

Essays on Integrated Agency

by

Angela M. Sun

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Philosophy)  
in the University of Michigan  
2022

Doctoral Committee:

Professor Sarah Buss, Chair  
Professor Chrisoula Andreou, University of Utah  
Professor Scott Hershovitz  
Professor James Joyce  
Professor Ishani Maitra

Angela M. Sun

angsun@umich.edu

ORCID iD: 0000-0002-1456-7652

© Angela M. Sun 2022

For my family: 妈妈, 黛黛, 润润, 兔子

## Acknowledgments

As a graduate student, my mind and days were filled with conversations with people who are much smarter than me. This dissertation exists because of them.

First and foremost, I want to thank my advisor, Sarah Buss. I sometimes say that my research is not about what it means to be a good person, but what it means to be good *at being* a person. In Sarah's case, there is no need for this distinction; she is one of the best people and the best at being a person of anyone I know. Over the course of working with Sarah on my dissertation, she taught me to always strive to get things right, but not to be afraid to make mistakes in the process. She taught me to take a firm stance on the issues I write about without losing nuance. She taught me to be read broadly in and outside of philosophy. She taught me to see the world differently: things that were once mundane (fallen leaves, snow, piles of books) are now so beautiful to me that I stop in my tracks to appreciate them. Above all, Sarah taught me never to forget how lucky we are to do this amazing thing called philosophy for a living.

I owe so many thanks to the members of my committee. Chrisoula Andreou graciously took me on as an advisee and helped me hone my views on resolutions and planning agency. She also provided invaluable feedback on a paper I wrote about procrastination. (Sadly, the paper didn't make it into my dissertation. I procrastinated too much while editing it!) I don't think I was ever fully confident in my philosophical abilities until I met Scott Hershovitz, who so clearly valued my contributions that I began to value them myself. My many conversations with Scott improved (and kept me motivated to finish) this dissertation. I owe my fondness for decision theory and its influence on my research to Jim Joyce. Jim patiently tolerated my antics (and vocal espousal of one-boxing in Newcomb's problem) as his student, advisee, and teaching assistant over the years. Not long after I got a call from Ishani Maitra telling me I had been admitted to the graduate program at Michigan, I knew I had found an intellectual role model. Ishani asked all the questions about my dissertation that I've found hardest to answer.

She always pushed me to get to the heart of the problems I was addressing, and I can't thank her enough for it. I also want to give special thanks to Jamie Tappenden, who despite not serving on my committee, frequently provided me with intellectual and professional support while I was writing my dissertation. Jamie's encyclopedic knowledge of philosophy, history, and literature (not to mention his good humor!) made him an invaluable confidant.

I am grateful to Michigan faculty, graduate students, staff, and spouses, past and present, especially Elizabeth Anderson, Abdul Ansari, Shelley Anzalone, Jude Beck, Gordon Belot, Jason Byas, Francisco Calderón, Kelly Campbell, Victor Caston, Mercy Corredor, Derrick Darby, Paul de Font Reaulx, Scott Dennis, Guus Duindam, Anna Edmonds, Maegan Fairchild, Gillian Gray, Emma Hardy, Johann Hariman, Rebecca Harrison, Malte Hendrickx, Danny Herwitz, Josh Hunt, Mariam Kazanjian, Gabrielle Kerbel, Meena Krishnamurthy, David Manley, Eduardo Martinez, Carson Maynard, Calum McNamara, Sarah Moss, Sherice Ngaserin, Lindy Ortiz, Sonya Özbey, Sumeet Patwardhan, Ariana Peruzzi, Josh Petersen, Peter Railton, Mica Rapstine, Julian Rome, Maddie Rome, Laura Ruetsche, Jonathan Sarnoff, Tad Schmaltz, Sarah Sculco, Janum Sethi, Eric Swanson, Valerie Trudel, Adam Waggoner, Maria Waggoner, Ken Walton, Brian Weatherson, Margot Witte, Katie Wong, Yixuan Wu, Sophia Wushanley, and Glenn Zhou. I am grateful to my Wellesley philosophy family—Erich Hatala Matthes, Mary Kate McGowan, Adele Watkins, and Claudia Yau—for showing me what a philosopher is and making me want to be one. And I am grateful to my Minorities and Philosophy co-organizers—Jesi Taylor Cruz, Carolina Flores, Yena Lee, Jocelyn Wang, Elise Woodard, and Jingyi Wu—for being paragons of philosopher-activists.

A few people need extra thanks. Aaron Glasser pushed and inspired me during my last year of graduate school. It's thanks to him that my dissertation about integrity *has* integrity—that I can identify the themes unifying my work—and that I can articulate why my work matters. My (very) long conversations with Cameron McCulloch about philosophy (but mostly about our love lives) were the starting points of so many ideas I explore in this dissertation. Cameron's generosity and unflappable disposition make him a truly remarkable friend. Some of my fondest memories of graduate school are of late nights in the philosophy department lounge with Laura Soter, drinking ciders and co-authoring papers on

population axiology. (For the record, I am still convinced that the Sun-Soter Thesis is the Theory of Everything). Laura has always felt like a kindred spirit, and I'm so grateful that we got to grow into philosophers together. Elise Woodard and I like to joke about being in philosophy sorority where she is my "big sister." I hope she can sense the affection in my voice every time I call her that. I think Elise saw potential in me when I cluelessly stumbled into Michigan philosophy that I never saw myself. She took me under her wing and encouraged me to seize every intellectual and professional opportunity available to me. I'll always credit Elise for my mental and physical health in graduate school; she introduced me to yoga, spin, and the remarkable healing power of venting to one's friends.

And, of course, I am grateful to my loved ones. Zachary Beck was my sounding board as I wrote my dissertation. The best parts of it are all him. Whenever I got too caught up with work, Courtney Schulz was there to remind me that friends are more important than school. I miss her and her singular devotion to others and to the planet every day, and I hope to always act in ways that honor her legacy. The one thing I have been grateful for in the aftermath of her passing has been reconnecting with my childhood friends—Aria Brunetti, Mikaela Orford, Emma Prud'homme O'Neill, Jude Shamsi, and Mia Zanic—who have held me up and kept me strong. My dissertation is about the threads that unify our actions and identities over time. My love for Rahul Walia is one of those threads for me. Through it all, he kept me smiling and laughing at how beautiful and absurd the world is. I've never met anyone who loves life as much as Rahul does. It's a honor to spend mine with him. Since I was a kid, my dad saw a bright future for me that I didn't always see for myself. I am grateful for the sacrifices he made so I could have a peaceful childhood, and for his support that helped put me through college and make it to the University of Michigan.

Before my dissertation, the biggest research project I had undertaken was writing my senior thesis at Wellesley College, which I did alongside my three best friends: Isabella Rosenberg, Rebecca Selch, and Victoria Uren. It's said that we are the products of the people we spend the most time with. I hope that's true and that some of Izzi, Reeb's, and Vicki's magic has rubbed off on me, because they are better than me in every possible way. They are my angels.

Finally, I want to thank my family: Mom and Diana (and, of course, Prince and Tuukka). My mom worked so hard to provide for my sister and me in a

country where she didn't speak the language because she wanted us to have options. Now, she is my biggest cheerleader. She attended almost every talk I gave virtually. (Folks at Michigan often ask me who the mysterious "Julie Wang" lurking in my Zoom sessions is). Her elation when I told her I was offered a tenure-track position at Washington & Lee made me happier and prouder than the offer itself. My sister, Diana, began inspiring me the minute I was born. I copied her taste in music and movies and "borrowed" her clothes. I wanted to play hockey because she played hockey; to write because she wrote; to be smart because she was smart. I wanted to be my big sister when I grew up—I still do. My mom and sister are everything good about me multiplied by a million. To them I dedicate this dissertation.

## Contents

Dedication	ii
Acknowledgments	iii
Abstract	viii
Introduction: Integrity as the Integration of Agency	1
Chapter 1: Practical Death	10
1 Dying for What We Believe In . . . . .	10
2 Core Commitments . . . . .	13
3 Integrity . . . . .	19
4 Practical Death . . . . .	22
5 Practical Mortality . . . . .	34
Chapter 2: Starting Over	36
1 Starting Over . . . . .	36
2 Commitments and the Self . . . . .	42
3 Conflicts Within the Self . . . . .	47
4 Practical Restructuring as the Rational Process of Integration . . . . .	52
5 Starting Over as a Method of Practical Restructuring . . . . .	55
6 Changing Is Hard . . . . .	58
Chapter 3: The Normative Power of Resolutions	61
1 Respecting Resolutions . . . . .	61
2 The Threat of Temptation . . . . .	64
3 Resolutions and Promissory Reasons . . . . .	71
4 The Bootstrapping Problem . . . . .	79
References	84



## Abstract

This dissertation offers an account of the role of integrity in our agency. I argue that the unification of the different facets of our agency into a coherent whole is essential for our self-governance: our ability to be the authors of our own lives and to act in ways that reflect what we stand for. When we are fragmented—when our commitments conflict, or we otherwise fail to live up to what they require of us—we experience inner conflicts that hinder our ability to be self-governing. However, integrity is not the only thing that matters for our agency. Throughout the dissertation, I remain sensitive to the limits of integrity's value, as well as to ways that non-ideal circumstances prevent many people from integrating their agency.

The first chapter, "Practical Death," argues that integrity requires that a person live up to the requirements of her core commitments: the commitments that are essential to who she is. I draw from literary, film, and historical examples to show that an agent who violates the requirements of her core commitments suffers a *practical death*: a condition characterized by psychological crisis, diminished capacity for instrumental reasoning, and undermined self-governance. Because practical death just is the state of lacking integrity, the account I offer in this paper demonstrates that integrity is importantly connected to our ability to be self-governing. However, I conclude by offering some reasons for thinking that agents who suffer practical death can reconstitute their integrity by unifying themselves around a new set of commitments.

The second chapter, "Starting Over," explores a phenomenon where, after a period of depression or personal turmoil, people sometimes express a desire for a "fresh start" or "clean slate." People suffering from internal conflict suddenly cast aside their commitments, sever ties with people they know, and go on adventures around the world. This paper uses this phenomenon of "starting over" as a basis for exploring the power we have to change the commitments that make us who we are through a process that I call *practical restructuring*. Because starting

over is a particularly effective way of resolving inner conflicts through practical restructuring, agents in the grip of internal conflict often desire to start over.

The third chapter, "The Normative Power of Resolutions," argues that resolutions are reason-giving: when an agent resolves to  $\phi$ , she incurs a normative reason to  $\phi$  over and above the reasons that led her to resolve to  $\phi$  in the first place. I argue that resolutions are important because, in the face of temptation, they allow us to stick to our plans and act in ways that reflect what we are truly committed to. On my view, resolution-making—like promising, forgiveness and consent—is a *normative power*: with it, agents have a remarkable ability to alter their normative circumstances through sheer acts of will. To establish my view, I compare the reasons we incur from forming resolutions to the reasons we incur from making promises. One upshot of my view is that it offers a ready response to the bootstrapping problem for mental attitudes, on which if mental attitudes gave us reasons, we could bootstrap any action into rationality simply by acquiring the relevant mental attitude.

## Introduction: Integrity as the Integration of Agency

In Chinese, there's a way of describing a person as someone who *huì zuò rén* (会做人). There's no perfect English translation for the phrase, but roughly, when you say that a person *huì zuò rén*, you say that she is *good at being a person*. This dissertation is about one facet of being good at being a person: having integrity.

This may be surprising. Integrity is typically regarded as a specifically *moral* virtue: a quality that makes someone *good*, not a quality that makes someone *good at being a person* (whatever that means). We think that a person with integrity refuses to compromise her moral principles even in the face of harm or temptation, protests injustice even if doing so will damage her reputation, and keeps her promises even when she stands to benefit from breaking them. I do not deny that integrity so understood is a moral virtue. However, I think there is a more general way of construing integrity that is essential to our theorizing about agency, self-governance, and the good life. On my view, a person with integrity lives up to the requirements of her commitments, whatever her commitments are. This entails that integrity is a matter of *integration*, of unifying the various facets of our agency—including our commitments, actions, beliefs, and intentions—into a coherent whole. And integrity is important to us not only or even primarily because we want to be good people, but because we want to be good *at being* people.

Human beings have a unique capacity for self-conscious reflection. We can *endorse* our actions—we can throw ourselves behind our motivations—or we can *reject* them. According to Christine Korsgaard (1996), our capacity for self-conscious reflection is what gives us authority over ourselves, and therefore what gives rise to normativity in the first place (p. 19).<sup>1</sup> At its core, this dissertation about using our ability take a reflexive stance—to endorse or reject our motivations for action—to *live well*. When we act on the basis of motivations we endorse, we act in ways that reflect what truly matters to us, and we need to act in ways

---

<sup>1</sup> See especially "Lecture One: The Normative Question." See also Frankfurt (1988b) for discussion of the role that self-consciousness plays in purposeful behavior.

that reflect what truly matters to us in order to live well.

My aim in this dissertation is to provide an account of the role of integrity in our agency. I argue that integrity is essential for our *self-governance*; in order to be the authors of our own lives, we must live up to the requirements of our commitments. When we are *fragmented*—when our commitments conflict, or we otherwise fail to live up to what they require of us—we experience inner conflicts that hinder our ability to be self-governing. In what follows, I lay out my account of integrity, explore some of its implications, and explain how the three papers in my dissertation defend the centrality of integrity in our self-governance.

\*\*\*

In addition to being used to describe people who live up to their moral principles, we commonly use "integrity" to indicate something's solidity or soundness, as we might say that a building has structural integrity. We also use "integrity" to indicate that something has not been corrupted, as we might say that a musician who refuses to corrupt her artistic values by "selling out" to a big record label has artistic integrity. But there is a sense in which an artist *necessarily* has artistic integrity; the phrase "an artist without artistic integrity" strikes us as self-contradictory. This is because we think that an artist who lacks artistic integrity—who consistently places competing, distracting, or corrupting values over the value of her art itself—can hardly be said to be an artist at all (Mills, 2018). Our identities give rise to reasons, values, and obligations, and a condition of having an identity is being appropriately responsive to and motivated by the reasons, values, and obligations that stem from it. According to Christine Korsgaard (1996):

A century ago, a European could admonish another to civilized behavior by telling him to act like a Christian. It is still true in many quarters that courage is urged on males by the injunction "be a man!" Duties more obviously connected with social roles are of course endorsed in this way. "A psychiatrist doesn't violate the confidence of her patients." No "ought" is needed here because the normativity is built right into the role (p. 101).

In order for a person to maintain an identity, she must live up to its requirements. For example, I identify as a daughter, sister, friend, philosopher, and reality television aficionado, and if I stopped living up to the demands of those identities—if I stopped calling my mom and my sister on the phone, supporting my friends,

reading and writing philosophy papers, or watching *Survivor*—I would lose them. But I also think that a stronger statement is true. The way I see it, I do not merely *identify* as a daughter, sister, friend, philosopher, and reality television aficionado. I *am* those things. If I ever stopped living up to their demands, I would not only lose them as identities; I would no longer be *me*. It is not merely the case that my identity would be fractured. *I* would be fractured; *I* would lack integrity.

Obviously, I am exaggerating when I say that my commitment to reality TV makes me *me*. (Those who know my longstanding, steadfast devotion to *The Bachelor*, however, might disagree.) But in the other cases, I don't take myself to be exaggerating much or at all. If I gave up philosophy, I might still be me—I and others might still recognize me *as* me—but I would be lost. I would have lost a commitment around which I had organized so much of my life and that made my plans and actions intelligible. And if I was ever forced to act in ways that irrevocably compromised my relationship with my mother or my sister, the person as whom I would emerge would in a very real way not be me. Harry Frankfurt (1999a) has argued that the tragedies we experience when we are forced to betray the commitments that define us "rarely have sequels"; because these tragedies fundamentally change the people we are, there is no "us" who goes on living afterwards (p. 139, n. 8). When Agamemnon at Aulis is forced to choose between his love for his daughter and his love for the army he commands, his "volitional unity" is ruptured, and there is a sense in which "the person he had been no longer exists" (p. 139, n. 8). Derek Parfit (1984) explains that there is no contradiction in saying that a person ceases to exist even though he goes on living by drawing a distinction between *numerical* and *qualitative* identity:

We might say of someone, "After his accident, he is no longer the same person." This is a claim about both kinds of identity. We claim that *he*, the same person, is *not* now the same person. This is not a contradiction. We merely mean that this person's character has changed. This numerically identical person is now qualitatively different (pp. 201-202).

If the person you are can cease to exist when your commitments are compromised, then living up to the requirements of your commitments—having integrity—is central to your being who you are. As Korsgaard (1996) points out, the idea that integrity involves living up to the requirements of one's own commitments squares with the word's etymology:

Etymologically, integrity is oneness, integration is what makes something one. To be a thing, one thing, a unity, an entity; to be anything at all: in the metaphysical sense, that is what it means to have integrity. But we use the term for someone who lives up to his own standards. And that is because we think that living up to them is what makes him one, and so what makes him a person at all (p. 102).

In order to have integrity—in order "to be a person at all"—we must act in ways that our commitments require and refrain from doing what our commitments prohibit. But there is a complication. Many of us think that a life that is rich, fulfilling, and worthwhile involves having numerous commitments that are diverse in nature. We want not only to have a career we care about, but also family and friends, and political, philanthropic, cultural, intellectual, aesthetic, or athletic projects. Because different kinds of commitments realize different values, and because we think that a life well-lived involves experiencing different kinds of value, we should care about having diverse commitments. This isn't to say that a person whose life is devoted to a single project will necessarily feel as though his life is deprived of value. For instance, Alex Honnold—whose equipmentless climb of Yosemite's El Capitan was chronicled in the documentary *Free Solo*—clearly gets all the life satisfaction he needs from rock climbing. In fact, in the film, he discusses his (what many have deemed questionable) unwillingness to do anything that might prevent him from pursuing his single-minded devotion to climbing, including committing to a woman he loves. "I will always choose climbing over a lady," he says (Chin & Vasarhelyi, 2018). However, as remarkable as Honnold's climbing achievements are, when you watch the documentary, it's hard not to feel as though he is missing out on so many things the world has to offer, like romantic love (and perishable food). Susan Wolf (1982) has argued that a life dedicated solely to moral pursuits—that didn't involve anything activities like "reading Victorian novels, playing the oboe, or improving [one's] backhand"—would be "a life strangely barren" (p. 421). But the same can be said of a life devoted to any single activity. A life spent pursuing a single project is missing something important, no matter how deep one's commitment to the project is.

