit Problem. Like Rawls, Nussbaum makes use of a modified inviolability conception of moral personhood and therefore cannot make sense of the moral ontological changes of oppression. The result is that, again like Rawls, the realistic utopianism of her theory is undermined, though for interestingly different reasons.

4.1 Nussbaum's Critique of Rawls

One of Nussbaum's starting intuitions is that the systemic mistreatment of disabled people is one of the many systems of oppression raising urgent questions of justice. "Until recently," she writes, "such people were simply not included in society."[36] Whether due to physical disability or mental, disabled people throughout history have been marginalized, stigmatized, often "hidden away in institutions or left to die from neglect" – they certainly were not considered equal, participating members of the public realm.[37] And yet, of course, disabled people are capable of flourishing lives, and this flourishing has moral significance insofar as it is grounded in what Nussbaum refers to as "the dignity of the human being, and of a life that is worthy of that dignity."[38] The extent to which social and political conditions impede that flourishing is therefore a matter of justice.

The widest articulation of Nussbaum's critique of Rawls is to say that she finds a contradiction between one of the founding intuitions of his project – that "each person possesses an inviolability founded on justice that even the welfare of society as a whole

[36] Nussbaum, *Frontiers of Justice*, 15.
[37] Nussbaum, 15.
[38] Nussbaum, 74.

cannot override"[39] – and the structure of the original position that is meant to generate principles that embody and safeguard this founding assumption. This contradiction is brought out by the consideration of disability, and in a sense her major claim is that Rawls does not live up to his own founding intuition about moral inviolability as it relates to the lives of disabled people.

This failure manifests in two ways: first, to recall, Rawls assumes that parties gather in the original position because they recognize that they can gain more from social cooperation than they can in its absence, but they are not indifferent as to how the benefits of this cooperation are to be distributed; his original choice situation is, like others in the social contract tradition, conceived of as one generating principles to regulate citizens who can enter into relations of mutual advantage. And yet, Nussbaum argues that the assumption of mutual advantage prevents the parties in the original position from adopting principles that acknowledge disabled people's "needs for special social attention, for the reasons of social productivity and cost that pertain to all people with impairments."[40] Not every interaction with a disabled person will be one of mutual advantage, and they may very well need a greater proportion of care and resources than a "normal" citizen does over a lifetime (although the fact that even "normal" citizens pass through the care-intensive stages of childhood and, if they're lucky, old age puts pressure on the idea that even "normally functioning" citizens are not vulnerable and needy in similar ways). But those choosing from within the original position are choosing for situations of mutual advantage. The resulting principles, then, will not be adequate to the

[39] Rawls, *A Theory of Justice*, 3.
[40] Nussbaum, *Frontiers of Justice*, 135.

need for certain principles that the consideration of disability reveals – disabled people are excluded from the set of people for whom the principles are chosen.

They are also, more troublingly, excluded from the set of people *by whom* the principles are chosen. In this way, Rawls fails to live up to the founding intuition about inviolability in a deeper sense. Nussbaum writes:

> People with mental impairments and disabilities…are disqualified from citizenship in a deeper way as well, because they do not conform to the rather idealized picture of moral rationality that is used to define the citizen in the Well-Ordered-Society.[41]

Rawls's original position, as I discussed in Section 3, is modeled after a conception of the person that treats rationality as the sole morally relevant factor in determining personhood. Indeed he holds that the justificatory force of the original position rests in its correspondence with this Kantian conception of the person. But this stipulation about what it takes to be qualified to enter the original position excludes cognitively disabled people whose mental capacities do not surpass the relevant threshold for moral personhood on the Kantian picture. This means that "mentally disabled humans…fail to be persons in the requisite sense."[42] They are structurally excluded from the original position, their claims relegated to claims of charity, not justice, to be considered only after the principles have already been selected. But Nussbaum notes that ensuring a flourishing life for the disabled is not merely a matter of supererogation; they, like all abled people, have rightful claims in this domain. Amending Rawls's original position to reflect this is no easy task, though: loosening the model of rationality will impact the

[41] Nussbaum, *Frontiers of Justice*, 135.
[42] Nussbaum, 332.

original position's justificatory force; it will also "[sever] ties with the classical social contract tradition"[43] insofar as it would include people who cannot be situated in relations of mutual cooperative advantage. If we are to maintain the justificatory force of the original position, then, as well as its ties to social contract theory, we are left with a theory whose core structure regards disabled people as second-class citizens.

In this way, we can understand Nussbaum as rejecting Rawls's use of the Kantian conception of the person to solve the Classification Problem. Rawls's view excludes disabled people from both the set of people *for whom* principles are chosen and *by whom* principles are chosen. But this conflicts with our considered judgments about the dignity of disabled people, according to Nussbaum. On the relevance of these considered judgments, she writes:

> Devotees of procedural approaches typically feel squeamish about the naked appeal to the idea of human dignity in the capabilities approach, in a way that they do not feel squeamish about the similar role of an idea of human inviolability and a related intuitive idea of respect for persons in Rawls's theory—simply because there are so many moving parts in between that intuitive idea and the final output that one can fail to notice how much work these intuitive ideas are doing.[44]

So to the extent that one worries about the role of considered judgments about disability, Nussbaum holds that this is a shared feature of the two views. As a result, it is not on its own a reason to reject her critique. If Rawls were to try to meaningfully incorporate this critique, though, Nussbaum claims he would have to give up too much – such a move would estrange him from the social contract tradition of which he is a part, including his

[43] Nussbaum, *Frontiers of Justice*, 119.
[44] Nussbaum, 174.

assumption of mutual advantage. It would also undermine the justificatory force of the original position.

4.2. Nussbaum's Proposed Solutions

Nussbaum views these twin shortcomings as an essential feature of the social contract tradition as presented by Rawls, especially in his reliance on a Kantian conception of the person to solve the Classification Problem. She therefore rejects the idea of an amendment to Rawls's project and instead offers a solution in the form of a pivot: her capabilities approach can be thought of as her attempt to respond to Rawls's same founding intuition about moral inviolability with a theory that is nonetheless not encumbered by his theoretical commitments. In this way, she would be able to solve the Classification Problem with a metric that accords with our intuitions about the dignity of disabled people. Nussbaum writes:

> The capabilities approach takes its start from the Aristotelian/Marxian conception of the human being as a social and political being, who finds fulfillment in relations with others.[45]

Where Rawls is a Kantian, Nussbaum is an Aristotelian/Marxian. By this is meant she focuses not on one particular feature of the human that, historically, has been thought to be the dividing factor between mere means and ends. Rather, she focuses on the totality of a life and the species-specific flourishing that that life is capable of given the right conditions. "We begin with a conception of the dignity of the human being, and of a life

[45] Nussbaum, *Frontiers of Justice*, 86.

that is worthy of that dignity."[46] We then move on to ask which diverse capabilities are required for this sort of flourishing and critically assess any political arrangements to see whether they safeguard these essential capabilities. The capabilities approach is in this respect quite different from Rawls's pure procedural justice, which leans heavily on the conception of the person on which the procedure is based in order to ensure just outcomes. Rather, Nussbaum's capabilities approach

> …starts from the outcome: with an intuitive grasp of a particular content, as having a necessary connection to a life worthy of human dignity. It then seeks political procedures…that will achieve that result as nearly as possible…. Justice is in the outcome, and the procedure is a good one to the extent that it promotes this outcome.[47]

Her approach therefore reverses the justificatory relationship between political procedures and principles of justice with respect to founding intuitions about human dignity – Rawls begins with intuitions about the inviolability of persons and constructs a *procedure* that reflects this inviolability; whatever principles of justice result from the procedure will, derivatively, be justified. Nussbaum, on the other hand, connects her starting intuition about dignity to the *principles* of justice, ensuring that they reflect this insight about the totality of a human life. Only then does she move on to procedures, and the procedures are derivatively justified only insofar as they can generate these independently selected principles with reasonable reliability.

The capabilities approach is thus unencumbered by the strict conceptual commitments of the social contract tradition and its original choice situation. According to Nussbaum, this allows it to generate principles of justice that include disabled people

[46] Nussbaum, *Frontiers of Justice*, 74.
[47] Nussbaum, 82.

as a matter of right, answering the Classification Problem in a more satisfactory way. It does so primarily through its Aristotelian/Marxian foundation – by focusing on the totality of a life, rather than only one element of it, it expands the group of beings who will count for the theory as having the requisite dignity for political consideration. Human life exists on a wide spectrum of differing levels of productive capacity, capacity for reciprocity, rationality, and for all of the other functions characteristic of the species. Disability is but one manifestation of this fact about us, as the cases of childhood and old-age reveal. Nussbaum's considered intuition about this fact is that each life, wherever it is on the spectrum, has a dignity that grounds claims of justice. By widening the scope of human dignity, she argues she is better able to respect the Rawlsian insight that each person has an "inviolability founded on justice that even the welfare of society as a whole cannot override."[48]

I regard Nussbaum's capabilities approach as a successful amendment to Rawls's solution to the Classification Problem. It begins from a reasonable considered judgment – that disability itself should not disenfranchise a person from the moral community. And it incorporates that considered judgment in a way that avoids the problems in Rawls that Nussbaum was right to point out; doing so was too theoretically costly for Rawls, because it required giving up on justification *and* on its contractarian founding assumption of mutual advantage. Nussbaum has no such contractarian assumption, so the intuition about disability does not pose any conflict. And the justification of her view as it relates to the Classification Problem is rooted in the *intuition* about disability, thus bypassing the justificatory problems that arise for Rawls's political *procedure* once its

[48] Rawls, *A Theory of Justice*, 3.

conception of the person is shown to be inadequate. The capabilities approach therefore marks a distinct advance in thinking about the Classification Problem as it relates to questions of justice and categories of oppression.

4.3 Nussbaum and the Transit Problem

Does Nussbaum's capabilities approach also solve the Transit Problem as it relates to disability? That is, does she account for the ways that the oppression of disabled people can undermine their moral personhood? In this section I will argue that she does not. This argument will proceed in stages: first, I'll discuss the concepts of the right and the good in the capabilities approach as they relate to her Aristotelian/Marxian conception of the person. I'll argue that this conception of the person gives us two "options" with respect to where we locate full moral personhood: either in a flourishing life, or in mere species membership regardless of the extent of flourishing. Second, I'll argue that each option poses serious difficulties for her view – locating it in a full flourishing life prevents her from making sense of differences among failures to flourish that reveal important differences between the right and the good, while locating it in species membership distorts the relationship between sociality and moral personhood. As a result, Nussbaum's conception of the person, which can be understood as having the same inviolability conception as Rawls, prevents her from responding to the deepest harm of oppression.

As I said above, Nussbaum is to be applauded for bringing in considerations of vulnerability to the Rawlsian account; in this respect, her work is a major advance from the Rawlsian conception of the person and its too-narrow understanding of the ways in

which vulnerability and interdependence suffuse human life, not as an accident but as an integral feature of what kind of beings we are. She does this through the use of an Aristotelian/Marxian conception of the person that expands the scope of dignity beyond the Kantian conception to included disabled people. The disabled have an inherent dignity, and the oppression of the disabled cannot change this fact. Indeed it is the very immutability of their dignity that Nussbaum uses to ground claims about the objectionable nature of their treatment. She locates this dignity in our animal nature:

> [Our] dignity just is the dignity of a certain sort of animal. It is the animal sort of dignity, and that very sort of dignity could not be possessed by a being who was not mortal and vulnerable, just as the beauty of a cherry tree in bloom could not be possessed by a diamond.[49]

The disabled and the nondisabled both possess animality just in virtue of being born into the species;[50] on Nussbaum's account, then, they both possess dignity, because this feature of the individual is a sufficient condition for having dignity. And since our lives cannot be understood apart from our animality, there is no sense in which we could lose the dignity that is grounded in it. In this sense, there is an element of inviolability to her view: we cannot lose our inherent dignity. But this is fully compatible with the idea that we are vulnerable and needy beings with bodies, susceptible to the ordinary influences of time, the world, and other people as we "grow, mature, and decline."[51] Inviolable dignity and animality are not opposed; indeed, they are (like the beauty of a cherry tree in bloom) best understood in relation to one another.

---

[49] Rawls, *Theory of Justice*, 132.
[50] See Rawls, 347.
[51] Rawls, 132.

This conception is, because of its foundation in our animality, thoroughgoingly social. She rejects the Kantian "idea that the core of our personality is self-sufficient rather than needy,"[52] instead describing the human as "a being who wants and needs to live with others."[53] Her description of the conception of the person in her capabilities approach reflects this:

> The capabilities approach takes its start from the Aristotelian/Marxian conception of the human being as a social and political being, who finds fulfillment in relation with others.[54]

The result is a list of protected capabilities that are almost all intertwined with social elements, including the ability to form attachments, engage in reciprocal interaction, and participate in political life. She gives priority of place to two of these capabilities in particular, one of which is the capability of affiliation: where one is "able to live with and toward others, to recognize and show concern for other human beings, [and] to engage in various forms of social interaction."[55] These are the capabilities that are required for someone "who seeks a good that is social through and through."[56]

So her conception of the person balances the inherent dignity of our animality with an understanding that this dignity is not fully realized unless it includes a thoroughgoingly social life. Where does moral personhood fit into this? That is, where does the concept of *right* fit in? On the topic of the right and the good, she has this to say:

---

[52] Rawls, *Theory of Justice*, 132.
[53] Rawls, 273.
[54] Rawls, 86.
[55] Rawls, 77.
[56] Nussbaum, *Frontiers of Justice*, 158.

> The capabilities are not understood as instrumental to a life with human dignity: they are understood, instead, as ways of realizing a life with human dignity, in the different areas of life with which human beings typically engage…Dignity is not defined prior to and independently of the capabilities, but in a way intertwined with them and their definition. (Of course the architectonic capabilities of sociability and practical reason play a guiding role throughout, as they do for the young Marx…). The guiding notion therefore is not that of dignity itself, as if that could be separated from capabilities to live a life, but, rather, that of a life with, or worthy of, human dignity, where that life is constituted, at least in part, by having the capabilities on the list. *In this way, the right and the good seem thoroughly intertwined.*[57]

As I have been using the concept of moral personhood throughout this dissertation, I take it to be a marker of *right*; it specifies the entitlements of the being in question. And though I have included the social as a precondition to the actualization of this right, I have not taken a stand on sociality's role in the *good*. That is, I have not committed myself to the idea that the good life is social; only that a certain social recognition is required in order to be a being for whom questions of the good life take on a moral character. So the right and the good have been regarded as separable in my arguments.

Contrast this with Nussbaum's explanation above: "In this way, the right and the good seem thoroughly intertwined."[58] A natural reading of this passage is to hold that, for Nussbaum, moral personhood is only fully actualized when both the right *and* the good have been satisfied; i.e., when a person's capabilities have been satisfied and they are in the midst of a flourishing life. This is the first "option" presented to us by her conception of the person that I mentioned at the outset of the section: moral personhood is located in this normative tangle. But there are serious problems with this interpretation, and they have to do with how this idea of moral personhood could make

[57] Nussbaum, *Frontiers of Justice*, 161-62 (emphasis added).
[58] Nussbaum, 161-62 (emphasis added).

sense of failures to flourish that have to do with oppressive conditions. The following is a discussion of these problems. After this discussion, I will consider whether we can save the concept of moral personhood in Nussbaum's view by grounding it in dignity itself, the second "option," without mention of flourishing. I will conclude that we cannot, because of separate but equally serious problems that arise on that interpretation. Either way we go, then, my argument is that Nussbaum's view cannot account for changes in moral status due to oppression and therefore cannot answer the Transit Problem.

Let's call the first option the "Combined Interpretation." How does the Combined Interpretation fare? When moral personhood is intertwined with both dignity *and* flourishing, there will simply be more instances of failures to attain full moral personhood. But among these more numerous failures, the Combined Interpretation cannot distinguish between cases that seem, intuitively, importantly distinct. Consider two cases:

> **Wanda**, who has no conception of the good, is abused by men and has few if any meaningful social connections based on mutual respect or even affection. Her life is generally terrible, inconsistent with the dignity of a human being.

> **William**, who has a conception of the good and budding social relationships, but is unable to pursue his life's dreams or deepen his social relationships because of an early death from a heart attack.

Both Wanda and William are the types of beings who could, if given the right conditions, flourish. They both fail to flourish though because something has gotten in the way. The result is that both of them fail to achieve full moral personhood on the Combined Interpretation.

But there is an important difference between Wanda and William: both of them involve failures of the good, but only Wanda's case involves a failure of right. Wanda's failure to achieve a flourishing life is due to the injustice of patriarchal oppression, and she has a rights-based claim against this sort of impediment. William's failure, on the other hand, is not something against which he has any sort of rights-based claim; it is in the realm of personal tragedy. No doubt, each is painful; but there is an additional political significance to Wanda's failure to flourish that is not present in William's situation.

The Combined Interpretation, however, cannot make sense of this difference. Where the right and the good are intertwined, the relevant fact for moral personhood is *whether* there has been a failure to flourish, not what the *cause* is of that failure to flourish. The result is that the Combined Interpretation of moral personhood flattens out distinctions among causes that are politically significant. In particular, it cannot see as a unique phenomenon those failures to flourish (and therefore failures to achieve full moral personhood) that are a result of injustice. This includes the injustice of oppression. So it cannot answer the Transit Problem, given that this problem is one that demands that a theory of justice be able to make sense of the changes in moral status that can occur as a result of oppression. So much for the first interpretation, then.

Is Nussbaum's conception of the person able to make sense of these moral-ontological changes if we favor the second interpretation? That is, if we scale the view back and ground moral personhood in mere dignity, leaving questions of flourishing aside, would we be able to answer the Transit Problem? On this interpretation, which we can call the "Dignity Interpretation," the answer is still no: multiple problems arise that have no easy answers.

The first problem with the Dignity Interpretation is that it requires us to have a distorted account of the explanatory relationship between sociality and moral personhood. On this interpretation, the social enters into the picture as a matter of the *good* and not of right. That is, despite the importance that Nussbaum gives to the social, in defining both what it is to be the kind of thing we are and which capabilities are necessary for our flourishing, on the Dignity Interpretation it comes in only after the rightful entitlements have been established – it is a way of further specifying what it means to live a "life that is worthy of [human] dignity."[59] Our dignity is conceptually prior, and an inquiry into the dignity *of which kind of being* tells us that sociality will be an important part in constructing the conditions that respect it.

Contrast this with my contingency of moral personhood view: there, the social is a *precondition* for us having the sort of dignity that would ground rightful claims. Of course, it may also turn out that sociality is also a part of our good – a flourishing life will likely include many of the actualized capacities that Nussbaum enumerates. But the important conceptual point is that this flourishing life takes on a moral character only when the social interacts with the individual in such a way as to generate a fully

[59] Nussbaum, *Frontiers of Justice*, 74.

actualized moral *person*. It is imperative that the individual, marked out by whatever individual feature is selected as the qualifying criteria for potential moral personhood, is taken up interpersonally in the right way. So the social enters on the ground floor of personhood, so to speak, rather than after personhood is up and running.

My first chapter is an extended argument in favor of the interpersonal conception of moral personhood, so it should come as no surprise that I reject the Dignity Interpretation – it confines moral personhood to the pre-social dignity inherent in us in virtue of our species membership. And because the social only comes in *after* moral personhood has been established, it is again a version of an inviolability view. In this respect, it contains the same moral rigidity that is found in Rawls's Kantian conception of the person. The Dignity Interpretation stops short of acknowledging our deepest vulnerability, which is that our social lives may go in such a way that even our dignity itself is impacted. Disabled people are brought into the realm of moral personhood, so long as oppression has not been *too* bad. This is unacceptable for a theory of justice that seeks to adequately capture the true extent of oppression's possible harms.

Neither interpretation of the concept of moral personhood in Nussbaum's account can adequately solve the Transit Problem, then. The Combined Interpretation leads to an inability to distinguish among failures to flourish that are due to injustice; but the pared-back Dignity Interpretation leads to a complete erasure of those instances of oppression that result in a loss of moral personhood. In both cases, we can see the inviolability conception of moral personhood at work, despite Nussbaum's attempt to distance her conception of the person from Rawls's.

To what extent, though, is this really a problem from the perspective of justice, especially justice for the oppressed? It may not make much of a difference as it relates to Nussbaum's actual recommendations – all of the capabilities she lists are ones that seem likely to eradicate oppression if implemented properly. Two worries should give us pause, though: first, Nussbaum considers the list of capabilities "as open-ended and subject to ongoing revision and rethinking."[60] She introduces this flexibility as part of her commitment to respect for pluralism. But this flexibility also introduces the possibility that the list of capabilities might change in different social and political contexts, and these changes might let certain oppressive conditions through the cracks. Whether this is a live issue, though, is an empirical question.

A second worry is the worry about where the boundary of moral personhood is drawn. We know that her notion of dignity is much more expansive as compared to e.g. Rawls's. But it is one that ignores the possibility of dignity's erosion under oppressive conditions. The result is that there is at least the theoretical possibility that someone on the other side of the moral border will fail to count as a being with dignity, and therefore fail to have an entitlement to the list of capabilities despite them being someone whom the capabilities approach is in spirit meant to protect. This would be a serious problem from the perspective of justice. Again, this is an empirical worry subject to seeing how things actually go. To be fair, it is probably very unlikely that this situation could come about – that the capabilities would all be satisfied for e.g. mildly disabled people who count as full moral persons, but simultaneously completely absent for severely disabled people who constitute an oppressed class and have had their moral personhood eroded.

[60] Nussbaum, *Frontiers of Justice*, 78.

But it is at least not an incoherent possibility, and it is one that Nussbaum's approach does not rule out from the start.

Even if these two worries turn out to be without merit, there is nonetheless what I take to be a conceptual error at the core of Nussbaum's account if we favor the Dignity Interpretation: the explanatory relationship between sociality and moral personhood is reversed. I view this as a problem for any theory of justice. It is an error that makes us unable to see the deepest moral-ontological harm of oppression; it is also an error that misrepresents the nature of personhood itself insofar as it denies that this type of vulnerability, too, is part of the human condition. I take it to be a reasonable desideratum of a theory of justice (among others) that it correctly describe the morally relevant features of our human nature. Especially in the case of a theory of *justice*, it should be able to correctly describe the full spectrum of ways that we can be harmed in interpersonal contexts; this is the only way that a theory will have a chance of developing principles that can prevent it. So while Nussbaum's list of capabilities will likely overlap substantially with the recommendations of whatever theory of justice *does* represent our human nature in the right way, and therefore these problems will likely remain in the theory and not in its practical application, I think as moral and political philosophers we should strive to reflect ourselves in the right way, if only for theory's sake.

4.4 Nussbaum and Realistic Utopianism

Because neither interpretation of moral personhood in the capabilities approach can account for the full scope of oppression, we find that the realism in Nussbaum's realistic utopianism is undermined. The capabilities approach obscures the possibility of

moral-ontological harm under oppression. So while her account makes strides towards a more realistic account of moral personhood than Rawls's by incorporating an Aristotelian/Marxian perspective, in the end there is one final human vulnerability that is not accounted for. My argument is that this vulnerability – the vulnerability to moral-ontological harm – is essential to the types of beings we are. As a result, Nussbaum's *realistic* utopianism is undermined.

Interestingly, for reasons discussed above, the utopianism in Nussbaum's capabilities approach does not seem to be undermined in the same way. The capabilities she lists are comprehensive and aimed at a flourishing life. And it is likely this focus on *flourishing*, rather than merely establishing a floor below which people cannot be allowed to descend, that makes it the case that the view remains utopian despite having the inviolability conception at its center. "[We] seek a higher threshold, the level above which not just mere human life, but *good life*, becomes possible."[61] The capabilities approach is therefore one that is concerned with "truly human functioning"[62] in the fullest sense of that idea. And because it ensures this flourishing via an outcome-based approach, eschewing the proceduralism of Rawls's theory of justice, we are able to directly check these outcomes against our considered intuitions about justice. I believe this more direct connection (again, contrasted with the many steps of a procedural account) makes us more sensitive to the full contours of a flourishing life. The list of capabilities, modified according to the plural demands of our diverse human community, seems genuinely utopian.

---

[61] Nussbaum, *Frontiers of Justice*, 181.
[62] Karl Marx, *Economic and Philosophic Manuscripts of 1844* (New York: Prometheus Books, 1988), quoted in Nussbaum, *Frontiers of Justice*, 74.

In this sense, Nussbaum's capabilities approach makes progress in the realm of justification, because it makes progress in its utopianism. Unlike Rawls's view, which I argued earlier would have to include additional principles if the parties in the original position considered the contingency of moral personhood, Nussbaum's does not obviously need additional principles to prevent the deepest harms of oppression. So the view matches our considered judgments about justice at least to this extent.

Still, even if utopian, it will lack a robust connection to *us* at the outer limits of our moral personhood. But this is the liminal space – the space between being and not being a moral person – that most requires principles of justice. I believe we have strong intuitions about what people in that space are entitled to, including the restoration of interpersonal conditions adequate to the task of repairing and actualizing their moral personhood. Being able to articulate these entitlements, even if one's theory is focused more on articulating the intertwining nature of the good with the right, is a crucial component of justification.

5. Final Remarks on Inviolability and Realistic Utopianism

I hope to have shown by way of two prominent examples that inviolability conceptions of moral personhood are present in even those political theories that attempt to get it right about oppression specifically. Furthermore, I hope to have shown that this inviolability conception of moral personhood undermines an important aim of political philosophy about which we have strong considered moral judgments – we think that if a theory of justice is to do anything, it should be able to properly account for *injustice* (even if only by way of contrast). Rawls and Nussbaum share this aim explicitly; I

believe it is also implicit in many others' thinking about justice. Nussbaum's critique, then, is not of the aim but of Rawls's method of achieving it. She focuses on various forms of oppression, in all of their complication and messiness, in order to advance the spirit of the Rawlsian project with a solution that has a greater likelihood of success (defined as producing principles that actually address these systemic forms of oppression). But by preserving the inviolability conception of the person from Rawls, even in its expanded form, Nussbaum detaches her theory from the people to whom it must be justified. This conceptual error, i.e. the tacit denial of the contingency of moral personhood, weakens the view and opens it up to critique from those for whom a theory's justification is of central importance. More generally, the inviolability conception of the person will stymie any political philosophical project that aims to fully account for and prevent oppression, because it moves on to solutions before getting a grasp on the extent of the potential problems. My hope is that this extended analysis will incentivize political philosophers to incorporate a modified conception of the person in which the interpersonal has moral-ontological significance if they are to produce justified theories of justice that respond appropriately to the injustice of oppression.