If leading a life that is rich in value involves having multiple commitments that are diverse in nature, then there is a friction between what it takes to live a life that is rich in value and what it takes to have integrity; the more commit-

ments one has and the more diverse those commitments are, the more likely it is that the requirements of the commitments will conflict. This is obvious when a person has commitments that logically contradict each other. A person who is committed both to unwavering moderation and to indulging her every desire can't have it both ways. However, it is also possible for a person's commitments to be reconcilable in *principle*, but irreconcilable in *practice*, due to the constraints with which we are inevitably confronted as creatures with finite time and resources. Here is an example. Many academics (and others with time-intensive jobs) struggle with the decision to start a family. Academia and parenthood realize different values: as an academic, you enjoy an exciting intellectual life; as a parent, you enjoy the unique rewards of watching a child you love become an adult. Experiencing both kinds of value would enrich your life. However, what is required of being an academic and of being a parent tend to conflict. Academia requires attendance at evening lectures and events when parents need to be at home; parenthood requires constant undivided attention to one's children when academics need time to write and read. To be sure, academia and parenthood are not *inherently* incompatible; there is nothing inherent in the concept of "academic" that requires attendance at evening lectures or in the concept of "parent" that requires constant undivided attention to one's children. Rather, these norms are products of our social circumstances. Nevertheless, these norms exert force on people who identify or wish to identify as academics and parents. Because the norms of academia and parenthood conflict, integrity would recommend choosing between them, lest one end up in a position like Agamemnon's. But integrity is not the only thing that matters for our agency. In addition to having integrity, it matters to us that we lead lives that are rich in value. This is the sentiment Sanni McCandless, Alex Honnold's girlfriend, expresses when she says that if Honnold wanted to, he could "have it all: a steady girlfriend and a climbing career" ([Chin & Vasarhelyi, 2018](#)). Who doesn't want to have it all?

Although my dissertation is about the importance of integrity, my aim is not to argue that integrity is to be pursued at all costs. But I do aim to call attention to ways that lacking integrity is bad for us. The papers in this dissertation show that we need to be integrated in order to be self-governing. When we are fragmented—when we fail to live up to the demands of our commitments—we fail to be the authors of our own lives. Because we are finite creatures who value

diverse experiences, integrity is difficult to achieve. But the difficulty we experience integrating our agency reveals something important. It makes what Sarah Buss (2006) calls "the task of living" clear. Our task is to figure out what to care about in the short time we have, what commitments will shape our plans and our lives given that we have limited resources at our disposal to pursue them.

One important implication of my account of integrity is that gives the concept a place in philosophical discussions where it is often excluded. The concepts of integrity and self-governance sometimes get bad raps among feminist philosophers, who for good reason think that these concepts presuppose an overly "atomistic" view of the self, a view on which we exist in a social vacuum, unaffected by our relationships to others.<sup>2</sup> However, my account of integrity suggests that these concepts have a place in feminist discourse. On my view, integrity is about living up to the requirements of one's commitments. But what a commitment requires can vary depending on one's social identities. For instance, the norms governing parenthood are more demanding for women than for men; mothers are expected to spend more time performing childcare duties than fathers are. The norms governing academia are also more stringent for women than for men; women in the male-dominated profession of academia are often held to higher standards in research, teaching, and service than their male counterparts (e.g. Hengel, 2017). This is why women in the academy (and in other time-intensive jobs) are forced to choose between career and family more frequently than their male counterparts. My account of integrity therefore sheds light on one thing that is terrible about oppression: it manufactures circumstances that make it especially challenging for minoritized people to have integrity without compromising richness in life.

\*\*\*

I have said plenty about what integrity is, but I have not yet been clear about why I think integrity is important. In the chapters that follow, I develop my view that integrity is essential for self-governance. On my account, we should care about integrity insofar as we care about being self-governing: being the authors of our own lives and acting in ways that reflect what we stand for. The papers in my dissertation show that agents who *fail* to live up to the requirements of

---

<sup>2</sup> See Stoljar (2018) for an overview of these criticisms.



their commitments and therefore lack integrity experience internal conflicts that hinder—and sometimes compromise—their ability to be self-governing.

The first chapter, "Practical Death," lays the conceptual foundation for my view of integrity as the integration of agency by providing an argument for why integrity requires that we live up to the requirements of our core commitments: the commitments that are essential to who we are. I draw from literary, film, and historical examples to show that an agent who violates the requirements of her core commitments suffers a *practical death*: a condition characterized by psychological crisis, diminished capacity for instrumental reasoning, and undermined self-governance. Because practical death just is the state of lacking integrity, the account I offer in this paper demonstrates that integrity enables our self-governance. However, I conclude by showing that agents who suffer practical death can rebuild their integrity by unifying their agency around a new set of commitments.

The second chapter, "Starting Over," explores a phenomenon where, after a period of depression or personal turmoil, people sometimes express a desire for a "fresh start" or "clean slate." People suffering from internal conflict suddenly cast aside their commitments, sever ties with people they know, and go on adventures around the world. This paper uses this phenomenon of "starting over" as a basis for exploring the power we have to change the commitments that make us who we are through a process that I call *practical restructuring*. Because starting over is a particularly effective way of resolving inner conflicts through practical restructuring, agents in the grip of internal conflict often desire to start over.

The third chapter, "The Normative Power of Resolutions," argues that resolutions are reason-giving: when an agent resolves to  $\phi$ , she incurs a normative reason to  $\phi$  over and above the reasons that led her to resolve to  $\phi$  in the first place. I show that resolutions are important because they allow us to stick to our plans and act in ways that reflect what we are truly committed to—to be self-governing—in the face of temptation. To establish my view, I compare the reasons we incur from forming resolutions to the reasons we incur from making promises. On my view, resolution-making—like promising, forgiveness and consent—is a *normative power*: with it, agents have a remarkable ability to alter their normative circumstances through sheer acts of will. My account of resolutions as reason-giving shows that an agent who unjustifiably acts against her

resolutions and therefore displays agential fragmentation lacks sensitivity to her normative circumstances. One upshot of my view is that it offers a ready response to the bootstrapping problem for mental attitudes, on which if mental attitudes gave us reasons, we could bootstrap any action into rationality simply by acquiring the relevant mental attitude.

\*\*\*

In Chinese, the phrase *huì zuò rén* picks out a set of characteristics that folks living in the English-speaking world might not immediately think make someone "good at being a person." Someone who *huì zuò rén* makes others feel at home in her company. She is a gracious host, intuitively knows what behavior a social situation calls for and acts accordingly, is kind, respectful, and outgoing, and fights for the check when it's time to pay for a meal (a practice I found very strange as a child, but that I eventually came to recognize as a ritual of care). Given the communitarian nature of Chinese culture, it's unsurprising that someone who is good at being a person is someone adept at fostering interpersonal relationships. One of the central virtues of Confucianism is benevolence: *rén* (仁).<sup>3</sup> *Rén* is about living well *through* others; according to Confucius (2019), a person who exhibits *rén*, "wishing to be established himself, seeks also to establish others [... and] wishing to be enlarged himself, he seeks also to enlarge others." It is in part by exhibiting *rén* that a community can attain the cherished value of harmony: *hé* (和). The Chinese character for *rén*, 仁, is a composite of 人, meaning "person," and 二, meaning "two." Moreover, 仁 (benevolence) and 人 (person) are homophonous: both are pronounced *rén*. In Chinese, "I" and "we" are indistinguishable; we cannot understand ourselves independently of our relationships with and conduct toward others.

I discuss Confucius' notion of benevolence here because I think that both Confucius and I are interested in the question of what makes someone good at being a person—how a person can use her human capacities to live well. Not long ago, a philosopher friend told me about traveling to a conference in Vancouver, where she spent the majority of her time shopping, eating delicious food, and getting massages rather than attending talks. When she told me this, I thought, "Damn. This woman knows how to live." Work is important, but so is relaxation,

---

<sup>3</sup> For an overview of the history of the term *rén*, see Chan (1955).

especially when you're visiting a city with food good as Vancouver's. If one of the aims of living is to maximize one's own pleasure (and I do think it sometimes is), my philosopher friend certainly does know how to live. But the aims of living are probably deeper than maximizing one's own pleasure. According to Confucius, one of the aims of living is achieving harmony. Someone who strives to make her community harmonious exhibits benevolence; therefore, someone who exhibits benevolence knows how to live. On my view, one of the aims of living is to be the author of one's own life. Someone who wants to be the author of her own life integrates her agency; therefore, someone who integrates her agency—who acts in ways that her commitments require and refrains from doing what they prohibit—knows how to live. She is good at being a person.

In my dissertation, I occasionally use language that some have described as problematically "managerial." I use phrases like "being strategic about your commitments" and "diversifying your life's meaning portfolio." Some readers have commented on how strange this is; it seems a bit perverse to speak about the things we care about most deeply as if they were commodities to be optimized. I grant that I tend to write bluntly about our commitments and life plans. But I think my managerial language captures something important about the care and attention with which we plan our lives. We want to lead lives that are meaningful, rich, and happy, lives that are full of people and things we care about. But no one is lucky enough to find themselves with commitments that make their lives meaningful, rich, and happy with no planning. For instance, at various points in our lives, we carefully contemplate choices between career paths; we make our decisions based on how much the jobs pay, how much time they'll allow us to spend with our families, and rewarding they'll be—which job will allow you to live the life you want. When we make a choice between careers—or about who to marry, where to live, what to spend our time doing—we are being strategic about our commitments. There may be a paradox of hedonism; the pursuit of pleasure may undermine the attainment of pleasure. But there is no paradox of *living well*. We can and do reflect on how to live—on how to be people. I hope that the essays in my dissertation provide some insight into what this reflection might look like.

## Chapter 1: Practical Death

*Abstract:* This paper argues that integrity requires that a person live up to the requirements of her core commitments: the commitments that are essential to who she is. I draw from literary, film, and historical examples to show that an agent who violates the requirements of her core commitments suffers a *practical death*: a condition characterized by psychological crisis, diminished capacity for instrumental reasoning, and undermined self-governance. Because practical death just is the state of lacking integrity, the account I offer in this paper demonstrates that integrity is importantly connected to our ability to be self-governing. However, I conclude by offering some reasons for thinking that agents who suffer practical death can reconstitute their integrity.

### 1 Dying for What We Believe In

Paragons of integrity are often thought to exhibit remarkable steadfastness in the face of imminent harm and even death. After Martin Luther nailed his "heretical" *Ninety-Five Theses* to the door of the Wittenberg Church, the Pope ordered Luther to travel to Rome to recant his writings. Luther famously refused the Pope's demand and stood by his criticisms of the Catholic Church, saying, "Here I stand, I can do no other." Oskar Schindler was a German businessman who risked arrest, destitution, and death to save the lives of over a thousand Jews in Poland during the Holocaust by employing them in his enamelware and ammunitions factories. After the war, he was asked why he acted as he did, to which he responded, "I had to help them. There was no choice" (Roberts, 1996, p. 91). In 2007, Ragheed Aziz Ganni, an Iraqi Catholic priest, was murdered after he refused to close his church and convert to Islam. Held at gunpoint, Ganni was asked to explain his refusal; he answered, "How can I close the house of God?" (Hattrup, 2018).

The idea that integrity is bound up with a readiness to die for one's commitments is reflected in many familiar adages. In an oft-quoted passage from Plato's

(1901) *Apology*, Socrates says that "the unexamined life is not worth living"—that he would prefer to die than live without engaging in philosophical inquiry (p. 77). Maritime tradition dictates that a captain go down with the ship; if a ship is sinking, the captain must do everything she can to ensure the safety of her passengers and be the last person on board to be rescued. "Get Rich or Die Tryin'," the title of rapper 50 Cent's debut album, epitomizes the hustler mindset. Consider, too, the official motto of New Hampshire embossed on the state's license plates: "Live Free or Die." The motto originated in the postscript of a letter written by American Revolutionary General John Stark; the full postscript reads "Live free or die: Death is not the worst of evils" (Stark, 1860, p. 21). (The French version of the motto, "*Vivre Libre ou Mourir*," was widely used during the French Revolution.)<sup>4</sup>

*Death is not the worst of evils.* For each of the speakers in the examples above, there is something worse than death. For John Stark (and the proud New Hampshireite), it is a life without freedom; for Socrates, a life deprived of philosophical reflection; for the ship captain, failing to take responsibility for her vessel and the lives aboard it; for the hustler, loafing around; for Luther and Ganni, betraying their religious beliefs; and for Schindler, standing by while innocent people are murdered.

I and most of the people reading this will probably never be forced to choose between preserving their lives and honoring their most deeply held commitments. It is unlikely that we will ever be threatened to renounce the things we hold most dear. (I am, however, becoming less and less sure of this. As I am writing, racial justice activists across the country are protesting and facing violent responses from police. More than ever in recent history, it seems as though many of us are at risk of persecution and even death for upholding our beliefs.) It is, nevertheless, hard for me not to wonder: if I was forced to choose between my life and my most deeply held commitments, would I choose as Luther, Ganni, and Schindler did? Or would I save my own life instead?

---

<sup>4</sup> A small side note: strictly speaking, the English translation of "*Vivre Libre ou Mourir*" is not "Live Free or Die," but "To Live Free or to Die"—the French phrase is written in the infinitival form, not the imperative. Expressing commitments as infinitivals has a different effect than expressing them as imperatives. When Star Trekkers aboard the *Enterprise* say "To boldly go where no man has gone before," they are expressing their commitment to exploration without instructing others to share it. In any case, my point in introducing these examples is simply to show that there is a tendency to express one's commitment to a project by communicating one's willingness to die for it. Thanks to Brian Weatherson for bringing this point to my attention.

Here is another way of putting the question. People who choose to die rather than violate their deepest commitments frequently describe their choices as a matter of *necessity*. (Think of Luther who could "do no other" and of Schindler for whom "there was no choice.") The necessity at issue here is *practical*, not logical or metaphysical. Luther could not recant, not because doing so would have been a logical contradiction or a violation of the laws of nature, but because he could not *will* himself to do so. According to Harry Frankfurt (1988b, pp. 181-184), for Luther, recanting was "unthinkable." His commitments made it so that recanting was "off the table" for him.

Some things are worth dying for—this is indisputable. The question that concerns me is not whether we can rationally choose to die for something we care about, but whether we truly count as being committed to something if we are unwilling to die for it. If I was in Luther's position, forced to choose between my life and my deepest commitments, I wish that I, too, would choose the latter, that the former would be "off the table" for me. But for me (and I suspect for my readers), the former would be very much *on the table*. Death may not be the worst of *all* evils, but it does seem to me like a worse evil than dying for some causes I care about—even some causes I care very deeply about. Does the fact that I would prioritize my life over my core commitments—or even entertain the idea of it—threaten my integrity? To have integrity, must there be actions that are unthinkable to me, actions that I would not perform under any circumstance, even if it cost me my life?

\*\*\*

This paper addresses three central questions. First: why do agents violate the requirements of their deepest (or what I call *core*) commitments? Because our commitments are the projects that give our lives meaning and make our lives worth living, it can be puzzling why a person would *choose* to violate their core commitments, especially if they aren't facing a threat of death. Throughout the paper, I will introduce case studies that will offer insight into why people violate their core commitments.

Second: what happens when we violate the requirements of our core commitments? I show that a person who violates the requirements of her core commitments suffers a *practical death*: a condition characterized by psychological

crisis, diminished capacity for instrumental reasoning, and undermined self-governance. I draw on literary, film, and historical examples to develop this phenomenological profile of practical death.

Third: in order for an agent to have integrity, must she be willing to die for her core commitments? Must she have a "bottom line" that she is unwilling to cross, even in the face of death? I argue that the experience of practical death that agents have when they compromise their core commitments should be understood as the experience of lacking integrity. Therefore, integrity consists (at least in part) in living up to the requirements of our core commitments, in staying practically *alive*. However, I offer some reasons for thinking that agents who suffer practical death can reconstitute their agency and their integrity.

## 2 Core Commitments

My focus in this paper is on the consequences of violating our *core commitments*: the projects in our lives that essential to who we are. I wish to distinguish core commitments from *mere commitments*—projects that are important to us and that give our lives structure, but that are not essential to who we are. The qualification that these are "mere" commitments isn't meant to suggest that they are necessarily frivolous or insignificant (though many of them, like my commitment to keep up with reality TV shows, may be). Rather, it is to emphasize that our core commitments are essential properties of who we are, while our mere commitments are accidental ones. (I elaborate on what it means for our commitments to make us who we are in §2 of Chapter Two.)

To illustrate the differences between core and mere commitments, suppose that Andrew is a graduate student in a philosophy program who dreams of one day becoming an academic. A few years into his program, his father becomes ill. If Andrew were to return home to care for him, he would need to drop out of his graduate program. Faced with the choice between caring for his father and completing his degree, Andrew immediately chooses to care for his father. There is no question for him that leaving his graduate program is the right thing for him to do. If he didn't leave his program and return home, *he would not recognize himself*. This shouldn't suggest that philosophy is unimportant to Andrew. He may grieve the loss of his dream of becoming an academic even though the choice

to abandon it is a no-brainer. Nevertheless, on my view, Andrew's commitment to being an academic is a *mere* commitment. Because he recognizes himself in the choice to violate the requirements of being an academic, being an academic is not essential to who he is.