CHAPTER 3
REVENGE

"One does not hate a hailstorm or a plague, one hates only men, not because they are material causes of material damage, but because they are conscious authors of genuine evil."

– Simone de Beauvoir

"And if you wrong us shall we not revenge?"

– Shakespeare, *The Merchant of Venice*

1. Introduction

In describing the political transformation that takes place at the end of Aeschylus' *Oresteia*, Martha Nussbaum writes, "Athena introduces legal institutions to replace and terminate the seemingly endless cycle of blood vengeance."[1] And this replacement is a successful one precisely because the orderliness of a criminal justice system, coupled with a "profound inner reorientation" of the once-bloodthirsty Furies, is meant to remove any *reason* one would have for seeking private vengeance. As Nussbaum writes, "The city is liberated from the scourge of vindictive anger, which produces civil strife and premature death. In the place of anger, the city gets political justice."[2] The construction of a legal

[1] Martha C. Nussbaum, *Anger and Forgiveness: Resentment, Generosity, Justice* (New York: Oxford University Press, 2016), 1.
[2] Nussbaum, *Anger and Forgiveness*, 3.

system that publicly adjudicates disputes on behalf of private parties gives wronged parties something to look towards for the first time – political justice. It is a political framework that shifts the balance of self-interest away from vengeful impulses.

This conception of the relationship between the state and private vengeance is not unique in the history of political philosophy. Indeed, we can find versions of it in almost every major strand. The social contract tradition incorporates this insight into its own political genealogy, by framing the social contract as that agreement which most benefits parties in the state of nature, precisely by unburdening them from both the threat of private violence and the need to pursue it for one's own benefit. And utilitarianism, at least in its later forms, can be interpreted as resting on this central point about the role of justice. If it is truly the case, as it is in *Eumenides*, that the eradication of private vengeance promotes the greatest happiness for the greatest number of people, then a utilitarian political framework would hold that citizens have *reason* to pursue such an organization.

There is a cluster of enduring assumptions in much of this Western political philosophical inheritance: first, that vengeance is a moral evil incompatible with freedom; second, that it is irrational; and finally, that vengeance has no place in our practical lives once sociopolitical organizations are up and running. Nussbaum herself explains the point about irrationality in describing anger's vengeful impulses as involving beliefs that are "false and incoherent, ubiquitous though they are."[3] But these assumptions, as I will argue, oversimplify questions regarding the status of revenge under conditions of political failure. What can wronged parties do when the state no longer adequately

[3] Nussbaum, *Anger and Forgiveness*, 4.

safeguards their moral personhood? When it seems that private vengeance is the only recourse one has in the aftermath of immense interpersonal harm?

In this chapter I explore these questions and develop an account in which revenge, conceived of in a particular way and under particular political conditions, is rational to pursue. This rationality is grounded in the moral good that it provides revenge-seekers: the actualization of a moral personhood that was eroded by the interpersonal conditions that constituted the state's failure. While I do not commit myself in this chapter to defending revenge *all things considered*, I make the still-robust claim that revenge provides *a* moral good and it is therefore rational to pursue it. This good is related to the more general good of actualizing one's moral personhood. In this way, I deny the first two assumptions of our philosophical inheritance, and I show that the last assumption – that revenge has no place in our practical lives once sociopolitical organizations are up and running – is a tautology once we understand the nature of revenge itself.

The chapter will proceed systematically, with each section representing one stage of the argument. I will consider a different philosophical view on the topic of revenge in each section – these views provide a sort of philosophical foil, a contrast against which my view of the rationality of revenge becomes clear (rather than a comprehensive taxonomy of the major views on the topic). The argument itself proceeds as follows:

1. As long as it is possible that revenge's successful pursuit can restore the moral personhood of the revenge-seeker, revenge is not necessarily irrational.

2. It is possible that revenge can be carried out successfully.
3. Successful revenge does indeed restore the moral personhood of the revenge-seeker.
4. We are not self-deceived about claim (3).
5. *Therefore* revenge is rational insofar as it aims at the restoration of the revenge-seeker's moral personhood.

Before we begin this argument, though, it is important to distinguish among a number of related concepts – this will allow us to have revenge as a distinctive phenomenon in view for the argument, delimited from kindred concepts like retaliation and retribution. The following section does this clarificatory work; the formal argument begins in Section 3.

2. Clarifying Revenge and its Relation to the State

What, exactly, is revenge? The revenge-seeker is, of course, looking for justice. They are someone who has been wronged, and they are in pursuit of some outcome that they believe will set that wrong "right" in some important way. Crucially, they believe that it is important that *they* be the one setting this right. There is an acute sense of disappointment for that revenge-seeker whose target dies before they can be reached – especially so where their death is a result of some non-moral event (like a spontaneous heart attack). The target has "gotten away" with something, from the perspective of revenge, where their negative life outcomes are the result of chance rather than others' intentional action.

It is another important component of the will towards revenge that it arises in conditions of political failure. Again, the paradigm revenge-seeker is someone whom institutions of accountability have failed – there were no formal consequences for the wrong that was done. In the most severe cases, this lack of accountability reaches to the legibility of the wrong itself, where the community in which it took place does not recognize it *qua* wrong. So the motivation for revenge arises where there is neither accountability for nor recognition of an egregious wrong. It is the desire to seek justice, which in turn is the desire to have one's moral personhood recognized in a certain way. In the functioning state, we have delegated this function to the state; in its absence, we take it upon ourselves.

Here I would like to distinguish revenge from a closely related phenomenon, retaliation. I take retaliation to be similar to revenge insofar as it seeks to impose a harm on a particular person in response to a previous harm that that person imposed on the retaliation-seeker. But retaliation need not, like revenge, view the harm it seeks as *making up for* what was originally done. One can successfully retaliate without viewing the retaliation as transforming what happened before, or transforming one's relationship to what happened before or to the person who caused it. Retaliation can instead, for example, simply aim at deterring future harms (by e.g. inflicting a disproportionate harm in response). In this way, retaliation has a wider set of possible justifications than revenge does.

Furthermore, the notion of proportionality that is at home in reflections on retaliation does not fit neatly into revenge. To be sure, and as in retaliation, we have a conception of when revenge is clearly inappropriate (e.g. killing someone in response to

them cutting you in line at the post office). But that impropriety seems to be derived from the triviality of the original transgression rather than the relative severity of the vengeful response. Impropriety can also come in for both retaliation and revenge where there is a failure to track the guilty party – where, e.g., innocent bystanders are targeted rather than the original transgressor. But proportionality does not seem to be one of the metrics by which we judge revenge's virtue. Contrast this with retaliation, about which judgments of proportionality are both coherent and commonplace. As one example among many, the international norms governing proportional response to state action make sense to us precisely because evaluating a retaliatory response for its proportionality is not a category mistake – proportionality is the boundary within which retaliation's aims, be they deterrence, intimidation, or retribution, must remain. Revenge, on the other hand, need not be sensitive to the fine-grained determinations of proportionality that are at home in retaliation; beyond the fact of meeting a harm with another harm, revenge seems unmoored from concerns about the specific details matching the severity of one to the other.

One of revenge's distinguishing features, as I began to discuss above, is that it necessarily makes reference to *systems* of accountability in ways that retaliation need not. I can retaliate against my roommate not doing the dishes by using some of her laundry detergent. And this is perfectly understandable *as retaliation* even though we have no third party tasked with the adjudication of our household disputes. But revenge is different: the harms that we seek revenge on account of are ones that, properly speaking, *should have been* dealt with by some other institutional body. A reflection on familiar cases of revenge suggests as much: they usually involve egregious harms that cross over into the

realm of the political by virtue of their severity. There is a sense in which the failure of an institution to properly deal with a certain wrongdoing opens up a space for thoughts of revenge to at least have a certain coherence. It is precisely the failure of the institution that renders the thought "…and so *I* must make it right" legible.[4]

There are two different forms of institutional failure that might be at play in revenge: either a certain moral wrong was not properly dealt with by the state where it should have been, and therefore we have been failed; or, more interestingly, the possibility of a certain type of egregious harm by definition means that the state has failed, even if the state has mechanisms for their adjudication. What might these latter wrongs be? Chapter 1 is an extensive look at these types of wrongdoing. They are ones that impact moral personhood so severely that part of what it means for a state to be just is to prevent the conditions in which these are possible. It is not enough, from the point of view of justice, that their aftermath be dealt with "properly." Rather, they must be prevented from the start. Slavery, widespread patriarchal abuse, and genocide are examples of these sort of moral atrocities. Revenge's coherence as a concept is tied to the way that the failure of the state, i.e. the unjust state, can breed conditions that lead the revenge-seeker to seek justice on their own. Retaliation, though it does employ the related concept of fairness (insofar as it is concerned with proportionality), does not employ the concept of justice in this way.

[4] On this point I differ from most philosophers writing on revenge, who allow for the concept to include cases that involve wrongdoing that would not properly be addressed by the state to begin with (e.g. cutting someone in line), therefore making the idea of state failure irrelevant to such cases. Where state failure is discussed, as in Gerry Wallace's "Wild Justice," *Philosophy*, 70, No. 273 (1995), it is only to note that revenge under such conditions might be more sympathetic. In general, then, the literature on revenge takes it to be a more expansive concept than I do. But I believe what philosophers like Wallace refer to as revenge is more accurately captured by the idea of retaliation, and that reserving the concept of revenge for those acts that are related in some meaningful sense to state failure is a better articulation of the do-it-yourself spirit at the core of the concept.

This is not to say that we cannot retaliate in circumstances where revenge would also become legible. If my roommate were physically abusive towards me, I may decide to set their car on fire. That is a clear-cut instance of retaliation. But I may also seek revenge: not only do I set their car on fire, but I do so precisely as a way of communicating something about the impermissibility of physical violence towards me. Or perhaps I do it just to remind myself that I am capable of standing up for myself. Or any number of other motivations – the point is that revenge is motivated by a desire to transform oneself or one's standing in relation to the wrongdoing that retaliation's motivations lack. It is a self-regarding motivation as much as it is other-regarding.

And in a sense revenge contains this self-regarding component precisely because of the institutional failure. If we think back to the conceptions of the state discussed earlier, there is a strong philosophical current that views the state as both a reflection and a safeguard of moral personhood. Take Rawls' use of the Kantian conception of the person, as one example among many: he describes his principles of justice as ones that all persons situated as free and equal could adopt in a choice situation in which morally irrelevant facts about them are unknown. And this choice situation is modeled after the Kantian conception of the person insofar as its structure is based on our rationality and our sense of justice. The principles of justice that result from such a choice situation are ones that both reflect and safeguard its citizens' moral personhood. Where a system like that fails, then the citizens are on their own with respect to their moral personhood. This is the space in which revenge emerges as one strategy among others to, perhaps, repair the moral impact that a severe wrong has had. Retaliation involves no such institutional

correspondence because, I am holding, it involves no necessary moral self-regarding element.

So much for revenge versus retaliation. What about the related concept, retribution? Retribution comes up most often in discussions about punishment as one possible justification among a few – it is the idea that wrongdoers deserve to be harmed because of what they've done, and that punishment satisfies this aim and is therefore permissible in the right contexts. Insofar as revenge responds to wrongdoing, then, and is aimed at the wrongdoer, it contains retributive elements. But revenge has a necessary personal element that retribution need not have. Indeed retribution in the context of state punishment, where it is most often discussed, is meant to be free of the personal vendettas that can characterize private retribution. As Gerry Wallace writes in "Wild Justice," "[judicial] disinterestedness is meant to remove some of the moral defects of revenge and *personal* retribution."[5] This personal element is seen most clearly in the different foci of the two concepts: where retribution focuses on what the wrongdoer deserves (as I've said already), the revenge-seeker aims "to recover something owing *to her* or to which she is entitled."[6] The fact that revenge aims at another is an instrumental fact about how the revenge-seeker has chosen to restore something for *herself*; it is not the end in itself. For these reasons, retribution can be carried out by disinterested, third parties in ways that revenge, properly speaking, cannot.[7]

I would like to briefly say more, now, on the types of wrongs that impact moral personhood to such an extreme degree that revenge's motivational force becomes

[5] Wallace, "Wild Justice," 367. See 367-369 for further illuminating discussion of the differences between revenge and retribution.
[6] Wallace, 368 (emphasis added).
[7] For more on this, and vigilante justice, see Section 5 of this chapter.

substantial. My full discussion of this, in relation to oppression and the significance of a first-personal phenomenology for understanding its full scope, is in Chapter 1. However, a brief outline of those arguments is appropriate here. In my discussion of Fanon's work, he looks to the Black man's experience of colonization and finds a persistent experience of becoming a *thing* – this experience is described variously as becoming nothing, becoming an object, shattering, and having death suffuse your existence. While much of Fanon scholarship has focused on this work as giving us a groundbreaking insight into the *psychology* of colonization, I argue in Chapter 1 that we can also interpret this work as giving us a moral metaphysics of colonization, and by extension of oppression more generally. That is, by taking the first personal perspective of oppression seriously, we can see that the full scope of oppression's harm is not merely a psychologically distorted self-conception, a getting-it-wrong about oneself, but also an actual *transformation* into the kind of being that no longer has the status of an end in themselves. What this means is that colonial oppression casts people out of the moral community not just in practice but also in fact. Contrary to the liberal tradition, which views moral personhood as immune to the vagaries of interpersonal treatment no matter how bad it may get, I argue that one's moral personhood is contingent in a very important sense. Further, I argue that this contingency must be understood if we are to understand the full scope of the harms of oppression.

Now, what does this have to do with revenge? Revenge, properly speaking, emerges under two kinds of political failure, as we saw above: both the failure to properly respond to wrongdoing, and the deeper failure consisting in the very possibility of certain types of more egregious wrongdoing. Oppression is a characteristic example of

this latter type of failure. It is, for example, part of what it means to say colonialism constitutes a failure of justice that it consists in its subjects experiencing their lives as a sort of living death. So, too, for the harrowing accounts of Nazi concentration camp inmates, for first-personal accounts of American chattel slavery. Oppression, then, is a political circumstance ripe for revenge. Not only do people find themselves in a position to want to "make things right"; they are unable to look to the state to do so because it is the state's very failure that put them in that place to begin with. Revenge, insofar as it has a DIY spirit with respect to our moral personhood, enters the scene.

What kind of making-it-right might revenge provide? If one of the characteristic features of oppression is its erosion of moral personhood, then it would be important if it turned out that revenge could restore moral personhood. And it might thereby become rational to pursue it, insofar as the restoration of one's moral personhood is both in one's self-interest and, I will argue, a moral good in its own right.

3. Revenge is Not Necessarily Irrational

Let us now begin the formal argument for the conclusion of this chapter: that revenge is a rational pursuit insofar as it aims at the restoration of the revenge-seeker's moral personhood. Our first question is whether this is even a possible outcome of revenge – if revenge's successful pursuit has any relationship to the restoration of moral personhood. If it does, then we will have cleared away the first hurdle for the argument, which is that revenge is often regarded as *necessarily* irrational. This assumption is captured well in the following statement by Jon Elster:

> The very definition of revenge…shows that it involves only costs and risks, no benefits. Rational individuals follow the principles of letting bygones be bygones, cutting their losses and ignoring sunk costs, whereas the avenger typically refuses to forget an affront or harm to which he has been exposed.[8]

Elster's view is representative of the general philosophical spirit towards revenge: it involves only costs and no benefits, and it is irrational both because it fixates on the past *and* misunderstands the potential causal role the future might have in redressing that past.

In this section, I would like to resist this assumption by weaving together a certain conditional: if we can show that revenge, where successful, can actually yield the benefit of restoring moral personhood, then it will be a possible object of rational pursuit. This would mean that revenge is not *necessarily* irrational – the conditional opens up a space of possibility in which revenge becomes a candidate for rational pursuit under the right circumstances.

To explore this possibility, let's consider Martha Nussbaum's discussion of anger. To be sure, anger and revenge are not the same: one is an emotion; the other is a particular kind of action. But of course the concepts are related, and the one often leads to the other. This tight link is best expressed in Aristotle's famous formulation of anger as "an impulse, accompanied by pain, for conspicuous revenge."[9] But one need not agree with Aristotle's strong claim to see the link – anyone who has ruminated in their own anger knows that, at some point, its own tempting logic can present revenge as a possible source of pleasurable relief.

This tight conceptual link emerges in the familiar critiques of revenge insofar as they are simply extensions of a critique of anger as a possible ground of (certain) rational

[8] Jon Elster, "Norms of Revenge," *Ethics* 100, no. 4 (1990): 862.
[9] Aristotle, *Rhetoric* (Indianapolis, IN: Hackett Publishing Company, 2018), ii.2.1378a30-1.

action. Philosophers vary with respect to how exactly these arguments unfold: in one school of thought, revenge is seen as the culmination of practical irrationality carried out by those whose judgment is clouded by anger, itself always an irrational motivation.[10] In another school of thought, anger is conceived of as a rational and indeed morally valuable motivation under certain conditions, but revenge is always viewed as a pathological variety that exceeds the bounds of the morally permissible. In all cases, then, the interplay between anger and revenge must be navigated – either conceptual boundaries must be erected to stave off revenge, or their very kinship is used as evidence of their shared irrationality. Nevertheless, the philosophical focus tends to be on anger itself, with the objections to revenge coming as a sort of logical endpoint to the main arguments. The consequence of this is that there has simply been more philosophical attention paid to anger than to revenge.

Nussbaum's views on anger do not fit entirely into either school of thought. She clearly rejects anger as a rational moral motivation (except in certain time-limited, forward-looking forms), suggesting she is squarely within the first school of thought on anger. But she rejects one of this tradition's central claims that anger "conceptually involves a wish for violent revenge."[11] Rather, she takes herself to be making the weaker claim that anger merely involves the wish that the wrongdoer suffer in some way or another. Nevertheless, she holds elsewhere that dwelling "even briefly" on fantasies of retribution, of which revenge is one type in her account, is "pointless." This pointlessness is akin to the same irrationality that she claims is characteristic of anger

[10] I am grateful to Agnes Callard for discussion of this overarching distinction, found in ancient Stoicism and Buddhism on the one hand, and Aristotle, the British moral sentimentalists, and Peter Strawson on the other. For more, see Agnes Callard, "The Philosophy of Anger," *Boston Review*, April 22, 2020, http://bostonreview.net/forum/agnes-callard-philosophy-anger.
[11] Nussbaum, *Anger and Forgiveness*, 22.

itself. So there is an irrationality common to both revenge and anger for Nussbaum despite her rejection of the idea that this is due to the one being defined in relation to the other. In this respect, her account still does have the feature I mentioned above: that developing an objection to anger generates a sort of "trickle down" objection to revenge, too.

Though both schools of thought hold revenge to be irrational, Nussbaum's does so through a more straightforward analysis of the irrationality of anger (without thereby collapsing the distinction between the two). For this reason I focus on her account of anger, though my ultimate focus is on revenge proper. My hope is that this practical choice, i.e. drawing conclusions about revenge from an account of anger, will not overlook some important and hidden-to-me conceptual distinction.

Let's look at a fairly long excerpt from *Anger and Forgiveness* in order to get the main points of the view on the table:

> I argue that anger includes, conceptually, not only the idea of a serious wrong done to someone or something of significance, but also the idea that it would be a good thing if the wrongdoer suffered some bad consequences somehow… I then argue that anger, so understood, is always normatively problematic in one or the other of two possible ways. One way, which I call the *road of payback*, makes the mistake of thinking that the suffering of the wrongdoer somehow restores, or contributes to restoring, the important thing that was damaged. That road is normatively problematic because the beliefs involved are false and incoherent, ubiquitous though they are… There is [another] case, however, in which the beliefs involved in anger make a lot of sense, indeed all too much sense. That is the case that I shall call the *road of status*. If the victim sees the injury as about relative status and only about that – seeing it as a "down-ranking" of the victim's self, as Aristotle put it – then indeed it does turn out to be the case that payback of some sort can be really efficacious. Lowering the status of the wrongdoer by pain or humiliation does indeed put me relatively up. But…it is normatively problematic to focus exclusively on relative status, and that type of obsessive

> narrowness, though common enough, is something we ought to discourage in both self and others.[12]

So the phenomena she has in mind are ones that respond to a serious moral wrong by imposing negative consequences on the wrongdoer. This will always be normatively problematic, on her view, because it either relies on an incoherent set of beliefs about the restorative potential of those negative consequences, or because it conceptualizes the harm as only about relative status and so seeks to repair it through an inversion. I would like to show that Nussbaum is correct about the reparative potential of status-focused consequences but does not take this fact seriously enough, in part because she underdescribes the full scope of what angry people might be seeking to repair. In particular, if someone motivated by anger, or vengeance in the cases with which I am concerned, is responding to a moral wrong so severe that their moral *personhood* were impacted, then the "road of status" can be an effective way of regaining moral personhood through the relative reordering of interpersonal moral status. Where successful, this status inversion does exactly what Nussbaum claims cannot be done via the "road of payback": it restores the important thing that was damaged. In this case, then, the road of payback depends on neither false nor incoherent beliefs. And therefore it is not normatively problematic at least for that reason. This leaves us with only her second objection, which I will object to as well on separate grounds. Let's look at this argument more closely.

First, the *road of payback*. I have already argued elsewhere that certain severe moral wrongs can erode one's moral personhood. So in these cases, it is *moral personhood* that

[12] Nussbaum, *Anger and Forgiveness*, 5-6.

has been lost, and which the individual has an interest in restoring. How might they do so? Though Nussbaum discusses the *road of status* as a second normative objection to anger, unrelated to the *road of payback*, the discussion itself is an effective outline of how one might go about restoring their moral status, and therefore achieving what the *road of payback* seeks. Of anger, she writes:

> Anger often, though not always, focuses obsessively on status or down-ranking, and its wish is to reverse positions, pushing the offender down below the previously down-ranked self.[13]

What kind of status, or rank, does she have in mind? In general, Nussbaum refers to relative *social* status of the sort that we are all ordinarily aware of in our social lives: where we are sat at dinner parties, whether a door is held open for us, and the like. And she refers to this sort of status anxiety with disparagement – it is a prideful focus on relative social ranking that can approach obsession, incompatible with a healthy and secure ego. She refers to these sorts of injuries as having "a narcissistic flavor" insofar as they consist in a relative social reordering rather than actual harm to one's well-being determined apart from one's social standing (what she calls a "eudaimonistic concern"). In short, it is a picture of the relationship that a fragile ego might have to our admittedly fragile social rankings – indeed it is a picture of how the fragility of our social rankings themselves generates, in the narcissistic personality, a fragile ego.

Nonetheless, Nussbaum points out that these status injuries *can* be repaired by down-ranking the other – via shame, humiliation, or the like. Where social status is a

[13] Nussbaum, *Anger and Forgiveness*, 197.

"zero-sum game," we can imagine it on a seesaw: when one goes up, the other goes down. On the rationality of this repair strategy, Nussbaum writes:

> …[T]his is the one way in which anger makes good sense: humiliating the offender really does remove the wrong, if the wrong is seen as a pure status-injury.[14]

I want to flag for later the "good sense" that anger makes in this case – angry people are perceptive enough of the mechanisms of social status to recognize that divesting someone else of it will increase one's own in a crude sort of way. And they sensibly recognize that, all other considerations being equal, when one finds oneself in a lowered position, ideally one climbs back up on the back of one's offender.[15] Though Nussbaum refers to this as a "normatively substandard kind of sense," insofar as it involves practices that are normatively objectionable for other reasons (e.g. humiliation and shaming), there is no denying that it is at least not incoherent.

Importantly, however, Nussbaum carves out a special sort of harm that she refers to as a "dignitary injury" and distinguishes it from narcissistic status injuries. These dignitary injuries consist in acts that go beyond a first layer of self-interest and instead down towards a deeper moral core. These are the ones, unlike dinner party seating arrangements, that "involve the denial of people's equal dignity."[16] They therefore are of special, non-narcissistic concern to each of us and, eventually, to the law. These include systemic harassment and discrimination, terror, torture, and the like. Unlike status-

[14] Nussbaum, *Anger and Forgiveness*, 197.
[15] I'm not sure what to make of the fact that often, in the absence of higher-status offenders to humiliate, people often turn the opposite direction, choosing instead to humiliate those who are already "lower" on the social hierarchy than they are.
[16] Nussbaum, 199.

injuries, over which it is normatively suspect to "obsess" for very long, there is good reason to sustain special concern about dignitary injuries.

But this dignitary realm, unlike status-injuries, is presented in Nussbaum's work as nearly immune to the vagaries of social conduct. Throughout the book, she holds that equal human dignity "belongs to all, inherently and inalienably, and is not a relative or competitive matter…[I]n that way it is utterly different from relative status." [17] Elsewhere, she writes that relative status is "easily damaged," whereas "inner worth or value…is not."[18] As a result, though she holds it *is* rational and normatively permissible to sustain concern over these dignitary injuries, it is not rational (even in the "normatively substandard" kind of sense) to deal with them via status remedies. To do so would be to misunderstand both the nature of the injury and the nature of moral personhood. On Nussbaum's account, pushing the offender's dignity down "does not push the victim's up." [19] A different route is needed.