Two concepts that are trying to get at the same idea that I am with core commitments are Christine Korsgaard's "practical identity" and Bernard Williams' "ground projects." (In what follows, notice how, strikingly, both authors invoke death in their descriptions of the concepts.) For [Korsgaard \(1996\)](#), practical identity is "a description under which you value yourself, a description under which you find your life to be worth living and your actions worth undertaking" (p. 101). Most people will value themselves under a jumble of descriptions—as parents, friends, philosophers, members of a religion, etc.—and their practical identity will therefore be multifaceted. Crucially, on Korsgaard's view, our practical identities are the sources of our reasons and our obligations. Our obligations spring from what our practical identities prohibit; for instance, because Andrew values himself as a son, and because to be son he cannot fail to care for his father, he has an obligation to return home. And according to Korsgaard, when an identity is *especially* important to us, it gives rise to *unconditional* obligations. Both Andrew's commitment to scholarship and to being a son factor into his practical identity; both give rise to reasons and obligations for him. However, only the obligations stemming from being a son are unconditional for Andrew. Korsgaard writes:

It is the conceptions of ourselves that are most important to us that give rise to unconditional obligations. For to violate them is to lose your integrity and so your identity, and no longer to be who you are. [...] *That is to be for all practical purposes dead or worse than dead.* When an action cannot be performed without loss of some fundamental part of one's identity, and an agent would rather be dead, then the obligation not to do it is unconditional and complete (p. 102, emphasis added).

Many people I speak to find Korsgaard's claim that to act against one's practical identity is to be "for all practical purposes dead or worse than dead" unacceptably hyperbolic. But this criticism stems from a misreading of Korsgaard's argument. Only in violating the requirements stemming from *the conceptions of ourselves that are most important to us* do we experience practical death. My distinction between core and mere commitments is intended to capture the idea that conceptions of ourselves can be more or less important to us, and I will argue that it is no



hyperbole to say that in violating the requirements of our core commitments we suffer practical death.

Next, consider Bernard Williams' notion of ground projects. For [Williams \(1981\)](#), a ground project is a project that is "closely related to [a person's] existence," "to a significant degree give[s] a meaning to his life," and whose pursuit is "a condition of his having any interest in being around in [the] world at all" (pp. 12-14). Like Korsgaard, Williams invokes death to explain the idea of a ground project:

For a project to play this ground role, it does not have to be true that if it were frustrated or in any of various ways he lost it, he would have to commit suicide, nor does he have to think that. Other things, or the mere hope of other things, may keep him going. *But he may feel in those circumstances that he might as well have died* (p. 13, emphasis added).

Like my distinction between mere and core commitments, Williams' distinction between mere and ground projects—the latter of which provide us with "reasons for living"—allows us to separate Andrew's interest in being a scholar from his interest in being a son. While he values both interests, the former would not be devastating for him to give up.

But I do wish to elaborate on the idea of core commitments beyond what Korsgaard and Williams offer in their accounts. Although an agent's core commitments are, as Williams says, a prerequisite for his interest in living, a core commitment need not be explicitly endorsed. Sometimes, we do not realize that we could not live with ourselves if we did something until we do it and experience the consequences. With perfect self-understanding, we would know which of our commitments were most important to us. But most of us do not have perfect self-understanding, and a project can be a core commitment in the sense I am interested in even if we fail to recognize its centrality to our lives.

Here are two examples to illustrate this point. In the film *Indecent Proposal*, David and Diana are a young married couple caught in the fallout of a bad investment and an economic recession. One day, a mysterious billionaire offers David one million dollars in exchange for a night with his wife. David discusses the offer with Diana, and together they agree that the money is too enticing to pass up. But as soon as Diana leaves for her night with the billionaire, David regrets the decision. He and Diana try to stay together, but David, consumed by jeal-

ousy and self-doubt, goes out of his mind. After he and Diana separate, David performs a series of impulsive, desperate actions that include harassing Diana at work, restaurants, and social gatherings. In this case, we might say that David did not recognize how important his commitment to fidelity was until he violated that commitment and reckoned with its consequences.

We find another example in Fyodor Dostoyevsky's novel *Crime and Punishment*. Raskolnikov, an impoverished student living in Saint Petersburg who kills a local pawnbroker for her money. After he kills her (and her half-sister, an unforeseen witness to the crime), he becomes manic: he hallucinates, falls in and out of uneasy bouts of sleep, and behaves so erratically that he draws suspicion from the local detective working on the case. He eventually confesses to his crime, but does not regret committing it or recognize how it caused his agency to fracture until much later, when he is serving his sentence at a Siberian labor camp. In this case, Raskolnikov didn't realize that something like "moral decency" was a core commitment for him immediately after he murdered the pawnbroker. It took him years to realize that his crime left him broken. The character of Raskolnikov demonstrates that we can continue to be self-deceived about how essential a commitment is to who we are long after we have compromised it. (Aptly, in Russian "*raskol*" means "schism." He is in every sense a man divided.)

I have introduced David and Raskolnikov as examples of agents who are mistaken about what their core commitments are, as they fail to realize that a project is a core commitment for them. There is another way we can be mistaken about what our core commitments are: we might incorrectly believe that a project is a core commitment when it is not. Suppose a lawyer devotes all her time to being the very best at her profession. Her work is her life. One day she is forced to give up her work, and when she does, she realizes that her career didn't matter to her at all. She easily fills up her life with new projects and doesn't experience any agential fracture. This is a case where an agent doesn't realize how *unimportant* a project is to her until after it is compromised.

There is more to be said about what commitments are. I provide a detailed account of commitments in §2 of Chapter Two. Because a thorough explanation of what commitments are isn't required for my argument in this paper, for now, I will provide just a brief sketch of my account. At a minimum, we *care* about our core commitments. Caring, as David Shoemaker (2003) points out, has emotional,

desiderative, and evaluative elements. When you care about something, your *emotions* are tied to the fate of that thing; you are vulnerable to "emotional ups and downs" corresponding to the up-and-down fortunes of it (p. 91). Additionally, when you care about something, you *desire* what promotes or preserves that thing, and you are disposed to *evaluate* that thing positively. Moreover, our commitments play an important explanatory role in our practical lives. They make our actions intelligible; it is in light of our commitments that our plans and the steps we take toward executing those plans make sense. For instance, my commitment to being a philosopher explains why I sit in front of a computer and write for days on end.<sup>5</sup>

Certainly we cannot be committed to something unless we care about it. However, care doesn't tell the whole story about core commitments. Andrew cares about his studies: his emotions are tied to his successes and failures as an academic, he desires what promotes his work, and he is disposed to evaluate academia as a worthwhile enterprise. Yet, on my account, scholarship is not a core commitment for Andrew. This is where the distinction between essential and accidental properties comes in. Our core commitments are the commitments without which we would cease to be who we are. Our mere commitments are projects we value even though we could go on living without them.

One question motivating this paper is the relationship between one's "bottom line"—actions that one would not be willing to perform in any circumstance—and one's integrity. With the idea of core commitments in mind, we are in a position to see what a bottom line amounts to. An agent's bottom line consists in those actions that would compromise her core commitments. If any actions are off the table for an agent, it ought to be the ones that, if performed, would compromise the projects she cares most about and that give her a reason for living.

\*\*\*

It may appear as though I have taken for granted up to this point that all agents *have* core commitments. But is that really the case? Suppose Beth's commitments

---

<sup>5</sup> This isn't to say that our commitments explain *all* our actions. My commitments cannot explain why I, say, scratch an itch. But as I discuss in Chapter Two, our commitments explain all our actions that matter for our self-governance. See pp. 43-46.

matter to her and give her a reason to live, yet she regards all of them as trivial. If all her commitments were compromised, she might be upset, but she knows she will eventually be equally happy with a new set of commitments; there was nothing about her original set of commitments that *made* her who she was. (Note that Beth, who would be just as happy if her commitments were replaced with a different set of commitments, is in a different from a person who would be just as happy if her commitments had been different *to begin with*. I am committed to my dog, Tuukka, and would be crushed if for some reason I had to give him up. However, if my life had gone differently and Tuukka had never joined my family—if my commitment to him had never existed—I know I would be just as happy with another commitment (or no commitment at all) in his place.) Someone like Beth does not have any core commitments. There is nothing she cares about that is essential to who she is. I will address discuss the implications of my view for agents with no core commitments in §5. But for now, I want to offer some reasons for thinking that it's harder to find agents with no core commitments than one might think.

The examples of practical death I discuss throughout this paper are "neat" in an important sense: they involve agents who have a single core commitment that, if compromised, would ruin them. But for most of us, our reason for living does not hinge on a single commitment. A person's core commitments might consist in a group of mere commitments, any one of which she could bear to lose, but that together give her a reason for living.<sup>6</sup> (Indeed, the picture I offer in this paper explains why it can be beneficial, practically speaking, to "diversify your life's meaning portfolio." In doing so, you ensure that no one misfortune can be the end of you.) An agent's core commitments might consist in a single core commitment, a group of core commitments, a group of mere commitments, or some combination of core and mere commitments. I believe there is value in beginning a study on practical death with neat, simple cases, where agents have single core commitments, since it paves the way for more challenging cases. But I do not mean for my focus on neat cases to suggest that cases like these are typical.

---

<sup>6</sup> Here, I follow Williams (1981), who writes, "Of course, in general a man does not have one separable project which plays this ground role: rather, there is a nexus of projects, related to his conditions of life, and it would be the loss of all or most of them that would remove meaning" (p. 13).

### 3 Integrity

What does it mean for a person to have integrity? On one view known as the *integrated self view*, an agent has integrity to the extent that she integrates the various aspects of herself—her desires, projects, attitudes, actions, etc.—into a coherent whole.<sup>7</sup> According to Alfred Archer (2017, p. 437), integrity on this view is to be contrasted with *fragmentation*: a person with integrity makes herself coherent, while a fragmented person is ridden with internal conflict.

One way that agents integrate the various aspects of their identities is by living up to the standards dictated by their commitments. (Agents with integrity "walk the talk," as the saying goes.) Korsgaard (1996) grounds this feature of integrity in the word's etymology:

Etymologically, integrity is oneness, integration is what makes something one. To be a thing, one thing, a unity, an entity; to be anything at all: in the metaphysical sense, that is what it means to have integrity. But we use the term for someone who lives up to his own standards. And that is because we think that living up to them is what makes him one, and so what makes him a person at all (p. 102).

The view that integrity is self-integration may appear to need complicating. On this view, integrity is a *formal* relation; it boils down to mere *consistency*. This view therefore leaves open the counterintuitive possibility that a person with deeply immoral or self-destructive commitments could have integrity.<sup>8</sup> A misogynist who diminishes and demeans women can have integrity in this sense, as long as he consistently lives up to his standards about how men ought to treat women. While writing this paper, I spoke to two dear friends who have battled eating disorders—though, by one friend's account, her battle was not so much against the eating disorder as it was the people who tried to stop the eating disorder from taking over her life. "Starving was easy. The days my mom and my doctors made me eat were the hard ones," she said. My other friend remarked that she had never felt as unified—as integrated—as she did while she had an eating disorder; everything she did served her goal of being thin and in con-

---

<sup>7</sup> For discussion of the integrated self view, see Archer (2017), Cottingham (2010), Pianalto (2012), and Taylor (1981).

<sup>8</sup> This criticism of the view of integrity as self-integration has been raised by Calhoun (2015, pp. 237-238) and McFall (1987, pp. 9-11), among others.

trol of her body. It wasn't until she started recovery that she felt aimless and confused, her agency fractured.<sup>9</sup> If integrity is about behaving in a way that is consistent with one's standards, then agents with immoral or self-destructive aims may have integrity—they may, indeed, be paragons of integrity—since on this view, the moral legitimacy of an agent's aims are irrelevant to her integrity. This seems deeply counterintuitive. As Lynne McFall (1987) points out, we cannot with a straight face say something like, "John was a man of uncommon integrity. He let nothing—not friendship, not justice, not truth—stand in the way of his amassment of wealth" (p. 9).

The view that people with immoral or self-destructive aims can have integrity may be counterintuitive. But I want to resist the conclusion that it's incorrect. Compare the debate over whether immoral and self-destructive agents can have integrity to the debate over whether agents with profoundly immoral projects (Nazis, terrorists) may nevertheless be *courageous*. If courage is just a matter of steadfastness in the face of danger, morally despicable people can certainly be courageous. But many have argued that courage is *not* just a matter of steadfastness in the face of danger: one cannot, on this view, be courageous in one's performance of morally bad actions (e.g., Foot, 2002; Kyle, 2017). As James Rachels (1999, p. 179) points out, in calling the Nazi soldier or the terrorist "courageous," we seem to *praise* them, and this seems wrong.

My preferred solution to this problem is to distinguish between weak and strong senses of courage. Courage in the weak sense is simply exhibiting steadfastness in the face of danger, whatever cause one is standing for. Soldiers defending morally deplorable regimes may be courageous in this weak sense. (We may also use "courageous" in this weak sense to describe someone with morally neutral projects. A person who loves bungee jumping and skydiving might be "courageous," but not in any particularly virtuous or morally laudable sense.) Calling a person courageous in the weak sense is a compliment, but only with respect to her disposition, not her morality. Courage in the strong sense, however, is a moral virtue; we might call this kind of courage *moral courage*. A morally

---

<sup>9</sup> In a memoir detailing her struggle with anorexia and bulimia, Marya Hornbacher (1997) expresses a similar sentiment about unifying herself around her eating disorder: "[Life] is only worthwhile insofar as it relates to your crusade. It was a kamikaze mission. Life and self are far less important than your single-minded goal. 'Thinness' was as good a name as any other for my goal. Twenty pounds, I said. No matter what" (p. 109).

courageous person is prepared to risk life and limb in the service of a morally commendable cause.

Rachels and others might object to this solution by pointing out that it is counterintuitive to call the Nazi courageous even in the weak sense. To praise the Nazi in *any* way seems inappropriate. I agree that it may be *inappropriate* to compliment the Nazi by calling him courageous, but I disagree that it is necessarily *inaccurate*. Pointing out that a Nazi showed courage inappropriately obfuscates his moral crimes, even though "courageous" (in the weak sense) might be an apt description of his character or actions. (Here, I think of how recent documentaries and dramatizations about serial killer Ted Bundy—such as *Conversations With a Killer* and *Extremely Wicked, Shockingly Evil and Vile*, the latter starring Hollywood heartthrob Zac Efron as Bundy—led many people to fixate on Bundy's good looks rather than his moral depravity. Even if Bundy is good-looking, it seems inappropriate to point that out in a complimentary fashion. Any comment that "changes the subject" from Bundy's morality seems remiss.)

Just as a distinction can be made between strong and weak senses of courage, a distinction can be made between strong and weak senses of integrity. To have integrity in the weak sense, one needs to live up to one's standards, whatever those standards are. To have integrity in the strong sense, one must additionally have standards that are morally admirable. Let us call integrity in the strong sense *moral integrity*.

One advantage of distinguishing between mere integrity and moral integrity is that it allows us to explain why we feel as though there is some way that people with morally questionable standards "do better" when they live up to those standards than when they fail to. Suppose there are two politicians whose platforms you consider morally problematic; one politician follows through on the promises they made their constituents while the other breaks many of them. Even if you are glad that the second politician fails to fulfill her morally dubious commitments, you might still condemn her for her hypocrisy. You might begrudgingly grant that the first politician has a backbone—that she sticks to her guns—even though you take issue with what she stands for. You might describe her as having integrity insofar as she lives up to her own standards, but not moral integrity.

Our core commitments are not always moral in nature. Indeed, an agent

who derived all her life's meaning from promoting the welfare of others—who would continue to find meaning in her life without sports, music, intellectual engagement, and other less morally worthy endeavors—would live, according to Susan Wolf (1982), "a life strangely barren" (p. 421). Most people (I hope) are committed to moral decency and would experience psychological distress if that commitment were compromised. But most people also have core commitments that are personal, not moral: a person might be committed to being a hockey player or pianist, or to collecting stamps. Unfortunately, we may also have core commitments that are self-destructive or *immoral*. Our agency does not fracture only when we compromise our moral commitments, but commitments of any kind. A committed pianist forced to give up her music might experience a similarly acute agential disruption as if she were forced to compromise her commitment to being a morally decent person.

If a person does something so morally atrocious that her identity as a morally decent person becomes irredeemable, we might say that she suffers a moral death.<sup>10</sup> But my interest in this paper is in a practical question: what happens to us when we compromise our core commitments, *whatever those core commitments may be*? My focus will therefore be on the notion of *mere* integrity, not moral integrity.

#### 4 Practical Death

In §2, I argued that our core commitments are the projects that give us a reason for living, and that our bottom lines consist in actions that would compromise those projects. In §3, I argued that integrity (in the weak sense) is a matter of integrating the various facets of one's agency—including one's commitments—and making oneself coherent. With these concepts in mind, I now turn to the three central questions motivating this paper. What leads agents to violate the requirements of their core commitments? What happens to an agent after she commits such a violation? And in order to have integrity, must we be unwilling to cross the bottom lines dictated by our core commitments, even in the face of death?

I'll begin by tackling the first question. As I've said before, given that our

---

<sup>10</sup> I won't elaborate on what exactly being morally decent amounts to. But for what it's worth, I think that the bar for moral decency is pretty low; an agent can make plenty of moral mistakes and still be morally decent.



core commitments are the projects that give our lives meaning, it can be puzzling why we would *choose* to violate our core commitments, especially if we aren't facing a threat of death. Literature, film, and psychology can provide us with some insight here.

One reason why an agent might compromise her core commitments is a lack of self-understanding; if an agent does not know what her core commitments are, she might act against them but only later realize what she has done. This partially explains the decisions of *Indecent Proposal's* David and *Crime and Punishment's* Raskolnikov to violate their core commitments, as they didn't realize that they were committed to fidelity and moral decency, respectively, until after they acted against those commitments.

This does not, however, tell the whole story about David and Raskolnikov. In addition to lacking understanding about what their core commitments are, both characters experience *pressure* to pursue a course of action that they would otherwise never consider. When the billionaire offers David a million dollars for a night with his wife, David and Diana are struggling to make ends meet. The billionaire's offer was therefore *coercive*: given David and Diana's precarious financial situation, the offer was too tempting for them to reject.<sup>11</sup> Raskolnikov, too, commits his crime under duress; when he contemplates murdering the pawnbroker, he is destitute and feels extremely guilty about the personal and financial sacrifices his family has made for him to attend university. It makes sense that, in general, an agent would only compromise her core commitments if she were under duress. If an agent freely breaches the requirements of one of her life projects, then it's hard to understand how she could have been committed to the project at all.