But this sets up a too-sharp distinction between status injuries and dignitary injuries, especially on a view (unlike Nussbaum's) in which one's dignity is shaped and maintained interpersonally. My Chapter 1 is all about the fragility of moral personhood – a fragility that is revealed under circumstances in which just the sorts of harms that Nussbaum discusses are present. This is a fragility that goes beyond one's moral character – the fragility of goodness as Nussbaum famously points out – and down to

[17] Nussbaum, *Anger and Forgiveness*, 27.
[18] Nussbaum, 26.
[19] Nussbaum, 27.

one's moral status altogether. Certain conduct can, under the right conditions, cast one out of the community of moral equals. It is not an inviolable status.[20]

Reflection on the fact that the law is meant to deal with these cases (and not e.g. mere status-injury cases in the sense that Nussbaum describes them) is further reason to recognize the fragility of moral personhood. Nussbaum herself writes, "It's not an accident that these serious well-being damages are the offenses that the law takes seriously."[21] And this is related to my observations earlier about the two ways an institution might fail: the more serious case, in which the possibility of certain types of serious moral harms means, by definition, that the state has failed, was meant to highlight the proper relationship a state has to the moral personhood of its citizens. Justice requires that the state, among other things, ensures that social and political conditions prevail which support full moral personhood. Indeed it is the particular nature of moral *person*hood that informs what a just state must look like. If moral personhood were inviolable, then it is unclear what work the concept of the just state could ever do for us.

Further, and as I've developed in Chapter 1, moral personhood is not only an individual matter. It contains an interpersonal element. This does not mean that it is a matter of mere social status, which can be offended by a misplaced curtsy. Nussbaum is right to point out that when mere status is truly the only focus of an individual (and not, say, the ways that slights to one's social status reflect deeper prejudice towards one's gender, or race, or ability), we are best off ignoring it and carrying on with our own projects. But where moral personhood is formed interpersonally, and *sustained*

[20] Though it is "inherent" in the sense that it is a status each individual has an innate right to attain. See Chapter 1.

[21] Nussbaum, *Anger and Forgiveness*, 164.

interpersonally, it is not a stretch to say that it is a matter of relative moral standing. This is a non-superficial status concern. Indeed part of what it means to strive for justice is to strive for moral and political equality (a relative status). And the relativity of this status does not take away from the fact that it is grounded in something deeply important about what it is to be a person: thought not inalienable, it is inherent, in the sense that each of us, in virtue of the kind of being we are, has an interest in the full development of our moral personhood. Whether that interest vests is a separate matter, contingent on the way things go interpersonally. If, for example, huge swaths of the population are treated as *mere means* in the Kantian sense, it is possible that they will suffer a dignitary injury so severe as to injure their moral personhood altogether. In this case, we can see the relationship between social status and moral status: contrary to what Nussbaum holds, the two are deeply intertwined.

This intertwining opens up the possibility that revenge may have at least the same sort of "normatively substandard" rationality that Nussbaum's *road of status* has. Just as retaliatory humiliation really does restore mere social status, it is possible that something (I am suggesting revenge) may really restore a deeper moral status that has been lowered or eradicated through interpersonal conduct. What this means is that the *road of payback* is not *necessarily* irrational: if the *road of status* really can restore what was lost, i.e. moral personhood, then it would be rational to pursue it. This also means that Nussbaum's objection to the *road of status* itself does not hold up in all cases. For when the road of status is pursued out of a concern for relative moral status, and not mere social status, then it has none of the interpersonal frivolity that Nussbaum is right to describe as undesirable in a well-functioning society.

The question, then, is whether revenge is capable of restoring moral personhood in a manner analogous to the way that humiliation restores social status. The analogy is made more difficult by the fact that, while I am holding that moral personhood is an interpersonal matter (like social status), it is not *only* that. We must also look to what is inherent about certain types of beings such that when things do not go *interpersonally* how they should, then something has gone wrong. What does it mean to say something has gone wrong, after all, if the being in whom normativity is grounded has not fully developed? It seems that the stifling or atrophy of moral personhood is itself an impediment to a moral critique of that phenomenon. Nonetheless, I will argue that the interpersonal element of moral personhood has to do with its development, while the normative force of the status has to do with kind-membership, and that combining these two elements of moral personhood over a lifespan is a coherent and normatively advantageous way to understand its full scope.

Also, a word about a second question: namely, whether the above question can be asked without also answering the question of whether it would be morally permissible to do so, all things considered. In the easier case of social status, we can observe that humiliation restores social status, even though humiliation is (let us assume) morally problematic for independent reasons. Separating the question of the moral permissibility of the means from the question of whether the means are able to achieve the desired end is easy to do. But in the case of moral personhood, the separation is not as easy to maintain. For if what we are restoring is moral personhood, then it makes sense to worry about whether *any* route can attain such a morally loaded endpoint, or if only a morally permissible subset of these routes has such potential. Put differently, how can we

consider revenge's emancipatory potential without first deciding whether this end-state, moral emancipation, can be undermined by avenues that are in conflict with it? Someone may object, for example, that moral personhood cannot be achieved by any means necessary – where the "cannot" is meant conceptually and not normatively.

Without fully answering this question, I will point out that we are concerned here with moral status and not moral character. If we were concerned with only the latter, then it would be a prima facie plausible view that one cannot become e.g. a virtuous person by non-virtuous means. But our question is different: it is not whether one can become virtuous, but rather whether one can become the sort of thing about which it makes sense to make these moral judgments. Being a candidate for moral evaluation is distinct from the particular ways that evaluation might end up. And it is at least not self-evidently impossible that one might become a candidate for moral evaluation via means that ensure the content of that evaluation will be negative.

The more difficult version of this critique is that becoming a candidate for moral evaluation has a deep relationship to the concept of freedom, and freedom is the sort of thing that is undermined by non-virtuous conduct of a certain extreme sort. Hegel and Kant hold some version of this view. I don't yet have an answer to this difficult question. My hope is that what follows will not be doomed by this sort of Hegelian worry.

4. The Rationality of Revenge and its Relation to Moral Personhood

4.1 Beauvoir on Revenge

In order to explore the question of whether revenge has the potential to restore moral personhood, I will now turn to Simone de Beauvoir's writings on the subject. I will

begin by exploring what I think she gets right about revenge (and which many other views lack) – namely, that revenge provides a special case of reversing the moral degradation that certain harms engender. And I will then turn to amending what I believe she gets wrong. That is, while Beauvoir holds that in the end revenge cannot achieve its aims of liberation because of an unavoidable internal contradiction in its motivations, I will argue that this contradiction only exists where revenge is contingent on "reaching the freedom of the evildoer" in a certain way – i.e., via the evildoer's own consciousness. If, on the other hand, revenge can reach the freedom of the evildoer without requiring his cognizance of it, then we can avoid this contradiction that Beauvoir is right to point out and maintain that it is possible to achieve a successful act of revenge. With this contradiction cleared away, we can then make headway on exploring how a successful act of revenge restores moral personhood.

Beauvoir describes revenge in these words:

> Hatred grasps at another's freedom insofar as it is used to realize the absolute evil that is the degradation of a man into a thing. And it calls immediately for a revenge that strives to destroy that evil at its source by reaching the freedom of the evildoer.[22]

In "An Eye for an Eye," Beauvoir weaves theoretical reflections on revenge with historical narratives of retributive violence witnessed in the aftermath of World War II – targeted killings of collaborators, "lynchings of snipers," and massacres of S.S. prison guards by those who were formerly held captive in the concentration camps.[23] She describes the hate that emerged in all decent people, the deep desire to see that those

[22] Simone de Beauvoir, "An Eye for an Eye," in *Philosophical Writings*, ed. Margaret A. Simons (Urbana, IL: University of Illinois Press, 2004), 248.
[23] Beauvoir, "An Eye for an Eye," 247.

who committed unspeakable acts of violence did not go unpunished. For what other stance towards them was possible at the time? "Vengeance," she writes, "answers to one of the metaphysical requirements of man."[24]

And Beauvoir defends this attitude, to a degree. It did not matter to her that retributive violence seeks neither rehabilitation nor repair – its punitive aim is precisely its point. On its justification in hate, she writes:

> Indeed hate is not a capricious passion. It denounces an abominable reality and imperiously demands that this reality be eradicated from the world. One does not hate a hailstorm or a plague, one hates only men, not because they are material causes of material damage, but because they are conscious authors of genuine evil.[25]

As with many other moral philosophers, Beauvoir notes that hate is sensitive to moral responsibility – it is a response only to those harms that have their origins in a freedom. And so the desire for revenge takes on a specific dimension: the act must reach not only the evildoer but the evildoer *qua* moral agent. Revenge must reach a freedom. In so doing, the act is able to achieve a concrete realization of a double reversal: where the original act of evil negated the freedom of the victim and therefore affirmed the evildoer as a sovereign consciousness, the act of revenge reverses this in order to reestablish the freedom of the victim. She describes it as follows:

> The affirmation of the reciprocity of interhuman relations is the metaphysical basis of the idea of justice. It is what vengeance strives to reestablish in the face of the tyranny of a freedom that wants to be sovereign.[26]

---

[24] Beauvoir, 247.
[25] Beauvoir, "An Eye for an Eye," 247-48.
[26] Beauvoir, 248-49.

Revenge, then, aims to restore a certain balance to the moral order – it aims to move a victim from a place of thinghood to a place of moral import. And it does so, according to Beauvoir, by reaching the freedom of the evildoer in a particular way. That is, the evildoer must, in his freedom, come to recognize the atrocity of his acts. He must come to *consciously understand* the reversal as justified, and so come to understand, from the position of his own freedom, the moral urgency of its violent constraint. Beauvoir writes:

> [T]he foreign consciousness must remain free with regard to the content of its acts, it must freely recognize its past faults, repent, and despair, but it must be an exterior necessity that compels this spontaneous movement.[27]

It is for this reason that the urge for retribution was stymied by, for example, Hitler's suicide, by his taking his death into his own hands – his freedom was preserved until the very end, and we have no guarantee that it was used to reflect on his evil, to repent, or to despair. Those among us who sympathize with the victims of his regime of torture feel that they were cheated in some sense of an opportunity to restore the moral order on their own terms – cheated of the very act which could accomplish that restoration most efficiently insofar as it would have constituted an assertion of their own freedom. The account of revenge presented by Beauvoir is one whose aim cannot be achieved unless its target is compelled *by his victims* to bring both his actions and himself under the description of evil. Anything else is a failure.

Beauvoir further holds that this necessary condition of revenge is precisely why it cannot ever be achieved. She describes a paradox at its heart: the contradiction between the free recognition of one's own evil and the "exterior necessity" that revenge wants as

[27] Beauvoir, "An Eye for an Eye," 249.

this recognition's cause. "It is because of this contradictory character that the intent of vengeance can never be satisfied."[28] The core of revenge's desire, for Beauvoir, is to impose one's own will on that of the evildoer's, thereby creating a painful space within which the evildoer freely comes to see their own wrongdoing. That this conscious recognition spans over time is all the better: it is the elongation of a free will coming to see the justified grounds on which its constraint is based. It is to understand, from the point of view of freedom, that one has simultaneously lost the moral basis on which to object to its erasure – and, thereby, to freely come to see that one deserves unfreedom. This is the moment, stretched out forever, that revenge takes as affirming "the reciprocity of interhuman relations."[29] Beauvoir ultimately points to the impossibility of this stretching out forever – revenge's logical endpoint – as evidence of its contradictory character. And for this reason she holds that revenge can never accomplish its own aims. "[W]e must stop seeing vengeance as the serene recovery of a reasonable and just order."[30]

But what if revenge harbored no contradiction? Beauvoir's account locates a contradiction in the content of revenge's aim – it exists in desiring that the evildoer become freely and consciously aware of his own evil. More succinctly, the contradiction arises at that point where revenge aims at the imposition of a particular free, self-conscious attitude. If we can discover an account of revenge that does not contain this as one of its aims, then we can avoid the contradiction and open up the philosophical space for considering whether revenge, when successful, restores the freedom of the revenge-

[28] Beauvoir, "An Eye for an Eye," 249.
[29] Beauvoir, 248-49.
[30] Beauvoir, 259.

seeker in the way that Beauvoir denies. This is the task I would like to embark on now, and it relates to Beauvoir in the following way: to recall, Beauvoir's account of revenge has two steps. First, revenge aims to (re)establish a sovereignty that was undermined by some morally reprehensible act or set of acts. Second, revenge seeks this sovereignty by imposing a will on another such that the other freely comes to see the necessity of their restraint, freely comes to repent and despair at understanding oneself as brought under the concept of evil. I think there is good reason to reject this second moment of revenge on Beauvoir's account, from a purely descriptive standpoint, and preserve only the first. That is, I think an accurate account of what is going on in paradigmatic cases of revenge reveals that it is not the free recognition that satisfies its aims but rather the imposition of the exterior will itself. A closer look at first-personal accounts of revenge will help us bring out this point.

4.2 Jewish Revenge

> *"We must fight as Jews," said Abba.*
> *"Do Jews have a different goal in their war against the Nazis?" asked Yurgis.*
> *"We have a special desire to avenge the things done to our families," said Abba. "It was not the citizens of Lithuania who were annihilated. It was the Jews, and the desire for revenge is so powerful that it must be held in high regard."*[31]

Abba Kovner was an early twentieth century poet and Polish Zionist. He was also an esteemed young leader of the Jewish underground resistance during the Holocaust, leading an armed uprising in the Vilna Ghetto in 1943 before joining Soviet-led partisan fighters in the nearby forests. There in the forests, Kovner led the all-Jewish partisan

[31] Rich Cohen, *The Avengers: A Jewish War Story* (New York: A.A. Knopf), 114.

group *Nokmim* ("The Avengers"), sabotaging German supply trains, killing Nazis, and killing any collaborator they encountered – Kovner himself was known to leave the note "Killed for betraying a Jew" on the body of any peasant he encountered, and killed, whom he knew to have turned in or killed one of his own soldiers.[32] The fact that those in his resistance unit were fighting *as Jews* was of great consequence to him – the above conversation was had with a fellow partisan leader who wanted to dissolve Kovner's Jewish unit and reabsorb it into one of the other nationalist units fighting alongside them in the Vilna forests. Kovner, of course, refused. By that time, he and his fellow resistance fighters had endured unimaginable trauma *precisely* because they were Jews – their families were tortured and killed as Jews, their friends, their lovers, their entire communities, tortured and killed as Jews. The causal connection that this status had to their collective trauma was something he was unwilling to cover up as they fought, in part because the fighting itself was not fully legible without it. Their fighting was not mere self-defense as threatened Lithuanians; it was the beginning of a long program of revenge *as Jews*.

Alongside him from the early days of the Vilna ghetto were Vitka Kempner and Rozka Korczak, two young women who eventually became esteemed lieutenants in their own right in *Nokmim*. During their time in the ghetto, they were often tasked with some of the most dangerous resistance missions, as they were able to navigate the world outside the ghetto walls without arousing as much suspicion as the fighting-age men in their ranks. Kempner and Korczak smuggled weapons, food, and Jews; they orchestrated armed resistance within the ghetto; blew up Nazi railyards; and served as a vital link in the underground communication network that spanned Europe at the time. Once in the

[32] Cohen, *The Avengers*, 86-87.

forests, they continued their fighting, sabotaging German troop movements and destroying the infrastructure of the nearby town in which the Germans were based.

They, too, understood what they were doing as revenge, especially as the resistance fighting entered a new stage after the war was over. Speaking about this transformation Korczak remarked, "We can now think of revenge with a plow as we once carried automatic weapons and grenades."[33] Revenge, for these resistance fighters, became a slow and steady grind. After all, millions of Nazis went unpunished after the war's end; Jewish refugees were being murdered upon their return to their hometowns, barely escaping death in the concentration camps only to be met with persecution by their collaborating neighbors; and German propaganda was advocating for the return of another Hitler. For these resistance fighters, there had been no justice in the war's end; only the beginning of a new chapter.

This beginning was to take the form of *Nakam*, the revenge plot dreamed up by Abba Kovner and advocated for across the surviving Jewish diaspora. In an early underground meeting with other survivors, Kovner decried the post-war conditions for Jews, the impunity of the vast majority of Nazis who occupied middle and lower positions in the genocide, and exclaimed: "We will do it for ourselves."[34] This enterprising spirit in the shadow of injustice is the hallmark of revenge. What, exactly, would they do for themselves? The *Nakam* program was a plot to kill former Nazis at-scale as retribution for the atrocities they committed during the war. Soon this plot crystallized into a plot of mass poisoning, in which either Nuremburg's entire water supply would be poisoned (though not the supply of those neighborhoods known to be

[33] Cohen, *The Avengers*, 189.
[34] Cohen, 189.

friendly to Jewish refugees), or, if that failed, a prisoner of war camp housing former Nazi officials would have their rations poisoned, instead.

Kovner's views on the mass nature of any acceptable revenge plot were summarized in an account given to a surviving family member of his comrade, Korczak:

> Abba believed that revenge should be anonymous. He did not approve of those Jews, mostly from Palestine, who traveled in Europe hunting Nazis. These agents had acquired a list of SS officers, names and addresses, and would, dressed as British soldiers, drive to some house, ask the man inside to ride with them to headquarters, then, beyond the city lights, say, "We are Jews," and shoot him. Abba thought such attacks too personal: they acknowledged the Nazis as individuals, allowing them to face their accusers, a luxury the Germans had never offered the Jews…The Germans, he said, must be killed in the same inhumane, factory-like manner in which they had killed the Jews.[35]

For Kovner and those who enlisted in his revenge plot, it was not important to the act of revenge itself that its targets recognize what was happening. This was true when he killed collaborationist peasants during the war (think of the note left on the corpse only after the killing) and it was true after the war: the cognitive and emotional life of the Nazis was not his main concern. Indeed, it was important to Kovner that their ignorance over their fate was preserved – announcing "We are Jews" before retribution took place was to grant the Nazis a humanity that he was intent on undermining. Contrast this with Beauvoir's claims about revenge's need for the target to bring himself under the description of wrongdoer, to finally *understand* what he had done. Kovner's revenge plot not only lacked this aspect, but he viewed such a forced understanding as antithetical to the project he was undertaking. This is not to say that the more dramatic revenge plots he looked down upon were not revenge; it is only that it shows that revenge does not

[35] Cohen, *The Avengers*, 192.

need such a dramatic recognition from its target. Kovner's plot was one that had the potential to succeed *as* revenge; this shows us that the phenomenon itself does not contain Beauvoir's second element as a necessary condition.

What *was* necessary, however, was that Jews were the ones carrying out this act of revenge. This direct causal link is exemplified in an anecdote from one of the young teenagers, Zalman Tiktin, sent by Kovner to procure weapons during the time of the Vilna ghetto:

> Each night, Abba sent Jewish partisans into the city to cut deals, bribe locals and buy equipment. That winter, the Germans caught a teenager in the rail yards breaking into a freight car loaded with machine guns. He was sixteen. In Gestapo headquarters, Zalman Tiktin was questioned by Murer and a sadistic Gestapo officer named Bruno Kittel. When asked whom he stole the weapons for, Tiktin said, "I stole them for you, because you killed my parents."[36]

To be sure, these weapons were also used for the self-defense of the Jews living in the Vilna ghetto. In that sense, Tiktin stole the weapons *for* the resistance. But that is not the reason cited for the theft – his response is an unequivocal retributive statement. And not mere retribution, which can be successfully carried out by any number of third parties, but personal retribution. Tiktin stole the weapons because *Tiktin's* parents were killed. Ordinary accounts of retribution lack this personal element (eg. retributive accounts of state punishment). The spirit of the later revenge plots was the same as Tiktin's: not only did Nazis have to be killed, but because Jews were the victims of Nazi atrocities, Jews had to be the ones doing the killing.

I mentioned earlier that there is a sense in which we feel the Jews were cheated by Hitler's suicide. Beauvoir's explanation of this sense of stolen retribution would be that

[36] Cohen, *The Avengers*, 56.

there is no reason to believe Hitler ever really *understood* what he had done – he was able to escape being compelled to see himself in the way that, all things considered, he ought to be seen. I would like to offer an alternate explanation of this sense of being cheated, one that focuses not on the failure to transform Hitler's self-understanding but instead on the fact that, in committing suicide, *he* was the cause of his own death (and not, say, a Jewish sniper). *This* is at the heart of that sense of stolen retribution. I believe the revenge stories from the Jewish resistance support this view. It was not important to them that their targets come to understand themselves differently. Rather, what was deeply important was that their targets not only "get" what they deserve, in some sense, but that the ones making sure this happened were the Jews themselves. In Beauvoir's terms, the resistance fighters sought to undermine the Nazis' freedom, but not necessarily via the Nazis' subjective experience. This causal element in particular, that Jews be the ones undermining the Nazis' freedom, was the most important for Kovner's resistance fighters. They believed that to overcome what they had endured *as* Jews required that they triumph in the same way. It is the continuity of the once forsaken identity that, for them, had the only chance of restoring a certain moral order to the world.

In the end, Kovner and his comrades across Europe had to abandon the original plan to poison entire waterways. They instead opted to poison the bread that was being sent to a prison camp housing former Nazi officers. Reports vary, but an estimated 2,000 Nazi officers became sick and hospitalized. No deaths were reported, but this was thought by the resistance at the time to be a cover-up intended to shield the British

guards from liability for their oversight. Writing about the revenge plots, Ruzka Korczak had this to say:

> [Revenge] is a cold decision not influenced by sentiments or emotion. This is fate. And since it is so, there is not much to ponder. Had Abba not come up with the idea, some Mr. Everyman would have done so.[37]

This is a far cry from the accounts of revenge that we are used to. Philosophical accounts almost invariably hold that revenge itself is the unbridled extreme of anger, that it is at the height of irrationality in part because of the way that this anger clouds one's judgment about the proper course of action in the aftermath of wrongdoing. Think of the bloodthirsty Furies, those dog-like creatures vomiting blood, moaning, and seeking vengeance. The moral chaos they bring is what must be driven underground and transformed in the *Oresteia*. But Korczak does not present revenge as moral chaos – instead, she presents it as having the orderliness of fate, the chill of calculated rationality. In fact, she sees it as so distanced from the subjective will that *anyone* could have come up with the plot – it did not hinge on the idiosyncracies of, e.g., Abba Kovner's moral psychology. Rather, it was a plot whose universality was rooted in the fact that it, as an act of vengeance, responded to what Beauvoir refers to as "one of the metaphysical requirements of man."[38]

However distant this portrait of Jewish revenge is from the accounts with which we are more familiar (ones that often come up in fictional depictions of revenge, where the moment of recognition is often the climax of the film), it nonetheless *is* an account of revenge. I think it is a shared, strong intuition that if anything counts as revenge, their

[37] Cohen, *The Avengers*, 199.
[38] Beauvoir, "An Eye for an Eye," 247.

story does. What I take this to show is that Beauvoir is incorrect about the necessary elements of revenge (and therefore about its necessary contradictions). The revenge plots taken by the Vilna resistance fighters were ones that contained only the first element as far as Beauvoir is concerned: they were plans to reestablish Jewish personal sovereignty, decimated during the Holocaust and its aftermath. But the plans did not aim at doing so by forcing a recognition in the subjective experience of its targets. What this shows is that revenge's aim does not have the subjective experience of its target as one of its necessary elements. Of course, that may very well be present (and perhaps subject to the self-defeating internal contradiction Beauvoir points out, though I do think that the Palestinian revenge plots upon which Kovner looked down are counterexamples to even this latter claim). But the point is that it is not *necessary*.

With this necessity cleared away, I take this to also show that there is at least the philosophical space for claiming that a successful act of revenge can, in fact, restore one's sovereignty and a certain balance to the moral order. If revenge is not doomed to failure due to an internal contradiction, then we can start to ask questions about what it might achieve through its success. I will argue in the next section that this success achieves precisely what Beauvoir thinks revenge (misguidedly) aims to do.

4.3 Revenge and Moral Emancipation

I will argue in this section that the general mechanics of revenge's success, defined as the moral becoming of the revenge-seeker, follows this trajectory: The acts of revenge that we are considering are ones that involve an imposition of one's will on a target for a very particular reason: as a way to make up for some prior wrong. To act on

this rationale involves a certain cognitive precondition: believing that those actions were indeed wrong, that the revenge-seeker deserved better, etc. And one cannot hold *these* thoughts without viewing oneself as, in at least some provisional sense, a moral being. I view the revenge-taking, i.e. the imposition of one's will on another for specific retributive reasons, to be the actualization of this provisional self-regarding position. To e.g. kill one's target *for those reasons* is to actualize oneself as a moral being, and not merely to impact the life of another. It does so by entering back into an interpersonal moral connection that was severed by the prior revenge-worthy conduct. Where one's moral status stands or falls based on interpersonal relations, it can be rebuilt in the same manner, even where the act of reconstruction consists in what would normally be considered morally impermissible conduct. Let's examine each moment in this moral becoming in more detail.

As I said above, acts of revenge involve an imposition of one's will on a target for specific retributive reasons – it is *because*, for example, the Nazi concentration camp guards tortured Jewish inmates that one of the Jews might dedicate their post-war efforts to hunting these former guards down. But endorsing this causal story requires another prior commitment: believing that the torture itself was wrong. This moral concept, in turn, requires that one view the behavior in question as occurring between two moral subjects. The conduct of the Nazi guards becomes morally relevant, indeed becomes *torture* and not some more neutral activity, only when we understand their victims as moral beings who deserve otherwise and not, say, inanimate objects. So revenge's retributive element involves a chain of cognitive preconditions.

To recall, my focus so far has been on revenge for conduct that is so egregious that it impacts one's moral personhood altogether. It is conduct that, from the first personal perspective, transforms a victim from a moral being into a mere thing. Entertaining the beliefs that are the first seeds of revenge, then, is already a substantial moral accomplishment: it is entertaining the idea that one is a moral being from the perspective of having that status compromised by others. This is part of what I mean when I say that revenge contains a self-regarding element: it is to view the other through a lens that is only possible when one views oneself as a moral being.