There are, however, some exceptions to this—cases where an agent might violate a core commitment without being under duress. First, an agent's circumstances might wear her down to the point where she doesn't realize her core

---

<sup>11</sup> There is disagreement over whether an offer can be so enticing that it is coercive. The disagreement stems from the fact that offers *enhance* one's freedom by giving one more options, and it is unclear whether having more options can restrict one's autonomy in a coercive way. For an argument that freedom-enhancing offers cannot be coercive, see [Feinberg \(1986, pp. 229-268\)](#). I disagree with Feinberg's position. Following Alan Wertheimer, I take it that if a person manipulates someone's options so that they are forced to choose between accepting the offer or suffering an unacceptable consequence, then the offer is coercive, even if the unacceptable consequence is the status quo. See [Wertheimer \(1987, pp. 222-241\)](#).

commitments have been compromised until it is too late. Suppose, for instance, that a college graduate with socialist leanings takes a job on Wall Street, vowing to change the system from the inside. Like a frog in a pot of slowly boiling water, she becomes increasingly worn down by corporate life until her socialist commitments are gone. She is not coerced into giving up her commitment; rather, her circumstances slowly chip away at her commitment until she can no longer recognize herself.

Second, an agent's habits and inclinations—such as an inclination to obey authority—may cause her to fail to be guided by her core commitments in the way she normally would be. Consider the subjects who participated in Stanley Milgram's (1974) experiments concerning the psychological pressure to be obedient. In the experiment, Milgram asked subjects to inflict electric shocks of increasingly higher voltage on a person in another room. Unbeknownst to the test subjects, the person in the other room was an actor, pretending to be hurt by the shocks. There were thirty-two shocks in total, ranging from slight to severe shocks; the final shock was sinisterly labeled "XXX." Sixty-five percent of subjects who participated in the experiment delivered all thirty-two shocks, even when the actor in the other room begged them to stop. In her analysis of the Milgram experiments, Hilary Bok (2008) argues that Milgram's subjects acted as they did because, subconsciously or not, they did not want to face a choice between disobeying the experimenter and continuing to do something they knew to be morally wrong. They were, as Bok writes, "paralyzed by their dilemma: unwilling to look beyond the mere existence of a conflict to the possibility that they might resolve it" (p. 181). Bok's analysis suggests that our unwillingness to meet conflicts head-on may cause us to simply continue with a course of action we're set on—to "act without choosing"—even if we find ourselves acting against our core commitments.

So far, my focus has been on cases where agents *choose* to cross their practical bottom lines. However, it is a sad truth that our core commitments may be compromised due to forces of nature or the actions of others. In her memoir, *Wave*, Sonali Deraniyagala (2013) recounts the devastating experience of losing her husband, two sons, and parents in the 2004 Indian Ocean tsunami. The memoir details her attempt to continue living in a world that took the people she loved most dearly, to continue living despite her core commitments being com-

promised. Susan Brison (1996) has written about how traumatic events—events that cause a person to feel "utterly helpless in the face of a force that is perceived to be life-threatening"—can shatter one's understanding of the world and one's place in it (p. 13). Strikingly, Brison remarks that survivors of trauma often use the language of death and dying to describe their experiences:

Survivors of trauma frequently remark that they are not the same people they were before being traumatized. As a survivor of the Nazi death camps observed, "One can be alive after Sobibor without having survived Sobibor." Jonathan Shay, a therapist who works with Vietnam veterans, has often heard his patients say, "I died in Vietnam." [...] What are we to make of these cryptic comments? [...] How can one die in Vietnam or fail to survive a death camp and still live to tell one's story (p. 12)?

Trauma victims report that they are not the same people they were before the trauma occurred—that their pre-trauma self has died. According to Brison, one kind of death suffered by trauma victims is the death of the "autonomous self": the locus of agency that freely makes choices and wills actions (p. 27). The symptoms of post-traumatic stress disorder—including intrusive memories, involuntary startle responses, and severe anxiety—may undermine the victim's control over herself. According to Brison, this loss of self-control largely explains what a victim means when she says that she is "no longer herself" (p. 28).

There are important differences between cases where agents choose to compromise their core commitments and where agents whose core commitments are compromised by external forces that I do not want to oversimplify. But I do think that looking to the experiences of trauma victims is instructive, one reason being that the circumstances that force agents to act against their commitments are often themselves traumatizing. Consider William Styron's (1979) novel *Sophie's Choice* and the dilemma faced by its eponymous character. When Sophie arrives at Auschwitz, a Nazi doctor says that she may choose one of her two children to be spared from the gas chamber; if she does not make a choice, both will be killed. Sophie spends the rest of her life coping with the horror of the choice she was forced to make; she suppresses memories of Auschwitz and numbs her pain with alcohol and a tumultuous relationship with an abusive boyfriend. (Indeed, it might be inapt to call Sophie's predicament a *choice*. All she could do in the face of such a dilemma was *pick*: she selected one of her children to be spared from death without reflecting on her reasons for making the decision. To take

into account the kinds of considerations she would need to in order to genuinely choose between her children would itself be unthinkable. And abstaining from picking would have been even worse, as it would have resulted in the deaths of both her children. The impossibility of her choice was part of her trauma.)

I hope I have given a clear enough picture of some of the circumstances that lead agents to violate the requirements of their core commitments. Agents may compromise their core commitments if they lack the self-understanding to recognize what their core commitments are; they are forced to act under duress; their circumstances slowly chip away at their commitments; or their habits or inclinations inhibit their ability to be guided by their core commitments. Sometimes, moreover, one's core commitments may be compromised by external forces. This list is not meant to be exhaustive. My aim here is simply to show that it is not so puzzling that agents so compromise their core commitments.

\*\*\*

I will now turn to the second question motivating this paper: what happens when an agent violates her core commitments? In what follows, I argue that when an agent compromises her core commitments, she experiences a cluster of symptoms that together I call *practical death*. I develop a phenomenological profile of practical death on which it is characterized by psychological crisis, diminished capacity for instrumental reasoning, and undermined self-governance. Finally, I argue that the experience of practical death lends credibility to the view of integrity as self-integration, that we ought to understand practical death as the experience of *ceasing to be integrated*. To begin, though, I want to offer one last example of practical death that I find especially illuminating.

In George Orwell's (1949) novel *1984*, Winston Smith is a government clerk in Oceania, a country run by the Party and its elusive leader, Big Brother. Citizens of Oceania are under constant surveillance, and anyone who does not fully conform to the Party's regime is persecuted by the Thought Police. Winston begins an illegal affair with another government worker, Julia, and they promise never to betray each other, even if they are captured by the Thought Police. When, eventually, they are inevitably apprehended by the Thought Police, they are brought to the Ministry of Love, a government division that enforces loyalty to Big Brother by torturing and humiliating alleged thought criminals. Over the course of several

months of torture, Winston confesses to his crimes, including the minute details of his meetings with Julia. But he never stops loving her, and never does anything he believes will harm her, and thus does not take himself to have betrayed her.

Eventually, however, he is brought to "Room 101," where prisoners are exposed to their worst fear as the final step of their "reeducation." Winston's worst fear is rats. O'Brien, the man torturing Winston, explains why this final step is necessary to force prisoners to give up their remaining commitments and dedicate themselves fully to Big Brother:

"By itself," he [O'Brien] said, "pain is not always enough. There are occasions when a human being will stand out against pain, even to the point of death. But for everyone there is something unendurable—something that cannot be contemplated. Courage and cowardice are not involved. If you are falling from a height it is not cowardly to clutch at a rope. [...] It is merely an instinct which cannot be destroyed. It is the same with rats. For you, they are unendurable. They are a form of pressure that you cannot withstand, even if you wished to" (p. 359).

When O'Brien holds the cage full of rats to Winston's face, Winston comes apart: he tells O'Brien that he doesn't care what happens to Julia, that he ought to torture her instead.

The effects of violating his commitment to Julia are devastating for Winston. After he is released from the Ministry of Love, he sits at the same café table drinking gin day after day, mindlessly absorbing the Party's propaganda broadcasted on television. He becomes utterly devoted to Big Brother and his love for Julia vanishes. Julia, who was also threatened with her worst fear in Room 101, experiences a similar agential dissolution. When Winston one day encounters Julia at a park, he puts his hand on her waist and perceives a new stiffness in her body, as if she were a corpse: "He remembered how once, after the explosion of a rocket bomb, he had helped to drag a corpse out of some ruins, and had been astonished not only by the incredible weight of the thing, but by its rigidity and awkwardness to handle, which made it seem more like stone than flesh. Her body felt like that" (p. 368).

The examples I have discussed throughout this paper—of David, Raskolnikov, Sophie, and Winston and Julia—provide a picture of what happens when a person's core commitments have been compromised, when they suffer what I call *practical death*. First, when these agents compromise their core commitments, they

experience *psychological crisis*. The type of psychological crisis they experience varies depending on their circumstances, the core commitment that was violated, and facts about their baseline psychology. David and Raskolnikov become frantic and unpredictable; Sophie becomes dependent on substances and tolerates demeaning treatment from her boyfriend; Winston and Julia become zombie-like. While the symptoms of psychological crises are diverse, a psychological crisis will involve a change in an agent's psychology that threatens her autonomy, the locus of agency that freely makes choices and wills actions. It is widely accepted that autonomous agency—however it is to be understood—can be undermined by forces such as "pain, obsessions, fear, depression, rage, and many other psychological and physiological conditions" (Buss, 2012, p. 661). The examples I have offered show that when an agent's core commitments are violated, she experiences a crisis that brings about the kinds of psychological and physiological conditions that undermine autonomous agency. When an agent experiences a psychological crisis, some of her actions are in an important sense not her own, but the result of external influences (depression, mania, substances) acting upon her. When agents say things like "I don't know who I am anymore," "I miss the person I used to be," and "I feel as though a part of myself has died," they are describing the experience of practical death. Agents who suffer practical death experience psychological crisis and lose self-control, as external forces become capable of overwhelming their agency. This is why agents who have compromised their core commitments may feel like their actions do not represent who they really are.

In addition to psychological crisis, the agents in the examples discussed in this paper experience a *diminished capacity for instrumental reasoning*. In *Crime and Punishment*, there are many striking episodes where Raskolnikov deliriously sets off toward some destination only to find himself meandering through the streets of Saint Petersburg and ending up somewhere other than the place he intended to be. (In one such episode, Raskolnikov finds himself at his classmate Razumikhin's apartment, as if his subconscious was seeking his friend's help.) In 1984, after Winston and Julia say their goodbyes for the last time after enduring the horrors of Room 101, Winston decides to accompany Julia to her subway stop, but abruptly abandons his intention: "He followed irresolutely for a little distance, half a pace behind her. [...] He made up his mind that he would accompany her

as far as the Tube station, but suddenly this process of trailing along in the cold seemed pointless and unbearable" (Orwell, 1949, pp. 369-370). In these episodes, Raskolnikov and Winston both struggle in their post-traumatic states to obey the norms of instrumental rationality that require agents to adopt suitable means to their ends.

These may seem like trivial occurrences. But I don't think it's a coincidence that the characters I have discussed struggle with instrumental reasoning after their core commitments are compromised. It is often thought that one's commitments determine what facts one can regard as reasons for action.<sup>12</sup> One's commitments determine one's character. Without commitments—without what Rawls (1982) calls an "antecedent moral structure"—one's identity would be too thin to choose from the courses of action one is confronted with at every turn, as no aims would appear to be more choiceworthy than any others (Buss, 2006, pp. 381-382). A person's commitments do not merely form her practical identity, but are crucial to structuring her practical deliberations. What Dostoyevsky and Orwell capture so powerfully in the passages I have cited is how aimless and confused an agent feels after she compromises her core commitments. A person who suffers practical death is an ineffective instrumental reasoner, even when it comes to decisions about how to act in the most mundane situations.

The third general feature of practical death that I wish to consider is its undermining effect on an agent's self-governance: her ability to be the author of her own life. The examples I have offered suggest that after an agent compromises her core commitments, she loses control over her behavior. This is why the death that occurs when we violate our core commitments is *practical* in nature; it is a death, in many ways, of a person's ability to decide what to do. But there is a more fundamental connection between self-governance and practical death, which will become clearer in Chapter Two, where I provide a full picture of self-governance. On my view, we are self-governing when we are governed by our commitments. Without commitments, there would be no threads unifying our actions over time—no possibility of a "story" of our lives for us to author in the first place.

I have provided a profile of practical death: the experience agents have when they violate their core commitments. But practical death is, I think, more

---

<sup>12</sup> See, for instance, Buss (2006), Frankfurt (1988b), Frankfurt (1999b), and Williams (1981).

than that: it is the experience agents have when they *cease to be integrated*. The profile of practical death I have offered supports the view that integrity is an integrated self, since it shows that agents who fail to integrate the various aspects of their identities and live up to the standards dictated by their commitments experience agential fragmentation.

The characteristics of practical death I have identified tell us some important things about integrity. Integrity is a background condition for psychological health. Integrity permits a person to reason well. Integrity is required for self-governance. Integrity is not trivial. "Practical death" is an apt name for the experience of ceasing to be integrated not only because of the experience's effect on one's capacity for practical reason, but because when one ceases to be integrated, one is *practically* (that is, *almost*) dead.

\*\*\*

I will now turn to the third question motivating this paper, which asks whether an agent can have integrity only if some actions are unthinkable to her—only if, that is, she is utterly unwilling to cross her bottom line.

In some of the cases I've discussed, agents clearly *choose* to compromise their core commitments. Even though David and Diana were under financial strain when the billionaire made his proposal, it was up to them to decide whether to violate their commitment to fidelity. But in other cases I've discussed, it may appear as though agents have *no choice* but to compromise their core commitments. Take Winston; does he *willingly* betray Julia? In one sense, he does. O'Brien gives Winston a choice: betray Julia or be swarmed by rats. Winston chooses the first option. But we might say that in an important sense, Winston has *no choice* but to betray Julia. O'Brien is right when he says that at some point, "courage and cowardice are not involved." If a mugger threatens you with "your money or your life," it seems perfectly reasonable to say that you had no choice but to hand him your wallet; when most of us consider this scenario, we think that all you can really do is give the mugger your money.

But that's not really the case, is it? "Your life" *is* an option on the table. My point is not that it is worth dying for our wallets (it's not), but just to say that, for some reason, when our lives are on the line, we often fail to regard death as a legitimate option. However, for Luther, Schindler, and Ganni—the paragons of



integrity discussed at the beginning of this paper—death *was* a legitimate option, and faced with a choice between death and compromising their core commitments, they chose the former. If we would fail to do the same—if death is always "off the table" for us—does that make us cowardly, narrowminded, and disunited?

The answer to this question might seem to be obviously "yes." On the view I have presented in this paper—on which an agent who compromises her core commitments ceases to be integrated and suffers a practical death—the answer is "yes." But even intuitively, our admiration for figures such as Luther, Schindler, and Ganni who stand by their commitments even in the face of death suggests that integrity is tied to standing for the things we care about most, even in the face of death (c.f. [Calhoun, 2015](#)). But I wish to offer a few reasons for thinking that integrity is possible even if we are not straightforwardly willing to die for our core commitments.

First, it's worth noting that for most of us, practical death is a distant possibility. We have no reason to fear that we will imminently face a choice between our lives and our core commitments. So even if we wouldn't die for our core commitments, our integrity is safe... for now.

This consideration might assuage some anxiety about the threat of practical death. But it doesn't resolve our uncertainty about the status of our integrity. If we would not die for our core commitments but are simply lucky enough never to be in a situation where we are forced to confront that unwillingness, then there appears to be something objectionably contingent about our integrity (c.f. [Korsgaard, 2009](#), p. 180).<sup>13</sup> We want our integrity to be safe, not just for now, but come what may.

Although I have argued in this paper that an agent ceases to be integrated when she compromises her core commitments, I do think that it is sometimes possible to have integrity (or something like it) even without a straightforward willingness to die for one's core commitments. This is because we often think

---

<sup>13</sup> Here, my view differs from Harry Frankfurt's. In his discussion of the integrity objection to utilitarianism, [Frankfurt \(1988b\)](#), pp. 177-180) argues that as long as a utilitarian does not foresee a scenario where his commitment to maximizing welfare would conflict with his personal commitments, then he can consistently maintain both without losing his integrity. While I agree that the utilitarian *can* retain his integrity, his integrity isn't "safe" in the way he might want it to be.

that negotiating conflicting commitments is one way that agents may *exhibit* integrity. This judgment doesn't change when one of those commitments involves preserving one's mental and physical health. Here is one example. In the years following the Jim Crow era, Joan C. Browning served as a Freedom Rider: a civil rights activist who rode interstate buses to challenge Southern States' failure to enforce integration laws. In a personal essay, [Browning \(2000\)](#) describes the important life experiences she had to sacrifice in order to be a Freedom Rider:

Being part of the Freedom Movement was a life-changing experience. In participating, I lost my only real opportunity for higher education, and I was alienated from my church. I experienced a lifelong separation from my large and loving family, and was set apart from the world in ways that affected all my relationships and employment options. For me, and for many other women like me, participation made us outcasts—women without a home (p. 82).

Racial justice is a core commitment for Browning. She made enormous personal sacrifices in order to be useful to the cause. But if Browning had chosen *not* to join the Freedom Movement—if she had chosen instead to pursue higher education, stay with her family, and find a home—it's hard to say that she would not have integrity, given how much her commitment demanded of her. This isn't to say that we should never be expected to make sacrifices for the things we care about. But there is something to be said about self-consciously negotiating the conflicts between the causes we are committed to and our own wellbeing. A person who abandons a cause as soon as her commitment to it conflicts with her mental and physical health clearly lacks integrity. But a person who *struggles* with that conflict displays *some* integrity, even though, at first blush, she may appear to be more psychically fractured than someone who calmly renounces her commitments at the first sign of friction. This suggests that integrity exists on a spectrum, that we can be more or less integrated depending on how firm we are about our bottom lines.

When we are negotiating a conflict between our core commitments and our wellbeing, how can we know that we are justified in choosing the latter? Here, I find Sarah Buss' ([2020](#)) discussion of moral courage to be a helpful starting point. Buss asks us to consider two scenarios. First, imagine a group of people who have almost no legally protected power over the conditions of their lives and no civil or political liberties. Their days are filled with hard labor; when they are

not working or sleeping, they enjoy the company of their friends and family. It appears as though we have no grounds for criticizing these people. Indeed, given their circumstances, we think that they have "made the best of a bad situation" (p. 2). Consider, next, the "Good Germans" who did not oppose Hitler because they had good reason to believe that they would be ostracized (or worse) if they refused to show support for his regime. Like the group in the first scenario, the "Good Germans" were at the mercy of those in power. However, we widely agree that the "Good Germans" were "cowardly accomplices to evil," blameworthy for "keeping their heads down" and "not being willing to stick out their necks" (p. 2). In both cases, the people involved compromised their commitment to liberty, but only in the second case do we think that the agents lack integrity. The key to figuring out how integrity can be consistent with prioritizing one's wellbeing over one's core commitments seems to lie in figuring out the morally relevant differences between these two cases—an enormous task that I will leave for a future project.

There is one more thing to say about the possibility of integrity even if one is not straightforwardly willing to die for one's core commitments. The account of practical death I have offered in this paper relies on an analogy to literal death. I have argued that when we choose to preserve our lives rather than uphold our core commitments, we *do* die, but in a practical sense rather than a literal one. I now want to point out one important sense in which the analogy fails. Unlike literal death, practical death admits of resurrection. It's true that after practical death, a person is never the same. There will always be a part of her that has died. But people who suffer practical death can and do find reason to live again, reconstituting their agency and their integrity. In *Crime and Punishment*, for example, Raskolnikov faces up to his moral crimes while imprisoned in a Siberian labor camp, reconstituting himself under the loving influence of his friend, Sonya. However, the most poignant examples of agential reconstitution can be found in the literature on trauma recovery. In her work, Susan Brison (1996) recounts her experience of surviving a rape and attempted murder while on a walk in the French countryside. The event irrevocably scarred her. Years later, she found the courage to bring a child into the world, and her relationship with her child helped bring meaning back into her life: "Having him [...] enabled me to rebuild my trust in the world around us. He is so trusting that he stands with outstretched arms,

wobbling, until he falls, stiff-limbed, forward, backward, certain the universe will catch him. So far, it has, and when I tell myself it always will, the part of me that he's become believes it" (p. 32).

## 5 Practical Mortality

I wish to conclude by briefly exploring one upshot of my use of death as a metaphor for the experience an agent has when she compromises her core commitments.

In §2, I touched on the possibility of an agent who regards all her pursuits as trivial. This agent has no core commitments; compromising any of her commitments or any given subset of them would not ruin her. If we suffer practical death only when our core commitments are compromised, then one way to avoid practical death is to avoid having core commitments altogether.<sup>14</sup> If you don't care about anything, then, practically speaking, you can never die. If integrity is something we want to have and practical death is something we want to avoid, then it may appear as though the smart thing to do is stop caring about things altogether. But we don't tend to do that. Rather, we make ourselves vulnerable to practical death by actively *seeking out* cares and commitments. We take up new hobbies, start new careers, and spend nights in bars with the hopes meeting someone we might fall in love with. If we want to protect our integrity, should we stop doing all that?<sup>15</sup>

Here, I find the literature on immortality instructive. Although death is a bad thing (or, at least, something we act as though we want to avoid), Bernard Williams (1973) has argued that immortality is *also* a bad thing, that a life without end would be "unlivable" (p. 100). According to Williams, if we lived forever, we would inevitably end up bored and distanced from our own lives, either because we would end up in a repetitive loop of experiences and relationships we've

---

<sup>14</sup> Dostoyevsky (2003) raises a similar question in *Crime and Punishment* when he writes, "Pain and suffering are inevitable for persons of broad awareness and depth of heart. The truly great are, in my view, always bound to feel a great sense of sadness during their time upon earth" (p. 315).

<sup>15</sup> Here, I think of the episode of *Game of Thrones* in which Grey Worm (commander of the "Unsullied," an elite army famed for the unflinching discipline of its soldiers) professes his love to Missandei: "You are my weakness. When Unsullied are young, the masters learn their fears. But I had no fears. I was never the biggest, never the strongest, but I was bravest, always. Until I meet Missandei from the isle of Naath. Now I have fear."

already had, or because our future projects would be so divorced from our current ones that we would have no reason, at present, to care about them. Therefore, Williams argues, we should desire mortality. An analogous argument can be made about practical mortality. Would it be worth giving up the projects and relationships that matter most to us—and the joys and hardships that come with them—in exchange for perpetual integrity? I, personally, don't think so. Like a life that went on forever, a life without commitments would be intolerably boring and unworthwhile.

There is a famous quote that goes, "I don't want to achieve immortality through my work; I want to achieve immortality through not dying. I don't want to live on in the hearts of my countrymen; I want to live on in my apartment." As Niko [Kolodny \(2013\)](#) points out, there's something about this quote that rings true: "Try as you might to console or distract yourselves with substitutes, what matters—and deep down we all know this—is simply not dying" (p. 3). But when we really stop to think about it, there's a lot more that matters than not dying. If practical immortality comes at the cost of all the meaning our commitments give us, it's hard to imagine that it's something worth having.

While writing this paper, I often found myself thinking about something my mother told me in high school to dissuade me—her naïve, teenage daughter—from dating. Strangely, her advice helped me understand why I desire practical mortality. She said: "If you never have a boyfriend, you can never have your heart broken." As usual, my mother was probably right. But when I look at my partner—and my mother, sister, friends, mentors: all the people I care about and by whom I am vulnerable to being deeply and irrevocably hurt—I do not for an instant regret opening my heart up to them. They make me mortal, and my life is immeasurably better for it.

## Chapter 2: Starting Over

*Abstract:* After a period of depression or personal turmoil, people sometimes express a desire for a "fresh start" or "clean slate." People suffering from internal conflict suddenly cast aside their commitments, sever ties with people they know, and go on adventures around the world. This paper uses this phenomenon of "starting over" as a basis for exploring the power we have to change the commitments that make us who we are through a process that I call *practical restructuring*. Because starting over is a particularly effective way of resolving inner conflicts through practical restructuring, agents in the grip of internal conflict often desire to start over. My discussion touches on various philosophical issues including self-governance, integrity, and the will.

### 1 Starting Over

According to Jean-Paul Sartre (2007), human beings are "condemned to be free" (p. 29); we are free in that we are responsible for everything we do, and our freedom is a condemnation because we do not choose our own existence. According to Christine Korsgaard (2009), human beings are "condemned to choice and action" (p. 1). We cannot escape the need to act. Even when we resolutely stand still, we make a choice about what to do; the need to choose and act is a "simple inexorable fact of the human condition" (p. 2). This paper considers one further respect in which it may seem as though we are condemned: this paper is about whether we are condemned to be *one person over time*.

In one sense, there is no doubt we are so condemned. We are the same people we were an hour, a month, or twenty years ago. We are condemned to be one and the same thing in a metaphysical sense as the people we were in the past. But interpreted another way, it's not so clear that each of us is bound to be the same person over the course of our lives. We are entirely different people now from who we were in our pasts. We look different now than we did when we were,

say, five years old, we know more now than we did then, and—most importantly for my purposes—we have *commitments* now that we didn't have then, and these commitments make us who we are. For instance, I am committed to being a daughter, sister, friend, philosopher, teacher, and reality television aficionado. If for some reason I lost these commitments, I would cease in a very real way to be who I am. And because I did not have these commitments when I was five, I am in a very real way a different person now from who I was then. Derek Parfit (1984) calls these two ways of understanding identity *numerical* and *qualitative* identity. He writes:

We might say, of someone, "After his accident, he is no longer the same person". This is a claim about both kinds of identity. We claim that *he*, the same person, is *not* now the same person. This is not a contradiction. We merely mean that this person's character has changed. This numerically identical person is now qualitatively different (pp. 201-202).

On Parfit's account, a person's qualitative identity consists in the things that constitute his character, that make him *who he is*. A person's qualitative identity excludes his superficial qualities, like the length of his hair or the number of freckles on his face (unless, of course, those things deeply matter to him, such that he would no longer recognize himself *as* himself if he cut his hair or if, after an especially sunless year, his freckles faded). Instead of "qualitative identity," I will call this locus of one's character one's *self*.<sup>16</sup> In what follows, I will use "who one is," "the person one is," and related phrases to refer to the self, rather than to numerical identity.

Parfit is not the only one to remark a distinction between numerical identity and the self. Korsgaard (1996) draws a distinction between "theoretical" and "practical" conceptions of identity: the former is "about what as a matter of inescapable scientific fact you are" while the latter is "a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking" (p. 101). Harry Frankfurt (2002) remarks that it makes perfectly good sense to speak of the "essential nature" upon

---

<sup>16</sup> I dislike the term "qualitative identity" because it suggests that this form of identity depends on all of one's qualities, whether they are deep or superficial. On my view, only changes to qualities that are *central* to who a person is can make her a different person than she was before; changes to a person's superficial qualities do not change who she is. For instance, the length of my hair isn't important to me. If I cut my hair, I would still be the same person I was before the haircut, even though I would, strictly speaking, be qualitatively different.

which a person's continued existence depends, even if the person's life does not (p. 124).<sup>17</sup> To be sure, there is a sense in which numerical identity is more important than—or, at least, metaphysically prior to—the self. As Sarah Buss (2006) points out, "It sounds like a joke, but apparently it needs to be said: living is essential to my being who I am in a way that 'doing philosophy' is not. I am *nothing* if I am not alive" (p. 386). But as long as we are living, we are tasked with making a life. And we need the notion of the self to understand what it means to make a life.

Clearly, we are not condemned to be the same self over the course of our lives in the same way we are condemned to be one and the same thing. We can and do change. Sometimes these changes occur instantaneously (as in the passage quoted above, where a man ceases to be who he was the moment he is injured in the accident); sometimes they occur gradually (as a person might change imperceptibly day by day, until one day she looks back at who she was and no longer recognizes herself as that person). But to what extent are these changes *up to us*? Do we have control over the people we are? My aim in this paper is to provide an account of how changes to our selves occur. In particular, I wish to explore the power we have to change ourselves through a process that I call *practical restructuring*. I contrast practical restructuring with other ways we change over time: *gradual evolution* (the "natural" or "spontaneous" way we evolve over time) and what I have called in other work *practical death* (the instantaneous loss of a commitment caused by compromising the requirements of that commitment). Unlike gradual evolution and practical death, when we engage in practical restructuring, we exercise control over what makes us who we are. Importantly, however, this control not direct. Practical restructuring is a process that involves acting in ways that *facilitate* changes to who we are; there is no guarantee that the process will work. Our power to change who we are is therefore constrained. This might seem like a bleak prospect; it may feel as though we *are* condemned to be the same people over time, after all. But at the end of this paper, I hope to show that the constraints limiting our ability to alter the commitments that make

---

<sup>17</sup> Here, Frankfurt is responding to a criticism from J. David Velleman (2002), who claims that we shouldn't take the idea that a person's commitments make up her essential nature literally, as such literalism can "easily lead to absurdity" (p. 98). Frankfurt explains that there is no absurdity, since we can coherently distinguish between a person's essential nature and what keeps her alive.



us who we are reveals something important about the persistence of our cares. Even if we could completely remake ourselves, I don't think we should want to, because it would make us problematically fickle.

The paper proceeds as follows. In the remainder of this section, I introduce a phenomenon called "starting over" that I take to be a paradigm case of practical restructuring. In §2, I present my account of the self, on which one's self consists in one's commitments. In §3, I argue that feelings of personal turmoil—the kinds of feelings that lead to a desire to start over—can be understood as conflicts between our commitments, and I explain why we are disposed to resolve these conflicts when they arise by engaging in practical restructuring. In §4, I present an account of practical restructuring as the rational process of acting in ways that will facilitate changes in our commitments, contrasting it with gradual evolution and practical death. In §5, I explain why starting over is a particularly effective way of engaging in practical restructuring. In §6, I discuss the limits of our ability to change who we are, even via practical restructuring.

\*\*\*

As a starting point for my discussion, I will consider a phenomenon that I take to be a paradigm case of practical restructuring. I call this the phenomenon of *starting over*. After a period of depression or personal turmoil, people sometimes express a desire for a "fresh start" or "clean slate." Seemingly out of nowhere, people cast aside their commitments, sever ties with people they know, and go on adventures around the world. As a woman living in the US, the idea of starting over immediately makes me think of the "*Eat Pray Love* phenomenon," where, after reading Elizabeth Gilbert's wildly popular 2006 memoir, unhappy middle class women across the country were inspired to quit their jobs, leave their husbands, and move to Bali in hopes of finding love. But starting over need not involve such grandiose travel. It might simply involve starting a new career (or applying to graduate school in philosophy?) because you believe that living in a new city, doing new things, and meeting new people will help you out of the rut you're in. The desire to start over during a dark time in one's life is pervasive. Perhaps that's why narratives about starting over have captured the public imagination.<sup>18</sup>

---

<sup>18</sup> There are too many novels and memoirs about starting over to count. But here are a few books with the general storyline that I have read recently: *The Red Car* by Marcy Dermansky (a novel

Let us get clearer picture of the phenomenon of starting over. What exactly are the "ruts" we find ourselves in that lead to a desire to start over? Why does starting over help us out of them? When Elizabeth Gilbert (2006) decided to leave her husband to travel the world, she had reached a tipping point: her misery had become intolerable. Her problem was not merely that she no longer wanted to be with her husband. Rather, she had become alienated from *all* the plans that made her life intelligible. She had built her "entire life" on the expectation that "after passing the doddering old age of thirty, [she] would want to settle down and have children," and she was "appalled" to discover, years later, that she didn't want any of those things (p. 10, emphasis added). Her revelation made it impossible for her to continue to use her commitments to understand herself and plan her life. She could no longer go on as she was.

Crises like Gilbert's are uniquely challenging in one important respect. Gilbert needed to change her commitments to go on living, but she could not rely on her commitments for any resources to make those changes.<sup>19</sup> She rejected her commitments to being a wife, homeowner, and future mother, but because these were exactly the commitments that made her who she was, she could not rely on them for help moving forward. The only commitment she *could* rely on moving forward was her commitment to live well, in a very broad sense. However, this commitment does not provide a clear picture of what her next steps should be; there are infinitely many life paths she could pursue that would be consistent with her commitment to living well.

In this way, we can contrast Gilbert's crisis with those similar to the one

---

about a woman who leaves her husband in New York to travel to San Francisco and claim a sports car she has inherited), *An Olive Grove at the Edge of the World* by Jared Gulian (a memoir about two Americans who leave their city lives behind to become olive farmers in rural New Zealand), *The Great Alone* by Kristin Hannah (a novel about a Vietnam War veteran who, upon losing yet another job, decides to move his family to Alaska), *Between Two Kingdoms* by Suleika Jaouad (a memoir by a cancer survivor who loses her sense of self after recovering from the disease, and drives across the country meeting other cancer survivors she had met online while in treatment), and *The Unlikely Pilgrimage of Harold Fry* by Rachel Joyce (a novel about a man who leaves his ruined marriage to walk 600 miles to deliver a letter to a dying friend).

<sup>19</sup> L. A. Paul (2014) makes a similar point in relation to a different phenomenon. Paul argues that you cannot rationally choose to undergo a transformative experience because you do not have access, from your present perspective, to the new perspective and accompanying set of preferences you would have after undergoing a transformative experience: your present perspective doesn't provide the resources you need to make a rational decision. Similarly, a person who is alienated from all her commitments has no commitments left to provide the resources she needs to leave those commitments behind.

experienced by Jean Paul Sartre's (2007, p. 30) student, who is torn between serving his country and caring for his aging mother. The student endorses both these commitments, but he cannot do what both commitments require of him. Like Gilbert, the student cannot go on as he was—he cannot remain ambivalent; he must make a choice about which commitment to prioritize. But the commitment he chooses will provide him with the resources to move forward. For instance, if he chooses to serve his country, he may no longer recognize himself as a son to his mother, but he *will* have a clear plan that reflects something he truly cares about; he will enlist. For Gilbert, however, leaving her commitments meant wading into a practical no man's land. She could not use her old commitments to figure out how to give up those very commitments in the way that Sartre's student can. "The only thing more unthinkable than leaving was staying; the only thing more impossible than staying was leaving," Gilbert writes (p. 12).

In a crisis like Gilbert's, a total upheaval of one's life might be necessary to go on. The adventures you go on when you start over tend to be exotic or arduous—certainly outside of your comfort zone. But the exoticism and arduousness of these adventures is the point. For the period of time while you are on your adventure, you don't rely on your old commitments to live your life. During her travels to Italy, India, and Indonesia, Gilbert reflected on her old commitments, but she didn't need them to organize her life. Being distanced from your commitments by embarking on such an adventure can draw things out of you and help you figure out how to go on. From the outside, starting over may look like a vacation: a temporary excursion to a foreign place. But when you start over, you have no intention of returning to your old life; the thought of going back to the way things were is intolerable.

Starting over may lead you to embrace a new set of commitments. Perhaps in decrying the life path you're on, you become someone new. Cheryl Strayed's 100-mile solo hike of the Pacific Crest Trail, which she details in her 2012 memoir *Wild* allowed her to move past her heroin addiction and her grief from mother's passing to become a writer. (Strayed has referred to her mother's death as her "genesis story": the moment her life started (Botton, 2012).) Often, however, leaving your commitments behind leads to a realization about how important or fulfilling those commitments were all along. I think here of Lorelai Gilmore in the *Gilmore Girls* revival, running away in anguish from her hometown of Stars Hol-

low and beau, Luke Danes, to hike the Pacific Crest Trail à la Cheryl Strayed's *Wild* (the book, not the movie, as she makes clear), only to realize how much she really did want to marry Luke and live in her old home. Lorelai starts over—she abandons her commitments and sets off for an adventure in California—but rather than discovering a new set of commitments, her experience leads her to reconnect with her old ones. Starting over can be successful whether it leads you to embrace a new set of commitments or to reaffirm your old ones; what is important, for reasons I explain in this paper, is that you end up integrated around a set of commitments.