Unfortunately this cognitive self-regard is not a sufficient condition to (re)becoming a moral person, because as I've discussed elsewhere moral personhood is an interpersonal phenomenon. It matters how I exist in the wider social-moral landscape. Nonetheless, this cognitive self-regard does function as an intermediary point in the emancipatory process. It involves, from the perspective of being an object, a splitting from oneself and considering oneself as a subject, too. This subject position is the position of moral personhood. From this position, the evaluation of others' conduct can be undertaken in a different light.

This provisional self-regarding position becomes actualized, then, when one in fact enters into an amended relation with one's wrongdoer. To entertain the possibility of asserting your will on a wrongdoer is one thing; to actually do it is another. Under more just conditions, i.e. political conditions that fall outside the scope of "oppression," there are a number of avenues to accomplish this. One plausible reading of a functioning justice system is that it allows a victim to sublimate their will into the will of the state and, in virtue of the state's punishment of a wrongdoer, thereby affirm their own

standing in moral relation with others. Or perhaps the wrong in question does not require state interference; still, there is a sense in which certain informal, reparative actions are means to reinstate a moral relation that was called into question by some wrongful conduct.

But at the core of what I have been holding re. oppression is that it is not only constituted by wrongful acts that get at the moral foundation of persons, but that its conditions are ones that lack the resources for the sort of moral adjudication I am discussing. That is, the conditions that make the concept of revenge relevant (indeed, coherent) are ones that, by definition, narrow one's options for morally reparative action. This is the moment at which violence enters the scene – not because revenge-seekers are particularly violent, but because the conditions in which violence is one of the few remaining option to assert your will over a wrongdoer are the conditions that give coherence to the concept of revenge. Therefore violence is a necessary feature of revenge *only because* it is a contingent fact that violence is one of the only options left under conditions of injustice for amending one's interpersonal relation with one's wrongdoer.

A crucial element of revenge is that it is done for the retributive reasons I have been discussing. Why is this crucial? Because it is what distinguishes the revenge-taker from the lightning strike: only where one is acting *for reasons, towards a moral agent* does the action forge a moral relation. Lightning strikes do not. And so the moment this moral relation is forged is the moment that the provisional self-regarding moral status becomes actualized. It amounts to an act of moral emancipation.

Becoming the type of moral being one has the capacity to become is a moral good; insofar as revenge is one avenue towards this moral becoming, it, too, is a moral good. Like courage or meaningful friendship, moral actualization is one of the goods at which we might aim in our practical lives. Of course this does not mean that revenge will be *the* good thing to do, all things considered, just as certain courageous acts might need to be foregone in the name of a more-pressing consideration. Like other moral goods, moral actualization may very well be outweighed by countervailing moral interests. But notice that *this* moral good, which revenge can achieve under special circumstances, is of a different status than courage or friendship: courage and friendship presuppose an already-actualized moral agent who weighs them against other moral goods; moral actualization, though, is prior. It determines whether someone is a candidate for the meaningful pursuit of the others. And there is a sense in which fully-actualized moral persons are blameworthy for failing to pursue courage or friendship in ways that non-actualized people are *not* blameworthy for failing to pursue moral actualization. There is a meaningful difference between the two types of moral goods, and it has to do with their respective relationships to the question of moral actualization.

Revenge's status as a moral good, then, is derivative of the more general moral good of moral actualization. It is good to become the types of beings we are; on my interpersonal account of moral personhood, our rationality grounds this potential that, under just conditions, joins with the interpersonal in such a way as to actualize one's full moral personhood. But as I've argued, some state failures (i.e. some instances of injustice) can rob one of this interpersonal moral becoming. Revenge, then, enters the scene as an alternate avenue towards this moral good. Because the most important aspect

of revenge is the causal element, and not the psychology of its target, it contains none of the self-undermining contradictions that Beauvoir points out. Rather, revenge is one way we can re-forge an interpersonal *moral* connection, and thereby become and simultaneously enter into a community of full moral persons. Of course it's possible that this act might also render one as having committed a moral wrong; nevertheless, it is a moral accomplishment to become the sort of thing for which this type of moral evaluation is newly apt.

5. Revenge: Constraints and Objections

With the main view in place, I would like to now briefly consider a number of its consequences and one especially compelling objection. One consequence of this view is that because of the self-regarding element at the core of revenge's rationality, the concept of revenge is limited to the *personal*. By this I mean that avenging someone *else's* harm will not, strictly speaking, count as revenge. This goes against the familiar usage of the term: we often speak of people avenging the death of their child, spouse, or other close relation in order to "get justice" for someone who can no longer do it on their own. In ordinary usage of the term, these scenarios are mixed in together with the personal ones that have been described in this chapter. But I believe there is value in separating them out and confining the concept of revenge proper to those instances where there is a direct (rather than an indirect) historical relationship between the revenge-seeker and their target. If revenge really is the moral emancipation of the revenge-seeker, and this emancipation requires that they be the ones targeting specifically their wrongdoer, then it

matters for the concept that the one inflicting revenge's harm is in need of moral emancipation precisely because of what their target did *to them*.

This does not mean that we have nothing to say about the other, more mediated cases. Vigilante justice, of which I take those to be examples, is a cousin of revenge insofar as it's a "taking matters into one's own hands." This is an important shared feature – it distinguishes both activities from the ones that are endemic to the successful state (e.g. filing a police report, turning the matter over to the judicial system, etc.). But the question of its rationality will have to depend on something other than the moral consequences of reenacting a different relationship with one's wrongdoer. Perhaps one can offer such justifications for the rationality of vigilante justice in the shadow of a failed state: deterrence is one plausible answer that would hinge on empirical questions about its efficacy; maintaining one's integrity (in the sense described by Bernard Williams) would be another.[39] I bring these possibilities up not to explore them, though, but rather to highlight the differences in rational justification for revenge versus vigilante justice and the utility in separating out the concepts along these same lines. While it departs from ordinary usage, I believe it benefits from an increase in conceptual precision.

Another consequence of this view is that many genuinely *personal* cases of taking matters into one's own hands will nevertheless not count as revenge. In this respect, too, it departs from ordinary usage. It is a crucial feature of my view that the rationality of revenge is grounded in its capacity to generate what the unjust state failed to protect.

[39] For more on this, see Brian Rosebury's interesting discussion of non-moral respect for revenge (what I would classify as vigilante justice) in Brian Rosebury, "Respect for Just Revenge," *Philosophy and Phenomenological Research* 77, no. 2 (2008): 451; and Bernard Williams, *Ethics and the Limits of Philosophy* (Cambridge, MA: Harvard University Press, 1985), 174-96.

Think back to the *Oresteia* – the establishment of the state was meant to do away with the need for revenge. In my terms, I think it's more precise to say that the establishment of the state rendered the possibility of revenge empty, as long as that state maintained conditions of justice. There was nothing that the state failed to address for which revenge could have filled in the gaps.

What are we to do, then, with all of the ordinary uses of the term 'revenge' that don't fit into this picture? They tend to break down into two types: either someone seeks "revenge" for a petty slight; or they do so for a genuine interpersonal harm when the state already has the means and procedures available to address it properly. Let's consider two examples:

> **Example 1:** Ted is in line at the apothecary when Paolo accidentally cuts in front of him. Because the sales clerk did not notice, they ring up Paolo next. Ted does not have any pressing appointments to get to, but he is mad at the slight and vows to take "revenge." He sneaks out while Paolo is paying, places a bomb under his car, and waits for it to explode when Paolo puts the keys in the ignition.

> **Example 2:** Rudy buys celebratory shots for everyone at the bar except his neighbor Céline, towards whom he has an inexplicable grudge. Céline becomes angry, orders and takes her own shot, then throws the empty shot glass at Rudy's head. Rudy goes to the hospital to get stitches and Céline is arrested. While Céline works her way through the justice system, which Rudy has every confidence will

respond appropriately, he develops and carries out a "revenge" plot to set her house on fire.

Example 1 is a case of a petty slight (by which I mean: no genuine harm); Example 2 is a case of a genuine harm for which the state has appropriate adjudicatory procedures. It is natural to use the term "revenge" to capture what Ted and Rudy do. But on my view, neither is a genuine case of revenge because in neither case is moral personhood at stake. A one-off, unintentional line-cutting is not the sort of thing that amounts to a harm, much less the type of harm that can erode moral personhood. And so Ted is not a candidate for the self-regarding element of revenge – he cannot participate in the activity because he is already a fully-actualized moral person. Just like a virtuosic piano player cannot participate in the activity of *becoming a piano player*, Ted cannot participate in the activity of becoming a full moral person. Ted mistakenly uses the term "revenge" to describe an act of wildly disproportionate retaliation.

Rudy does suffer a genuine harm; his bodily autonomy and right to safety was violated by Céline's throwing of the shot glass. But he, too, misuses the term "revenge," because there is no need for him to seek justice on his own. As I stipulated, Rudy himself has every confidence that the justice system will respond to Céline's crime appropriately. And as I've discussed earlier in this chapter (and at length in Chapter 2), the state (in the social contract tradition) is the third-party safeguard of our moral personhood. Recall that there are two types of state failures that result in a failure to safeguard moral personhood: failing to respond adequately to wrongdoing; and allowing for the possibility of certain more egregious wrongs, even if the state has an adjudicatory

mechanism for it. Rudy is not in either of these situations; the state is dealing with Céline's crime well, and luckily it did not amount to the type of harm that can erode moral personhood. So Rudy has no need to seek it on his own. Rudy, like Ted, is engaged in an act of retaliation.

The upshot of limiting the conceptual bounds of "revenge" in this way is we are able to preserve the self-regarding element at its core *and* preserve the conceptual link with failures of justice. Though all of the concepts I have discussed in this chapter – revenge, retaliation, retribution – have something to do with justice, only revenge involves a robust notion of it. Retaliation and retribution are concerned with justice in the minimal sense that Aristotle describes – a sense of fairness, i.e. treating similar cases similarly. But revenge involves justice in the fuller sense that it is about securing the conditions for human flourishing. This requires us to have a conception of the person (as in, for example, the Kantian core of the social contract tradition), and it is this conception that the concept of revenge responds to. Ordinary usage of the term does not reflect this, but we can.

Let us now consider a pressing objection against this view. It should go without saying by now that the preceding account of the rationality of revenge is entirely rejected by views like Nussbaum's, and that this rejection is accompanied by a clear stance on the question of whether revenge could ever constitute a moral good. In a sense, revenge itself is not even taken seriously by these views in its own right. Rather, it is rejected in virtue of its location at the far side of anger, and the possibility that the phenomenology of revenge may differ in important instrumental respects from anger is not entertained. Their conclusions are straightforward: revenge is irrational and morally impermissible,

and revenge-seekers are blameworthy insofar as they go down that path, regardless of how extreme the conditions were that generated that desire.

A more complicated objection, however, comes from another angle: one that takes revenge seriously, but denies that it can succeed in the sort of moral emancipation I have described. George Orwell captures this sentiment best in "Revenge is Sour," where he sidesteps the question of revenge's moral permissibility and instead focuses on its phenomenology. There he concludes that the type of moral emancipation I am discussing is actually an artifact of fantasy, and that "when the thing becomes possible, it is merely pathetic and disgusting"[40] – far from the picture of empowerment and moral becoming that I have painted. Let us take a look at this claim, and ask whether it undermines the view I have put forward.

Orwell, like all of the others discussed in this chapter, considers revenge in the context of the aftermath of the Holocaust. He toured one of the prisoners-of-war camps holding Nazis in South Germany (much like the one poisoned by Abba Kovner's Avengers) and saw firsthand how the Jewish guards kicked and berated the imprisoned Nazi officers, not out of any imperative of self-defense in the moment but rather to obtain a certain vengeful catharsis that had been out of reach for so long. But he concluded:

> It is absurd to blame any German or Austrian Jew for getting his own back on the Nazis. Heaven knows what scores this particular man may have had to wipe out; very likely his whole family had been murdered; and after all, even a wanton kick to a prisoner is a very tiny thing compared with the outrages committed by the Hitler regime.[41]

[40] George Orwell, "Revenge is Sour," in *In Front of Your Nose* (vol. 4, *The Collected Essays, Journalism, and Letters of George Orwell*), eds. Sonia Orwell and Ian Angus (Boston, MA: Nonpareil Books, 2000), 5.
[41] Orwell, "Revenge is Sour," 4.

In this respect, Orwell departs from the moralism in accounts like Nussbaum's. Surely we cannot *blame* the Jewish guards for taking advantage of their newly reversed positions with respect to the Nazis. But Orwell also notes the following:

> But what this scene, and much else that I saw in Germany, brought home to me was that the whole idea of revenge and punishment is a childish daydream… Properly speaking, there is no such thing as revenge. Revenge is an act which you want to commit when you are powerless and because you are powerless: as soon as the sense of impotence is removed, the desire evaporates also.[42]

So he withholds both blame *and* endorsement. What is going on? Like Beauvoir, Orwell holds that revenge contains a sort of contradiction. But where Beauvoir locates the contradiction in revenge's twin incompatible aims, Orwell finds it in its preconditions. Desiring revenge requires that one exist in a state of powerlessness; once one finally gains the power to act on that desire, it vanishes. Why? Orwell points to the pitiful conditions of the Nazis once they were subordinate – "disgusting specimen[s]," "merely neurotic," so far from the monster one imagines conquering from the depths of one's oppression. "Somehow the punishment of these monsters ceases to seem attractive when it becomes possible: indeed, once under lock and key, they almost cease to be monsters."[43] His core claim is that the oppressed do not know how pitiful their oppressors can appear in a certain light once stripped of their immense institutional power. Their fantasies, therefore, are based on an incorrect commitment to the *permanence* of the monstrosity of their oppressors, even through shifting power dynamics. And this is not all bad – perhaps fantasies of revenge buoy the spirits of the oppressed, providing a comforting sense that justice will be served one day. But Orwell's point is that this is

[42] Orwell, "Revenge is Sour," 4-5.
[43] Orwell, 5.

only a provisional, immature, and mistaken position. Revenge's satisfaction, in the very psychological sense, relies on a perception of one's oppressor that only exists when one is in no position to follow through with it; once one is, there is nothing to satisfy.

Orwell goes further and claims that the absence of this shift in perception when the power dynamics are reversed is pathological: "Only the minority of sadists, who must have their 'atrocities' from one source or another, take a keen interest in the hunting-down of war criminals and quislings." There is something detached from reality, still residing in fantasy, where people carry out a revenge plot long after its targets are a real threat. For all others, Orwell holds, the impermanence of viewing one's oppressor as a monster is the mark of a healthy psychology. It means that one is properly acquainted with "the meaning of war," with the true horrors of its violence. The ability to exit the fantasy of revenge and enter the reality and horror of what revenge calls for, *especially* with respect to the recognition that "punishing an enemy brings no satisfaction," is the mark of rationality – it shows that one understands that the imperative of revenge vanishes the moment one is able to carry it out, and their motivational structures reflect this fact.

How do these claims hold up against, for example, the rationality that I have been holding was at the core of the Avengers' revenge plots? There are two important questions to address: First, does revenge contain the contradiction Orwell points out? And second, for those for whom revenge contains no contradiction, should it? Since Orwell's contradiction is an empirical matter (and not a conceptual contradiction as in Beauvoir), we can simply look to the actual world to answer the first question. It is true that many of the stories of revenge with which we are familiar are fantastical – film and

literature so often stand in for catharsis when we wouldn't dare carry out our own revenge fantasies. But the Jewish revenge stories after the war show that *some* instances of revenge are real and perfectly comprehensible: the Jews' desire to assassinate former Nazis did not vanish when they were finally able to carry it out; if anything, it gained a clarity of purpose that was lacking in the chaos of wartime fighting. Orwell attempts to explain away these obvious counterexamples by suggesting that they were merely performative. Of the Jewish guard who insisted that the prisoners repeatedly narrate their sins, he wrote:

> I concluded that he wasn't really enjoying it, and that he was merely – like a man in a brothel, or a boy smoking his first cigar, or a tourist traipsing round a picture gallery – *telling himself* that he was enjoying it, and behaving as he had planned to behave in the days he was helpless.

Perhaps this was true of the guard about whom Orwell made this observation. But if we take e.g. Kovner and his comrades at their word, we do not have a sense that they were self-deceived about either their desires or what would constitute their satisfaction. Their story was not a case in which a fantasy of revenge developed from a place of total passivity; they fought against the Nazis at every step of the way. Neither was it a case in which the continued revenge after the war was merely reproducing an intention set prior to the war's end; indeed their reasons for formulating revenge plots evolved along with the changing post-war conditions. To be sure, the idea was planted from within the horrors of war; but it gained traction in response to the post-war murders of Jewish refugees returning to Poland; the German calls for the return of Hitler; the Nazis' widespread impunity "as they strip[ped] away their uniforms and resume[d] normal

lives."[44] Revenge's desire, in their case, remained stable in the face of changing power dynamics.

Should it have, though? On this second question, there is no obvious internal contradiction to both desiring revenge and being able to carry it out. The preceding paragraphs hopefully have shown this to be the case. So if the normativity in the question is about rational coherence among our aims, there is no issue. Orwell's point was not that revenge generates a contradiction in aim, but that it rests on contrary preconditions (a shifting landscape of power). Is there another sense in which one "should" amend one's vengeful desires?

Orwell does suggest that the constancy of a revenge drive is the mark of an immature "sadist." Though he doesn't blame anyone who finds themselves in this position, he nonetheless implies that it would be better (healthier) if they moved past it; that truly understanding the horror of violence would lead one to retire their revenge plot once it was not "needed." My guess is that Kovner et al. would vehemently disagree: that the sanitized trials, the calls for forgiveness and healing, for orderly adjudication were the things that were removed from the horrors of war. That revenge itself was the only response that properly respected what the Jews had really endured; a true "eye for an eye." Regardless, Orwell's point may very well be true; perhaps all things considered, the psychologically mature thing to do is to set aside one's desire for revenge once one is able to carry it out, precisely because one understands the horror of violence. Recall, though, that we are talking about revenge for atrocities that shatter moral personhood; and the idea of psychological maturity is proper to fully actualized moral beings. Further,

[44] Cohen, *The Avengers*, 189.

it's not clear that it would be rational to prioritize aiming towards psychological maturity over moral emancipation. Practical deliberators are always faced with the delicate balancing of disparate (and often incompatible) moral goods. However these considerations turn out, though, it's important to note the following: perhaps Orwell is correct that revenge is an immature activity; this is consistent with its being an activity of moral emancipation.

While Orwell's discussion of revenge is noteworthy for his refusal to blame revenge-seekers, his assessment of revenge's contrary preconditions is not empirically universal. More importantly, even if he were correct about revenge's immaturity, that does not mean that it is best consigned to fantasy. Real moral advantages can be attained through revenge, and it is rational to consider these moral goods in practical deliberation.

6. Concluding Remarks

I have attempted throughout this chapter to take a closer look at revenge and consider whether its pursuit might be rational under certain conditions. And I have argued, contrary to the familiar positions in the history of philosophy on this topic, that revenge is rational to pursue in the following case: where one has suffered a wrong so egregious that it impacts their moral personhood; where there has been a failure of the state to safeguard that moral personhood (this point is closely related to the previous one); and where revenge, properly targeted, has the potential to actualize that damaged moral personhood. Revenge's rationality does not answer the ethical question of whether we *should* pursue it. That is a question for another time. But it does explain how some people who *do* pursue it can take themselves to be doing something coherent and,

indeed, morally beneficial – in the face of limited options, revenge-seekers aim at a moral emancipation that is rooted in the types of beings we are. Its persistence in our popular imagination, and in our historical records, is by my lights evidence *not* of the irrationality or evil of humankind but rather of the endurance of the human spirit and the imperative of becoming the types of beings we have a right to become.

Abstaining from answering the moral question has also allowed us to see revenge in its natural habitat, so to speak – to see revenge as a robust moral phenomenon that derives its rational coherence from the interpersonal account of moral personhood in which it resides. In the introduction to this dissertation, I presented revenge as the way *in* to this alternative ethical account; the dissertation itself, though, began with the alternative account and ended in revenge. But with this full account in retrospective view now, i.e. one in which we regard the contingency of moral personhood as a vulnerability essential to the human condition, we can see more clearly revenge's place in a life of practical rationality. We are not full moral persons, no matter what. That much was argued for in Chapter 1. And part of what it is to say this is to commit yourself to the idea that the state has failed when our vulnerability to moral erosion becomes actualized – that oppression and the erosion of moral personhood go hand in hand. But this aspect of state failure is only visible if we as political philosophers shed our commitment to inviolability conceptions of the person that deny the contingency of moral personhood. That is what Chapter 2 was about – a detour into the political philosophical literature to show that some of our most promising attempts to account for the injustice of oppression fail to recognize the very vulnerability that explains its deepest harms. For this reason, and unless we incorporate an interpersonal conception of the person into

our theories of justice, our specifications of what justice really requires are going to fall short. In this theoretical gap, and in the real-world injustices to which this gap corresponds, emerges the practice of revenge – revenge responds, in the absence of justice, to the erosion of moral personhood. It is one rational strategy for recovering, or perhaps building for the first time, a fully-actualized moral personhood. And as the types of beings we are – that is, as beings whose nature is to become full moral persons, striving for a flourishing life and subject to moral evaluation along the way, if only the interpersonal conditions are right – this is an end we have a right to pursue. Revenge, in this light, is consistent with our practical rationality because it aims at becoming the type of being for whom this feature takes on moral significance. Far from being a denial of our natures, then, it is in the right circumstances a resounding affirmation.

BIBLIOGRAPHY

Anscombe, Elizabeth. "The Dignity of the Human Being." In *Human Life, Action, and Ethics: Essays by G.E.M. Anscombe*, edited by Mary Geach and Luke Gormally, 64-70. Exeter, England: Imprint Academic, 2005.

Arendt, Hannah. "Personal Responsibility Under Dictatorship." In *Responsibility and Judgment*, edited by Jerome Kohn, 17-48. New York: Schocken Books, 2003.

Aristotle. *Rhetoric*. Translated by C.D.C. Reeve. Indianapolis, IN: Hackett Publishing Company, Inc., 2018.

Baldwin, James. *No Name in the Street*. New York: Vintage Books, 2007.

Beauvoir, Simone de. "An Eye for an Eye." In *Philosophical Writings*, edited by Margaret A. Simons, 237-260. Urbana, IL: University of Illinois Press, 2004.

Brison, Susan J. *Aftermath*. Princeton, NJ: Princeton University Press, 2011.

Callard, Agnes. "The Philosophy of Anger." *Boston Review*. April 22, 2020. http://bostonreview.net/forum/agnes-callard-philosophy-anger.

Cohen, Rich. *The Avengers: A Jewish War Story*. New York: A.A. Knopf, 2000.

Collins, Patricia Hill. "Black Feminist Epistemology." In *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*, 251-272. New York: Routledge, 2000.

Du Bois, W.E.B. *The Souls of Black Folk; Essays and Sketches*. Chicago, A.G. McClurg, 1903. New York: Johnson Reprint Corp, 1968.

Elster, Jon. "Norms of Revenge." *Ethics* 100, no. 4 (1990): 862-885.

Fanon, Frantz. *Black Skin, White Masks*. Translated by Richard Philcox. New York: Grove Press, 2008.

Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press, 2007. See esp. chap. 7, "Hermeneutical Injustice."

Kael, Pauline. "The Current Cinema." *The New Yorker*, March 20, 1971.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. London: Oxford University Press, 2012.

Loden, Barbara. "Barbara Loden Revisited." Interview with Madison Women's Media Collective. *Women and Film*, nos. 5-6 (1974): 67-70.

Marx, Karl. *Economic and Philosophic Manuscripts of 1844*. New York: Prometheus Books, 1988. Quoted in Martha C Nussbaum, *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, MA: Harvard University Press, 2006.

Nagel, Thomas. "The Value of Inviolability." In *Morality and Self-Interest*, edited by Paul Bloomfield, 102-116. Oxford: Oxford University Press, 2011.

Nussbaum, Martha C. *Anger and Forgiveness: Resentment, Generosity, Justice*. New York: Oxford University Press, 2016.

Nussbaum, Martha C. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, MA: Harvard University Press, 2006.

Oliver, Mary. *New and Selected Poems*. Boston, MA: Beacon Press, 1992.

Orwell, George. "Revenge is Sour." In *In Front of Your Nose* (vol. 4, *The Collected Essays, Journalism, and Letters of George Orwell*), edited by Sonia Orwell and Ian Angus, 3-6. Boston, MA: Nonpareil Books, 2000.

Porges, Stephen. *The Polyvagal Theory: Neurophysiological Foundations of Emotions, Attachment, Communication, and Self-regulation*. New York: W.W. Norton & Company, 2011.

Raganelli, Katja, dir. *I Am Wanda*. 1980; Munich: Diorama Film Munich, 1980. Film.

Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.

Rawls, John. *Justice as Fairness*. Edited by Erin Kelly. Cambridge, MA: Harvard University Press, 2001.

Rawls, John. "Kantian Constructivism in Moral Theory." *The Journal of Philosophy* 77, no. 9 (1980): 515-572.

Rawls, John. *The Law of Peoples*. Cambridge, MA: Harvard University Press, 1999.

Reynaud, Bérénice. "For Wanda." In *The Last Great American Picture Show: New Hollywood Cinema in the 1970s*, edited by Thomas Elsaesser, Alexander Horwath, and Noel King, 223-247. Amsterdam: Amsterdam University Press, 2004.

Rosebury, Brian. "Respect for Just Revenge." *Philosophy and Phenomenological Research* 77, no. 2 (2008): 451-471.

Rousseau, Jean-Jacques. *The Social Contract and Discourses*. Edited by G.D.H. Cole. New York: E.P. Dutton and Company, Inc., 1950.

Tessman, Lisa. *Burdened Virtues: Virtue Ethics for Liberatory Struggles*. Oxford: Oxford University Press, 2005.

Young, Iris Marion. *Justice and the Politics of Difference*. Princeton, NJ: Princeton University Press, 1990.

Wallace, Gerry. "Wild Justice." *Philosophy* 70, no. 273 (1995): 363-375.

Williams, Bernard. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press, 1985.