Starting over is not always successful. Old habits die hard; it's easy to slip back into familiar patterns of behavior without endorsing them. Here is one example. In Dashiell Hammett's novel *The Maltese Falcon*, private investigator Sam Spade recounts a tale of a well-to-do man named Charles Flitcraft who, after a near death experience at a construction site, abandons his wife and children and disappears.<sup>20</sup> A few years later, Mrs. Flitcraft hires Spade to figure out where her husband went. Spade discovers that after leaving his wife, Mr. Flitcraft traveled for a few years before settling down and marrying a woman very similar to his ex-wife and building almost *exactly the same life* with her that he had with his ex-wife. (Spade recounts the Flitcraft parable to his client, Brigid O'Shaughnessy—who has betrayed his trust in the past—to convey his belief that people never really change and that he will remain mistrustful of her.) The Flitcraft parable is a pessimistic reading of the power of starting over. Even if we can leave behind our commitments and embark on an adventure, it's no guarantee that we'll end up in a better place.

## 2 Commitments and the Self

Because this paper concerns our ability to remake ourselves, I'd like to begin by explaining what I take the self to be. One reason why philosophers are interested in providing an account of the self is to understand *self-governance*. The concept of self-governance is famously elusive. As a starting point, we can say that being self-governing is more than just doing something voluntarily. Rather, to be self-governing, our behavior must be under our control in an important way that I

---

<sup>20</sup> Thanks to Jonathan Sarnoff for pointing me to this example.

hope to explain in this chapter. We *fail* to be self-governing when we are not the authors of our actions, an authority forces us to act in some way, we have no choice but to pursue a course of action, or we are moved by impulses or compulsions "in spite of ourselves."

The idea of self-governance often comes up in discussions of autonomous agency and moral responsibility. The thought goes: we are responsible for actions we perform autonomously, and autonomous agency is presumably self-governing agency. In what follows, I wish to provide an account of the self and what it means to be self-governing independently of its relationship to autonomous agency and moral responsibility. I want to separate the idea of self-governance from that of autonomous agency; as I will explain, there is reason to think that the two concepts describe separate dimensions of agency and action (c.f. [Jacobs, 2003](#)).<sup>21</sup> While autonomous action might be a precondition for moral responsibility, self-governance is an ideal for agency. My aim is not to provide an account of what makes an agent eligible for moral praise or blame; rather, it is to understand what it means to be the authors of our own lives, what it means for our actions to represent who we are and what we stand for. There are many actions that we perform autonomously and should be held responsible for even though they do not represent who we are or what we stand for. For instance, Sarah [Buss \(2021\)](#) points out that we do not stop governing ourselves "whenever we scratch an itch." On my view, it is more accurate to say that we do not stop acting *autonomously* when we scratch an itch. Autonomous action refers to a thinner sense of our control over our behavior than self-governance, which concerns our ability to author our lives and to act in ways that reflect what who we are.

Many philosophers believe it is in virtue of one's ability to *identify* with one's actions that one is self-governing. On these views, agents are self-governing when they act from a perspective that is importantly their own. Consider the unwilling addict introduced by Harry [Frankfurt \(1971, pp.12-13\)](#). The unwilling addict has conflicting first-order desires: he wants both to take a drug and to

---

<sup>21</sup> Jonathan [Jacobs \(2003\)](#), following a Kantian stream of thought, argues that autonomy is "the basis for the distinct status of rational agents" and therefore the basis for understanding (and being bound by) moral requirements (p. 222). Self-governance, on the other hand, is not status-determining; instead, we are self-governing when we are the authors of our lives—when "we fashion ends, we act upon values, we pursue interests and concerns that we conceptualize, and with regard to which we exercise deliberative rationality" (p. 227). This suggests that it is worthwhile to theorize self-governance separately from autonomous agency.

refrain from taking it. But he is not neutral in this conflict: he wants the latter desire to be effective, to constitute his will (or, as Frankfurt puts it, he has a *second-order volition* to refrain from taking the drug). It's true that the addict acts intentionally when he succumbs to temptation and takes the drug in spite of his higher-order volition. But because he does not identify with his action—because his action does not match his second-order volition—there is an important sense in which his action is not his own. Therefore, the unwilling addict is not self-governing. Sripada (2016) advances a "deep self" view of self-governance, on which we are self-governing when our actions emanate from our deep selves: a privileged subset of our attitudes that are fundamental to who we are. According to Sripada, one's deep self consists in one's *cares*, which are to be understood in terms of their characteristic functional role: cares have *motivational* effects (a care about something is a source of intrinsic motivation to promote the wellbeing or achievement of that thing), *commitmental* effects (when you care about something, you're intrinsically motivated to continue caring about it), and an *evaluative* aspect (when you care about something, you're disposed to evaluate it in a positive light) (pp. 1209-1210). Because cares are to be understood in terms of their functional profile rather than in any explicit endorsement of them, it's possible for one's self to consist in attitudes one doesn't realize one has or even that one *denies* one has.<sup>22</sup> Pace Watson (1975), we don't necessarily care about the things we tell ourselves we do in "cool and non-deceptive moment[s]" (p. 215).

On my view, one's self consists in one's *commitments*, which are to be understood similarly to Sripada's cares.<sup>23</sup> Therefore, it is possible to have commitments one is unaware of or that one would even deny one has; the contents of our minds (and hearts) may be inaccessible to us. Raskolnikov of *Crime and Punishment* and David of *Indecent Proposal*—two of the characters I introduced in Chapter One—illustrate this: they do not realize what they are committed to until after they act in ways that the commitment prohibits. It is also possible to believe that one is committed to something without in fact being committed to it, as a hypocrite lacking self-awareness might tout what he thinks is a genuine commitment to philanthropy even though he never actually acts to promote the welfare of others.

---

<sup>22</sup> Other deep self views of self-governance have been proposed by Arpaly and Schroeder (1999) and Watson (1975).

<sup>23</sup> I use the term "commitments" rather than "cares" as Sripada does mostly to remain consistent with my other work.

Our commitments play an important explanatory role in our practical lives. They make our actions intelligible; it is in light of our commitments that our plans and the steps we take toward executing those plans make sense. For instance, my commitment to being a philosopher explains why I sit in front of a computer and write for days on end. This isn't to say that *all* our actions can be explained by our commitments. My commitments do not explain why I scratch an itch. But, as previously discussed, those aren't actions that are relevant to my authoring my life in a meaningful sense. While scratching an itch might be an action that is performed autonomously, it does not manifest my self-governance. The actions that are central to my self-governance, however, *can* be explained by my commitments.

Note, however, that it is not the case that *all* my commitments will play an explanatory role in my actions. I am deeply committed to not killing people: I am motivated to act in ways that achieve this (by not killing anyone) and to sustain this commitment if it ever wanes, and I am disposed to evaluate my refraining from killing positively. Although this commitment is one of the deepest ones I have—I am certain I would suffer a practical death if I ever killed someone—it does nothing to explain my actions, as I never really need to act in ways to satisfy the requirements of the commitment—not killing comes naturally! It does, however, explain certain *omissions*, like why I don't drive after I've been drinking.

In Chapter One, I drew a distinction between *mere* and *core* commitments. I argued that, despite their names, a "mere" commitment is not necessarily a commitment to something superficial and a "core" commitment to something of profound value. Rather, a mere commitment is an accidental property of one's self, while a core commitment is an essential one. If you compromise one of your mere commitments, you may be seriously upset, but you won't be ruined; you will still be able to recognize yourself as yourself. But if you compromise one of your core commitments, your self would cease to exist. Although our core commitments are the most central to who we are, our mere commitments are also part of our selves. We are self-governing when we are governed by *any* of our commitments, not merely when we are governed by the commitments that are essential to who we are; all of our commitments, whether they are mere or core commitments, are part of our "deep selves." A person who is committed to her intramural hockey team is self-governing when she plays in the team's games, even if her commitment to the team isn't a crucial part of her identity. Indeed, many

of us do not have any one core commitment; a person's core commitment may be made up of several mere commitments, any of which she could bear to lose, but that if lost together would cause her to suffer a practical death. It's also possible for a person to have no core commitments at all—that is, for there to be nothing she could lose such that she would suffer a practical death. But these agents can still be self-governing, as long as they are governed by the commitments they *do* have.

On my view, one's self (as I am understanding the self in this paper) is made up of one's commitments. We are self-governing to the extent that we act in accordance with our commitments (and don't act in ways that violate them). In other words, we are self-governing to the extent that our selves—our commitments—govern what we do. This account captures something important about our intuitions concerning our authorship of our lives. To be self-governing, it's not enough that we simply act autonomously. A life that consisted in scratching itches and opening fridges would not be a self-governed life. Rather, a self-governed life is one spent pursuing projects that matter to us, projects we're committed to. Certainly, there might be other uses of "self" that my account does not explain. My aim has just been to show that there is an important sense in which we *are* our commitments.

There are two objections to my account of self-governance that I want to touch on before I move on. First, one might object to the idea that a person is self-governing as long as she acts in accordance with her commitments on the basis that we have little control over the commitments we acquire. If we only have the commitments we do as a result of habituation, indoctrination, or a lack of other options, then it seems implausible that self-governance is a matter of acting in accordance with our commitments. For instance, members of oppressed social groups sometimes develop commitments that prevent them from flourishing in a process known as adaptive preference formation. For instance, a woman living under patriarchy and in poverty might desire to undernourish herself in order to feed her husband (Khader, 2011, p. 74). She identifies wholeheartedly with this desire. Yet, we have reason to hesitate in saying that she is self-governing; the oppressive conditions under which she formed her desires appear to undermine her ability to be the author of her own life. Natalie Stoljar (2018) calls instances like these "feminist 'hard cases,'" as feminist philosophers disagree over whether



women in these cases experience diminished self-governance. On one hand, oppressive circumstances that severely constrain the life paths available to its victims appear to undermine self-governance (e.g. [Oshana, 2006](#)). On the other hand, if women who form adaptive preferences under patriarchy endorse their adaptive preferences and self-consciously reject alternative life paths, it seems as though can be self-governing in spite of their adaptive preferences; according to [Uma Narayan \(2002\)](#), we should not think of women who form adaptive preferences under oppression as "compliant dupes of patriarchy" (p. 420). We can and should criticize the circumstances under which adaptive preferences are formed. However, I do think it's useful to have a concept that describes a person's pursuit of her commitments, even if her commitments were formed under oppressive conditions; this is what I hope the concept of self-governance does. Without such a concept, we run the risk of characterizing victims of oppression as "dupes."

Second, one might object to view because it entails that we are self-governing when we act in accordance with commitments even when we don't explicitly endorse those commitments. If a person has a commitment she isn't aware of, but that nevertheless serves as a source of motivation for her, it may seem strange to say that this agent is self-governing; self-governance seems to require *deliberate* pursuit of one's goals. I agree that self-governance requires deliberate pursuit of our goals in the sense that a self-governing agent deliberately pursue a course of action that is in accordance with her goals. I simply deny that the goal must be explicitly the realization of her commitment. David in *Indecent Proposal* is self-governing in his marriage to Diana, even though he doesn't recognize that monogamy is a fundamental part of that commitment. Certainly, knowing what your commitments are will *help* you be self-governing; knowing what your commitments are makes it more likely that you will act in accordance with them. But knowing what your commitments are isn't necessary for your self-governance.

### 3 Conflicts Within the Self

In the previous section, I argued that one's self consists in one's commitments, and that an agent is self-governing when she is governed by her commitments. In this section, I argue that when a person's commitments conflict, her self is

fragmented, and it is in moments of fragmentation that she may have a desire to reintegrate her agency by starting over.

Sripada (2016) introduces a distinction between *homogeneous* and *mosaic* conceptions of the self. On a homogeneous conception, there can be no conflicts within one's self. Any apparent conflict will always disappear on closer inspection; one will always find that at least one of the apparently conflicting commitments is not genuinely a commitment. By contrast, on the mosaic conception, our selves are "potentially complex, heterogeneous things" and we may have genuine commitments that are in tension with one another (p. 1226). Sripada favors a mosaic conception of the self. I agree with him; I think that we can and do find ourselves in predicaments involving conflicting commitments. But I wish to add an evaluative component to Sripada's mosaic conception of the self. Although it is possible for our commitments to conflict, I think that there is an important sense in which such conflicts are *non-ideal*.

There are various relationships that can hold between our commitments. Frankfurt (1988b) explains that a person is *wholehearted* with respect to a commitment when she fully endorses it as a source of motivation. He contrasts wholeheartedness with *ambivalence*. A person is ambivalent with respect to a commitment when she has commitments that conflict, such that "there is no unequivocal answer to the question of what the person really wants" (p. 165). Consider, once again, Sartre's student, who is torn between serving his country and caring for his aging mother. Because the requirements of each commitment conflict with the requirements of the other, the student cannot fully endorse either commitment as a source of motivation.

When an agent is ambivalent, her self is fractured; the commitments that make her who she is are irreconcilable. In order to be *integrated*, an agent's commitments must be reconcilable, so that she can wholeheartedly endorse them as sources of motivation. This means that, at a minimum, her commitments must be logically consistent; a person who is committed both, say, to unwavering moderation and to indulging her every desire can never have it both ways. However, it also means that her commitments must be reconcilable given the constraints with which she is inevitably confronted as a creature with finite time and resources.<sup>24</sup>

---

<sup>24</sup> Frankfurt (1988a) makes a similar point in his discussion of the possibility of having commitments as a utilitarian. One apparent problem with utilitarianism is that it doesn't allow agents

Because people experience different constraints on time and resources, it may be possible for one person to be both, say, a patriot and a son, given his particular circumstances, even though the same is not possible for Sartre's student. As long as a person is wholehearted in her commitments and her commitments do not *presently* conflict, her agency is integrated.

The internal conflict we experience when we fail to reconcile our commitments impedes our ability to be governed by them. One way it does this is by requiring personal costly code-switching. Here, I think of the experiences of many immigrants in the US caught between the conflicting cultural values of their old and new homes.<sup>25</sup> The persistence of such conflicts can take a toll on one's ability to wholeheartedly embrace one's heritage. The pressure Western culture imposes on immigrants to assimilate causes ambivalence and—in the worst cases—total erasure, when immigrants become "whitewashed" in the course of striving to integrate their commitments.

Some philosophers approach this problem differently. María Lugones (1987, 1990, 1991) has argued extensively for the value of understanding oneself heterogeneously, of constituting one's identity differently in different cultural contexts in a process she calls "'world'-travelling."<sup>26</sup> For instance, she discusses her own experiences as a Latina (and therefore a member of a culture where lesbianism is unthinkable) and a lesbian (and therefore a member of a culture that does not put Latinx values at its core) (Lugones, 1991):

I do not know whether the two possibilities can ever be integrated so that I can become, at least in these respects, a unitary being. I don't even know whether that would be desirable. But it seems clear to me that each possibility need not exclude the other *so long* as I am not a unitary but a multiplicitous being (pp. 138-139).

I agree with Lugones that there is value in understanding oneself as a multiplic-

---

to have a personal character, as agents must be willing to perform every action in *some* circumstance if it is the best of the options available to them; as Frankfurt writes, on a utilitarian view, "nothing can be ruled out in advance" (p. 177). Frankfurt argues, however, that even though in principle a utilitarian must be ready to perform any action even if it violated her values, as long as she is convinced that a situation would never arise in which she would have to perform that action, she can genuinely hold those values (p. 180).

<sup>25</sup> See Mullin (1995) for several examples that convey the same point about personally costly code-switching.

<sup>26</sup> See also Ganeri (2021), whose book discusses the ideas of "heteronymic subjectivity" and "the possibility of multiplicity in subjectivity" in Fernando Pessoa's poetry.

itous being who travels through "worlds," constituting oneself differently in each local context. If one's self totally shifts from context to context, one won't experience ambivalence: when one enters a new "world," one leaves the commitments one had in the old "world" behind. But multiplicity isn't always possible or desirable. For instance, women in professional settings should not need to speak in a deep voice, wear muted colors, or otherwise "check their femininity at the door" in order to be taken seriously. It is a sad fact about sexism that women often experience ambivalence with respect to their desires to express their femininity and to be professional. There is obviously no inherent conflict between femininity and professionalism. Although it may appear as though my claims about the value of integrity undermine the experiences of marginalized people, I think it does the opposite. My account sheds light on one thing that is terrible about oppression: it manufactures circumstances that prevent marginalized people from enjoying the benefits of integrated agency.

Another kind of case where we see ambivalence impede an agent's self-governance involves agents who *repudiates* a commitment. In these cases, an agent is motivated to promote the object of her commitment and to sustaining her commitment to it, and disposed to evaluate it positively, even though, on a higher level, she *rejects* the commitment because it conflicts with another commitment. I think here of children of abusive parents who feel compelled to care for their parents even though they wish they wouldn't or think they ought not to—who care for their parents "in spite of themselves." Repudiating one's commitments amounts to a kind of temporally extended akrasia: when a person wishes that a commitment would not motivate her actions, but she consistently acts in accordance with the commitment in spite of herself. Because her behavior is in some sense not under her control, she is not fully self-governing.

Next, consider agents who are *uncertain* about their commitments. Think, for instance, of a young adult who can afford to go to college and thinks that going is the "right thing to do" even though she isn't certain it's what she really wants, or a person who is approaching "that age" when people tend to get married and "settle down," and therefore proposes to her partner even though she is uncertain she really wants to marry them. Their commitments to being a student and to being a romantic partner organize and give shape to their lives. They might even wholeheartedly believe that the commitments are worthwhile and meaningful.

Yet, they are not wholehearted in their pursuit of the commitments. There is a sense in which they are simply following a life script someone else has written.

But it's also possible that being uncertain about your commitments will have a different effect. Rather than being a source of internal conflict, uncertainty about your commitments might be a source of excitement: in uncertainty there is a promise of discovering who you are.<sup>27</sup> Suppose a young adult goes to college unsure about her major or even if higher education is the right path for her. On one hand, her uncertainty might cause her to flounder—to stay enrolled in courses only because she feels as though it's the "right thing to do." On the other hand, she might see her uncertainty as an opportunity to figure out what she really cares about, or what she wants to "do with her life."

While exciting, this kind of uncertainty about one's commitments is unsustainable. We can't live our lives solely for the promise of what the future holds. But there is value in spending some parts of one's life in a state of commitment flux. These moments give us opportunities to figure out what we care about, and therefore who we are; it is in these moments that a person might start over.

\*\*\*

On the picture I have provided, it is possible to be ambivalent, to have commitments that are in tension with one another. I therefore endorse Sripada's "mosaic" conception of the self. But I added an evaluative element to his account: for the purposes of self-governance, it is better to be integrated than disintegrated. On my view, various psychological and emotional afflictions can be understood in terms of a disintegrated self.

Sripada's choice of the word "mosaic" to describe the self is significant. In a mosaic, individual tiles or stones of various shapes, colors, and sizes are placed together to form a pattern that is on the whole harmonious and beautiful, despite the heterogeneity of its parts. This is the kind of self so many of us strive for. We want to live a life that is rich and full of things we care about—for our lives to look like mosaics. But, as discussed in the Introduction, the more commitments we have and the more diverse they are, the harder it is to unify them. Those of us who strive both to live a life that is rich and to integrate our agency strive for a mosaic—rather than shattered—self.

---

<sup>27</sup> Thanks to Adam Waggoner for pushing me to elaborate on this point.

#### 4 Practical Restructuring as the Rational Process of Integration

In §2, I argued that one's self consists in one's commitments. In §3, I argued that conflicts between one's commitments involve internal conflicts that hinder our capacity to be self-governing. In this section, I want to provide an account of how we resolve internal conflict through a process called *practical restructuring*. Practical restructuring involves acting in ways that facilitate changes in our commitments; it can bring us closer to our commitments or help us reconcile them when they are in conflict, and therefore promote agential integration. It will turn out that the phenomenon of starting over I described at the outset of this paper is one particularly effective method of practical restructuring. This is why so many people experiencing internal conflict have a desire to start over.

\*\*\*

Suppose you are in an unhealthy relationship. Perhaps your partner is judgmental, or jealous; although you love them, you often betray yourself and your responsibilities to your friends, family, and career in order to preserve the status quo. Your commitment to your partner conflicts with your other commitments. As the conflict grows more severe, you decide to end your relationship with your partner.

However, ending your *relationship* with your partner is not the same as ending your *commitment* to your partner. You can't decide to stop caring about someone, for your emotional ups and downs to suddenly stop being tied to theirs and to stop being motivated to act in ways that promote their interests. All you can do is act in ways that make it more likely that you will *eventually* stop being committed to your partner. Magazines are rife with advice on "getting over your ex"; common suggestions include cutting off communication with them, going on a vacation, and redecorating your home. We need to do things like this precisely because we cannot will ourselves to stop caring about people; rather, we must *do* things to help us get over them. Breakups are difficult in large part because we stop communicating with our (ex-)partners but do not stop loving them.

When you act in ways that help you get over your ex, you engage in practical restructuring: you act in ways that facilitate changes in your commitment to them. On my view, engaging in practical restructuring is the one way we can exercise *power* over who we are. To see why this is the case, we can contrast practical

restructuring to what I take to be the two other ways one can become a different person.

Consider, first, *gradual evolution*: the non-rational, passive process by which one's commitments gradually change over time. Gradual evolution is non-rational and passive in that it isn't intentional or goal-directed: it happens "spontaneously." As we go about our lives and have new experiences, we cannot help but evolve in some ways. Gradual evolution is a long process: our commitments change little by little, day by day, until we eventually find that we are different people from who we were before. A person might gradually evolve to become more integrated. She might, for instance, resentfully accept a job she despises, but gradually evolve to enjoy her work and embrace her career as part of her identity. A person might also gradually evolve to become *less* integrated. She might eagerly accept a job she loves, but gradually evolve to hate her work and reject it as part of her identity. There may be no particular moment this change occurs. Eventually, she simply discovers that she is no longer the person she used to be.

The other way a person can change is through *practical death*: the instantaneous loss of a core commitment caused by violating a requirement of that commitment. As I discussed in Chapter One, practical death typically occurs when one is "tested," forced to choose between upholding one's commitments and some other reward or relief. In George Orwell's *1984*, O'Brien offers Winston a choice: betray Julia or encounter his worst fear (being swarmed by rats). The effects of choosing the former are devastating for Winston. He quite literally becomes a shell of his former self: although he occupies the same body that he did before the betrayal, he is a completely different person. Practical death may also occur "accidentally": a person may not realize how deeply he is committed to something until after he suffers the repercussions of violating his commitment. This is what happens to Raskolnikov in *Crime and Punishment*, who does not realize that he is committed to moral decency until he is forced to cope with the psychological consequences of murdering an innocent person. Unlike gradual evolution, which is temporally extended, practical death occurs instantly—the moment one violates the requirements of one's core commitments. When a person fails to live up to the requirements of her commitments and suffers a practical death, she loses her ability to identify with her commitment and therefore becomes less integrated.

(There is also the possibility of an instantaneous change like practical death, but in the other direction, so that a person immediately becomes more integrated than she was before. I think that epiphanies and experiences of divine inspiration might have this feature. But I will save any exploration of this phenomenon for a future project.)

Obviously, we cannot will our gradual evolution; gradual evolution just is the non-goal-directed process by which we change. Less obviously, we cannot will our practical deaths. You suffer a practical death if you are *forced* to choose between upholding your commitment and sparing yourself from harm and find that you are unable to will yourself to choose the former. Although you are given a choice, there is an important sense in which your inability to uphold your commitment isn't "up to you." You find yourself at the limits of your will, unable to push it any further.<sup>28</sup> If you chose to violate a core commitment without being coerced or under duress, it's hard to imagine that it was ever a core commitment for you in the first place. And if you violate your commitments "accidentally" because you are mistaken about what they are, the practical death you suffer as a result is, again, not "up to you." The violation was an accident, after all, not a result of your willing.

Although you cannot will gradual evolution or practical death, there are things you can do to facilitate (and, conversely, prevent) them from occurring. If you want to become a music lover, you can attend more concerts. In exposing yourself to more music, you might find that you gradually come to appreciate music more than you did before. And if you are a music lover and want to stay one—that is, if you want to avoid gradually evolving into someone different—then you can continue attending concerts, and perhaps decline invitations to par-

---

<sup>28</sup> The same is often true for people who are unable to will themselves to choose the *latter*. When Martin Luther refuses the Pope's orders to recant his criticisms of the Catholic Church—saying, "Here I stand; I can do no other"—he is unable to will himself to recant his writings. Even if he wished to spare himself the repercussions of refusing to recant, he could not will himself to do it. According to [Frankfurt \(1988a\)](#), for Luther, recanting was "unthinkable." Therefore, in an important sense, Luther's inability to recant wasn't "up to him." There is an interesting question of why there is an asymmetry in our treatment of these cases. On one hand, there is an intuition that a person who cannot will herself to sacrifice her life for her core commitment should not be blamed for her actions. On the other hand, there is an intuition that a person who cannot will herself to sacrifice her commitment for her life deserves *praise*. This is why Martin Luther's refusal to recant is often brought up in discussions of integrity. Although I won't be able to address this question here, I wanted to flag it for future discussion.



take in activities that might lead you to take on commitments that conflict with the requirements of being a music lover. For instance, if a friend invites you to join her weekly poker match that will sometimes prevent you from going to concerts, you might decline the invitation. Although we cannot will changes in who we are, we can act in ways that *facilitate* (and prevent) changes in who we are. This is practical restructuring: the goal-directed process by which we act in ways that bring us closer to the people we want to be.

One form of practical restructuring is what Agnes Callard (2018) calls *aspiration*. On Callard's view, aspiration is the process of acquiring a new value. (To use my terminology, aspiration is the process of acquiring a new commitment.) It is, however, just one instance of practical restructuring, since it involves the rational acquisition of a very specific value. Callard explicitly limits the scope of aspiration to cases where agents have sufficient antecedent access to the value they are seeking to acquire; she doesn't think that a young adult who sets out for Europe to "find herself" is an aspirant (p. 7). But even if this young adult is not aspiring in Callard's sense, she *is* engaging in a form of practical restructuring. Specifically, she is engaging in the kind of practical restructuring that is the focus of this paper: starting over.

Starting over is more "complete" than aspiration: aspiration is the rational process by which a person comes to value something new, while starting over is the rational process by which a person *becomes* someone new. Moreover, the aims of starting over may be more indeterminate than those of aspiration: when a person aspires, she strives to understand a particular value, but when a person starts over, she may not have an idea of the what values she is seeking to acquire. Callard argues that unlike aspiration, embarking on a process of self-discovery by, for instance, going on an adventure "rarely feel[s] like work" (p. 7). I hope the account of starting over I offer in this paper shows that isn't true. Remaking oneself *is* work. Putting oneself back together when one is fragmented—integrating one's agency—is an achievement.

## 5 Starting Over as a Method of Practical Restructuring

On my view, starting over is an instance of practical restructuring. When you are experiencing internal conflicts, a change in your routine and environment can

help you integrate your agency. In starting over, you may be searching for a way to reconcile the conflicting commitments that have caused you inner turmoil—perhaps some distance from your usual circumstances will give you the perspective you need to resolve these conflicts. Alternatively, in starting over, you may be trying to figure out what your commitments are in the first place. In any case, when a person starts over, she seeks to integrate her agency: she wants a clearer picture of who she is.

The aims of starting over are usually vague. When people start over, they say they are "soul searching" or "figuring out who they are." Often, it isn't possible to be any more specific than this (if you knew where your soul was, you wouldn't need to search for it, after all). There are a lot of different ways soul searching can look. A person who wants to start over might embark on a 1000-mile solo hike, end a relationship that has for a long time been at the center of her life, or apply to a graduate program in hopes of pursuing a new career. Any one of these paths might help her "find herself." Because starting over is a vague project in this sense, there are many means a person can rationally adopt to her end of "finding herself." There is one sense in which it matters a great deal which path she takes: the path will determine her future commitments—the kind of person she becomes. But in another sense, the way she chooses to start over *doesn't* matter. Her aim is to find herself. But the truth is that there are many selves to be found in different places. This is why starting over is imbued with possibility. There are many versions of ourselves out there to look for, and we aren't sure which one we'll end up finding.

Starting over, then, is a form of practical restructuring with a vague goal; aspiration, by contrast, is a form of practical restructuring with a well-defined one. The vagueness of the aims of starting over may make it seem like a futile enterprise. But there are several reasons why starting over is an especially effective method of practical restructuring.

First, starting over involves a total disruption of one's routine and environment and is therefore a powerful method of breaking patterns of problematic thought and behavior. Verplanken et al. (2018) argue that people going through a period of significant change (like moving house or starting a family) get "in the mood for change" and are more likely to be receptive to behavior change interventions. According to Borkovec and Sharpless (2004), the more novel our

environment, the more we are able to control our responses to stimuli, rather than simply respond in the way we are used to. When we engage in practical restructuring, we strive to understand or change the commitments that make us who we are. It's unsurprising that distancing ourselves from our commitments by changing our routines and environments makes that process easier.

A second reason why starting over is an effective method of practical restructuring is that changing one's environment often involves meeting new people whose ways of life are different from those one is familiar with. These people might serve as "mentors" whose ways of life we can imitate or emulate. [Callard \(2018\)](#) discusses at length the value of mentors in the process of aspiration. On Callard's view, when a person seeks to acquire a new value, she necessarily starts from a place of lacking understanding of that value. This is why mentors are important: they serve as models for aspirants to emulate. Mentorship looks different for starting over than it does for aspiration. A person who aspires to be a music lover can easily identify mentors for her to emulate—she need only look for a music lover. Because the aims of starting over are vaguer than those of aspiration, mentors for people starting over are less readily identifiable, but more diverse. A person who starts over may look to others who have previously uprooted their lives in order to find themselves for inspiration. She may also find mentors in the people she meets as she starts over who expose her to different ways of life. She might discover in these people the person she wants to make herself into.

A third reason why starting over is an effective method of practical restructuring has to do with the *adventures* that typically make up a person's experience of starting over. Starting over is difficult: breaking up with a long-term partner and living on one's own can be devastating; going on a 1000-mile solo hike is no joke. And while it takes place over an extended period of time, starting over is not permanent. When a person starts over, she intends to find a solution to her inner turmoil and eventually end up with a clearer picture of who she is. These two features of starting over—difficulty and temporal boundedness—prompt a comparison to *games*. In his account of the value of games, C. Thi [Nguyen \(2019\)](#) argues that when we play games, we temporarily take on "disposable" ends for the sake of the experience of pursuing them. These experiences are (supposed to be) fun because they require us to engage with difficulty in a controlled setting ([Hurka, 2006](#); [Nguyen, 2019](#)). Importantly, on Nguyen's view, by providing op-

portunities for us to pursue disposable ends in controlled environments, games allow us to experiment with different forms of agency. Game designers create agencies for players to inhabit: they create spaces for players to move, obstacles for them to overcome, as well as the powers and motivations they have to overcome those obstacles (p. 437). One of the values of games is that they allow us to try on agencies for size and teach us which ones to apply in the right circumstances.

Starting over is not a game. It requires huge investments of time and emotion, and often money. But I think there is a meaningful connection to be drawn between participating in games and starting over. Because starting over often involves going on an adventure, it forces us to undergo challenges that are outside of our comfort zone, that push us to occupy new agencies and overcome different obstacles than what we are used to. In doing so, it can give us insight into what we're capable of. In the following passage, Cheryl [Strayed \(2012\)](#) reflects on how the physical capabilities she discovered she had while hiking the Pacific Crest Trail helped her recognize her emotional capabilities:

I was amazed that what I needed to survive could be carried on my back. And, most surprising of all, that I could carry it. That I could bear the unbearable. These realizations about my physical, material life couldn't help but spill over into the emotional and spiritual realm. [...] It had begun to occur to me that perhaps it was okay that I hadn't spent my days on the trail pondering the sorrows of my life, that perhaps by being forced to focus on my physical suffering some of my emotional suffering would fade away (p. 92).

The trials we undergo when we embark on adventures are applicable in unexpected ways to the challenges we face in our ordinary lives. The revelations that people who start over report having reveals something important about the structures of the agencies we inhabit. If we can translate our survival of physical challenges into beliefs about our abilities to survive emotional ones, it's no surprise that the experience of starting over is therapeutic for so many people.

## 6 Changing Is Hard

Up to now, I may have given the impression that I am optimistic about our capacity to engage in practical restructuring and exercise control over the integration

of our commitments. I have offered examples of people who found solace after distancing themselves from their old commitments by starting over. But I want to be realistic about the limits of the power we have to change the commitments that make us who we are.

Let us return to the *Maltese Falcon* case I outlined in the first section of this paper. Recall that after a near-death experience, Mr. Flitcraft abandons his wife and children and travels for a few years before settling down with a new woman and building a life that was almost identical to the one he previously had. We aren't told what happened during his travels. All we know is that something about narrowly missing being crushed by a construction beam prompted him to reflect on his life and decide to leave his family:

His second wife didn't look like the first, but they were more alike than they were different. [...] I don't think he even knew he had settled back naturally in the same groove he had jumped out of in Tacoma. But that's the part of it I always liked. He adjusted himself to beams falling, and then no more of them fell, and he adjusted to them not falling (Hammett, 1930, p. 64).

Perhaps Flitcraft came to the realization that he was a family man through and through; after roaming around for a few years, he figured out that what he really wanted was exactly what he previously had. A less cheerful possibility is that he simply slipped back into the life he was used to; despite his best efforts, he couldn't make the changes to his life that he wanted. Either way, his commitment to having a family stuck with him, even when he tried to shake it off.

After going soul-searching, people tend to return home to the lives they were previously leading. After her travels to Italy, India, and Indonesia, Elizabeth Gilbert returned home to New York and resumed her career as a writer. In John Updike's novel *Rabbit, Run*, the protagonist, Rabbit Angstrom, a washed-up high school basketball star now in midlife, leaves his wife and newborn child and takes up with another woman, Ruth, only to return to his wife, leave his wife for a second time, go back to Ruth, and leave Ruth for a second time. (Needless to say, *Rabbit, Run* is a rather frustrating read.) In his philosophical self-help book, *Midlife*, Kieran Setiya (2017) describes the "nostalgia, regret, claustrophobia, emptiness, and fear" he felt at the prospect of spending his life in academia, caught in an endless cycle of researching, writing, and publishing (p. 2). Yet, Setiya remains a philosophy professor at MIT (a fact that I am pleased about,

since I am a big fan of his work). Many of us will go through periods of personal turmoil during which we doubt our commitments. But after these periods pass, our previous commitments beckon us back. To use Detective Sam Spade's terminology, there may be times in our lives when we feel as though the beams that hold us together—that keep us integrated—begin to fall. Although this sometimes motivates us to build completely new structures, more often than not, the beams simply *stop falling*. Traveling the world, moving somewhere new, or going on an adventure might help us discover who we truly are, but, for better or for worse, who who we truly are may be exactly who we were all along. Explaining exactly why it's so hard to successfully practically restructure one's self is beyond the scope of this paper, but I hope to one day have some answers.

The limits of our ability to change who we are may seem like something to lament, especially to those of us who are currently experiencing the kinds of internal conflict that prompt desires to engage in practical restructuring and start over. To be clear, I do think these limits are lamentable when our commitments are immoral, self-destructive, or otherwise harmful. But I also like being the kind of person who struggles to aim to break off from my commitments. Thank goodness my philosophical musings have led me to conclude that I am likely to return to my current commitments if ever, in a state of crisis, I abandon them. I hope to not be fickle, to be the kind of person whose devotion and kindness to loved ones is permanent. And I hope that my loved ones' devotion and kindness to me is permanent, too.

## Chapter 3: The Normative Power of Resolutions

*Abstract:* This paper argues that resolutions are reason-giving: when an agent resolves to  $\phi$ , she incurs a normative reason to  $\phi$  over and above the reasons that led her to resolve to  $\phi$  in the first place. I argue that resolutions are important because, in the face of temptation, they allow us to stick to our plans and act in ways that reflect what we are truly committed to: to be self-governing. On my view, resolution-making—like promising, forgiveness and consent—is a *normative power*: with it, agents have a remarkable ability to alter their normative circumstances through sheer acts of will. To establish my view, I compare the reasons we incur from forming resolutions to the reasons we incur from making promises. One upshot of my view is that it offers a ready response to the bootstrapping problem for mental attitudes, on which if mental attitudes gave us reasons, we could bootstrap any action into rationality simply by acquiring the relevant mental attitude.

### 1 Respecting Resolutions

In *Being and Nothingness*, Jean-Paul Sartre (1956) introduces us to a gambler who, fearful of financial ruin and the disappointment his losses will cause his family, resolves to quit gambling. But the next day, as he approaches the game table, he "suddenly sees all his resolutions melt away" (p. 69). He realizes that his resolution to stop gambling is totally inefficacious; it has no binding effect on him. He thinks:

[Y]esterday I even had a synthetic appreciation of the situation (threatening ruin, disappointment of my relatives) as *forbidding* me to play. It seemed to me that I had established a *real barrier* between gambling and myself, and now I suddenly perceive that my former understanding of the situation is no more than a memory of an idea, a memory of a feeling (Sartre, 1956, p. 70, emphasis in original).

So, the gambler takes a seat at the table and asks to be dealt in.

There is a strong intuition that Sartre's gambler has done something wrong.<sup>29</sup> One might think that the gambler has acted wrongly because he has acted against his better judgment, that this is a classic case of *akrasia*. But that diagnosis wouldn't be quite right. The gambler does not act against his better judgment. When he approaches the game table, he unambiguously changes his mind about what he has all-things-considered-reason to do; in the grips of temptation, he decides that he *ought* to gamble. The right thing to say in this case seems to be that the gambler's change of mind, though "genuine," is *unjustified*, and it is in unjustifiably changing his mind that the gambler has acted wrongly.

Not all of us are trying to overcome gambling addictions. But we all find ourselves in situations that are like the gambler's in relevant ways. Every project we undertake is spread out in time and requires planning for us to complete. Think of writing dissertations, fighting injustice, going for hikes, and cultivating relationships with family and friends. If every urge or inclination caused us to change our plans, we would be unable to make meaningful commitments and pursue the kinds of projects that enrich our lives. Every time in the past year I resolved to work on my research but changed my mind when it came time to sit down and do it (there were, I am sad to admit, many such occasions), I acted wrongly in a smaller but similar way to the gambler.

Some philosophers have tried to explain what agents like the gambler do wrong by appealing to the benefits of planning agency that they forgo in failing to respect their resolutions. Michael Bratman (1983, 1987, 2000, 2007) that without the ability to follow through on our resolutions, we would lose pursue projects that are spread out in time. Other philosophers have tried to explain what agents like the gambler do wrong by remarking that we are creatures with limited cognitive resources who do not always have the time, information, or clarity of mind to deliberate at the moment of action. According to Luca Ferrero (2010), one benefit of our planning agency is that it allows us to divide deliberative labor intertemporally, delegating decisions about what to do *now* to our *past selves*, who are

---

<sup>29</sup> I use "wrong" not in the moral sense, but simply to indicate that there is some way that the gambler's behavior has gone awry. However, someone with Kantian intuitions and strong views about what we owe to ourselves might believe that the gambler *has* acted wrongly in the moral sense, for if we have obligations to cultivate capacities that are distinctive of our human nature—including, perhaps, our *planning* capacities—we may wrong ourselves when we fail to respect resolutions that we made in our own best interest (Kant, 2017).



better positioned to make those decisions than our present selves are. When the gambler ignores his resolution to refrain from gambling, he fails to respect his better-positioned past self to whom the decision about what to do was delegated.

While I think it's true that agents who fail to respect their resolutions forgo the valuable benefits of planning agency, I don't think that this provides a satisfying explanation for the intuition that Sartre's gambler acts wrongly *right then*, when he feels his resolve melt away. There are certainly advantages to being the kinds of creatures who follow through on resolutions. But this doesn't explain why we are wrong to disregard our resolutions in *any particular instance*. After all, it's implausible that one slip up could compromise one's planning agency. As long as the gambler gets back on track, there's no reason to think that he has forgone any of the benefits of planning agency. My aim in this paper is to explain why it can be wrong to fail to respect resolutions in a particular instance. I argue that agents who unjustifiably act against their resolutions fail to be *responsive to reasons* in the way that well-functioning agents ought to be. On my view, resolutions are *reason-giving*: in resolving to  $\phi$ , one incurs a normative reason to  $\phi$  over and above the reasons that led one to resolve to  $\phi$  in the first place. In other words, the fact that one resolved to  $\phi$  ought to count as a consideration in favor of  $\phi$ ing at the time of action. Therefore, an agent who unjustifiably acts against her resolutions lacks sensitivity to her normative circumstances.

There is evidence in natural language and patterns of reasoning that resolutions are reason-giving. Suppose my partner asks, "Why are you doing yoga at six in the morning?" "Because last night I told myself I would" seems like a perfectly satisfactory answer. More often than not, the mere fact that a person resolved to  $\phi$  is enough to explain her  $\phi$ ing. Of course, my partner might ask why I made the resolution in the first place. I would cite the reasons that led me to form the resolution to practice morning yoga in the first place: I want to develop a habit of exercising first thing in the morning, get better at clearing my head, etc. But, to be perfectly honest, at six in the morning, those reasons will be inaccessible to me. Like the gambler, "my former understanding of the situation [would be] no more than a memory of an idea, a memory of a feeling." What *will* be accessible to me is the memory of forming a resolution, and I have to trust that my past self made the right decision in forming it. My resolution to do yoga in the morning is the reason *on which I act* when I do yoga in the morning.

One reason why philosophers have been hesitant to explore the possibility that resolutions are reason-giving is worries about *bootstrapping*: if resolutions gave us reasons, the worry goes, then we could bootstrap any action into rationality simply by resolving to do it. On the account I offer in this paper, the reasons we incur from forming resolutions are similar in important ways to the reasons we incur from making *promises*. A comparison between the reasons we incur from resolutions and from promises will reveal that resolutions, like promises, may bootstrap, but not problematically. If the comparison is apt, my account of resolutions has a ready response to a major objection to the thesis that resolutions are reason-giving.

The paper proceeds as follows. In §2, I defend the value of our power to form resolutions, explaining why we need it to shield our capacity for self-governance against the threats of temptation. In §3, I develop a comparison between the reasons we incur from resolutions to the reasons we incur from promises. Because promises are widely regarded to be reason-giving, the comparison demonstrates that we have good reason to think that resolutions are, too. On my view, resolution-making turns out to be a *normative power* alongside phenomena like promising, consent, and forgiveness. With the ability to make resolutions, we have a remarkable ability to alter our normative circumstances through sheer acts of will. In §4, I use the comparison between resolutions and promises to show that my account of resolutions does not lead to problematic bootstrapping.

## 2 The Threat of Temptation

Resolutions are often thought to be a special kind of intention formed to stand firm in the face of future contrary inclinations (Holton, 2009, p. 10). Sartre's gambler does not merely intend but *resolves* to quit gambling because he anticipates that he will be *tempted* to return to the game table. When you make a New Year's resolution—to call your parents more, save money, or pick up a new hobby—you form an intention to do something where some barrier threatens your success. The barriers in question might be substantial. Perhaps you have been calling your parents infrequently because, after balancing your childcare duties with your professional ones, you find yourself with very little free time. Sometimes, however, the barriers in question are trivial. The barrier preventing you from picking up

a new hobby might be laziness or the allure of watching another episode of TV. Think, too, of showing resolve in the face of an obstacle. When you resolve to jump off the diving board, you decide to stop deliberating and *just do it*. We show resolve when we commit to a course of action *and* to refraining from further deliberation about the matter because reopening deliberation might cause us to change our minds.

On this view, resolutions are a species of intention. Like resolutions, intentions more generally are known to resist reconsideration and revision; they are characterized by what [Bratman \(1987\)](#) calls *stability*. Unlike resolutions, however, intentions are not always formed specifically with the aim of resisting temptation that threatens to cause us to deviate from our plans. This is why it seems silly to call an intention to, say, make pizza rather than pasta for dinner a resolution. In cases like this, there is no temptation that threatens to throw us off course; we simply need to choose between pizza and pasta and for the intention we form to be stable so we have something to eat for dinner.

Another difference between resolutions and mere intentions concerns the processes by which we form them. As Richard [Holton \(2009, p. 53\)](#) points out, many intentions are formed automatically. When you see slow traffic ahead you form an intention to switch lanes, and you scan your mirror for an opportunity to do so without giving any conscious thought to the matter. We do not only form intentions automatically in mundane cases. One feature of expertise is the ability to automatically form intentions to perform highly refined, complex actions in familiar scenarios; a professional hockey player might recognize a play her teammate is starting and get in formation without any explicit thought ([Bergamin, 2017](#); [Bermúdez, 2017](#); [Logan, 2018](#)). Resolutions, by contrast, must be formed explicitly; one cannot resolve to do something without giving it conscious thought.

Why do we need a special kind of intention to shield our plans from the forces of temptation? In this section, I argue that, as creatures with temporally extended projects, we need a capacity like the one to form resolutions in order to make self-governance possible. Without something like the capacity to form resolutions, we would not be able to carry out plans to pursue projects that matter most to us when temptation threatens to throw us off course.

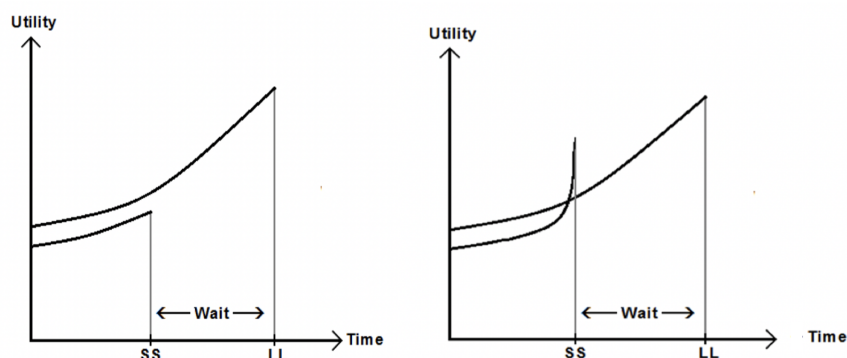
\*\*\*

Temptation is widely regarded to cause our desires—and in some cases our judgments about what we have most reason to do, all things considered—to shift. Suppose you set some time aside this evening to watch TV. Because you would like to get some work done afterwards, you resolve to limit yourself to one episode. But, cozy on the couch, you are tempted to watch a second, and when the streaming service automatically starts playing the next episode, you succumb to temptation. You decide that in light of your heightened desire, you now have most reason to watch more TV rather than work.<sup>30</sup> As discussed in §1, the judgment shifts induced by temptation make it difficult to explain why your behavior is irrational. If temptation heightens your desire to watch TV and causes your all-things-considered judgment of what to do to shift, then it may be rational for you to continue watching TV, even though it goes against your resolution not to.

We see the kind of irrationality in other forms of preference shifts. Consider temporal discounting. People often judge a good to be more valuable as they approaches the time of its consumption. Some forms of discounting are rationally permissible—strictly speaking, there’s nothing wrong with preferring to eat a chocolate bar *now* than in four hours—but others are not. Consider the figures below, which contrast *exponential* and *hyperbolic* discounting:

---

<sup>30</sup> I am introducing a new example of temptation, rather than continuing to use Sartre’s gambling addict, because addictive temptation may be different in important respects from standard cases of temptation. According to [Holton \(2009, pp. 103-111\)](#), while standard temptation should be characterized in terms of judgment shifts, addictive temptation should be understood in terms of a disconnect between an agent’s desires and what they judge best; an addict will act on the basis of desires regardless of whether they judge that action to be best. I think that Sartre’s analysis of the gambler—especially his description of the gambler’s thought process when he abandons his resolution—is illuminating, and I wanted to include it as a central case in the paper. But I also want to respect the possibility that there are fundamental differences between addictive and regular temptation.



Exponential vs. hyperbolic discounting<sup>31</sup>

Suppose that you are deciding whether to receive a reward of \$30 in one week or \$50 in one month; \$30 is therefore a "smaller but sooner" good (SS) and \$50 is a "larger but later" good (LL). If you are an exponential discounter, your preferences for SS and LL are represented by the figure on the left. Although your desires for the two goods grow as you approach the time you receive the monetary reward, they grow at the same rate. Because LL never becomes more desirable to you than SS—because your *comparative ranking* of the two goods remains fixed over time—you will be patient and hold out for LL despite your increasing desire for SS. However, if you are a hyperbolic discounter, your preferences are represented by the figure on the right. In this case, your desire for SS *spikes* immediately before consumption, surpassing your desire for LL. Therefore, you will succumb to impatience and choose SS when it is offered to you, rather than hold out for LL. Experiments conducted by [Ainslie and Haslam \(1992\)](#) suggest that most people are hyperbolic discounters, not exponential ones. The allure of having something *now* causes not only our preferences but also our judgments about what to do to shift. When faced with the possibility of receiving a reward *now*, we decide that we ought to take it, even if there is a promise of a bigger reward later. There is a strong intuition that hyperbolic discounting is irrational. However, if you judge that you have most reason to accept SS over LL *at the time of choice*, it is difficult to say why that is the case.

To explain why it's irrational to give into temptation and impatience despite the judgment shifts it causes, we can begin by appealing to a distinction drawn by [Bratman \(2007\)](#) between "policies" and "singular valuing." A policy is a rule—or

<sup>31</sup> I borrow these graphs from Jim Joyce.

what Bratman calls a "general intention"—you impose on your practical reasoning (p. 272). You might, for instance, have a policy of limiting yourself to one episode of TV on evenings you also have work to get done. According to Bratman, you ought to respect this policy even if, faced with the vivid and immediate prospect of a second episode, you temporarily come to value two episodes "just this one time." Bratman argues that a person's policies ought to be prioritized over her singular valuing in her practical reasoning because of their "agential authority" (pp. 265-268). A person's policies constitute a point of view that is, in an important sense, *her* point of view: because they establish where she stands—because they establish her "practical framework"—they have a privileged status as being peculiarly her own (p. 265). Often, a person's present evaluative ranking has agential authority; she ought to act in ways that get her what she currently wants most. But this isn't always the case. According to J. David Velleman (1992), "[a] person can be alienated from his values, too; and he can be alienated from them even as they continue to grip him and to influence his behavior" (p. 472). Because our agency is temporally extended and we often have singular valuing when we are faced with temptation, we need to find ways to respect the agential authority of the policies that structure our ongoing practical reasoning. This isn't to say that you cannot have a policy of watching only one episode of TV on evenings you need to work *except on certain special occasions*, on days that really are different from other days. You would, of course, need to be careful to make sure that today *really is* different from other days and that you are not acting on the basis of a singular valuing disguised as a policy.

The idea that some perspectives or attitudes better represent "who we really are" than others has philosophical precedent. In §2 of Chapter Two, I discussed various views of moral responsibility on which we are responsible only for actions that emanate from our "deep selves": a privileged subset of our attitudes that are fundamental to who we are (Arpaly & Schroeder, 1999; Frankfurt, 1971; Sripada, 2016; Watson, 1975). On these views, a person is not morally responsible for actions that emanate from her "shallow self"—like the unwilling addict's drug use or a kleptomaniac's compulsive shoplifting—because those actions are not, in an important sense, *hers*; they do not represent who she is. As with all philosophical views worth taking seriously, these positions are controversial. But I've provided a sketch of them here to show that there is nothing new about the idea that

the perspectives and attitudes we identify with have a privileged status in our practical and moral lives, even if we do not always act in accordance with them. This is the threat of temptation: when temptation causes a spike in desires with which we do not identify, we act in ways that are not representative of who we are and what we stand for.

In this sense, we can compare following through on resolutions to what [Arpaly and Schroeder \(1999\)](#) call *inverse akrasia*. An inverse akratic is someone who performs the *right* action despite their all-things-considered judgment to do otherwise. For instance, Huckleberry Finn does the right thing when he refrains from turning in Jim, a runaway slave, despite his judgment to do otherwise. According to Arpaly and Schroeder, we can praise inverse akratics like Huck Finn even though we shouldn't blame regular akratics because Huck Finn acts on the basis of desires that are "well-integrated" into his overall personality; Huck Finn's motives for refraining from turning in Jim—"compassion, loyalty, a sense of friendship"—represent who he is, even though he does not explicitly endorse them (p. 177). Similarly, when we follow through on resolutions, we act in accordance with reasons that are well-integrated into our agency, that represent who we are and what we stand for. This is true even if there is something "akratic" about following through on the resolution insofar as we act against our *current* all-things-considered judgment of what to do.

One question that arises at this point is why we should *care* about acting in accordance with the perspectives and attitudes with which we identify. What's wrong with giving into temptation? You may resent your desire to watch a second episode of TV because it makes it harder for you to get your writing done. But you do get *something* good out of giving into your temptation to continue watching TV: the pleasure of a second episode. What reason do you have to prioritize your "deep" desire to get writing over your "shallow" desire to watch another episode of TV if you get something good either way, or even if you would get *more* pleasure out of watching more TV than you would writing?

The answer to this question lies, I think, in our interest in being self-governing. We *want* to be the authors of our own lives, be in charge of what we do, and act in ways that reflect what we stand for. In previous chapters, I have argued that in order to be self-governing, we must act in accordance with our commitments; we must have integrity. The challenge is that there are often

things that are easier or more fun to do than what our commitments require. You can be committed to writing a novel but want to spend the evening watching TV; you can be committed to doing yoga in the morning yet prefer to sleep in; you can be committed to monogamy but pursue a crush on someone other than your partner. If we want to act in accordance with our commitments—if we want to be self-governing—we must find ways to stick with our commitments in the face of temptation.

Perhaps temptation only poses a problem for those of us who have an *interest* in being self-governing. Mikhail [Valdman \(2010\)](#) argues that there's no reason why we ought to have such an interest. On his view, while there is intrinsic value in having our decisions reflect our commitments, there is no intrinsic value in making the decisions ourselves. Therefore, if you were given an opportunity to relinquish your self-government to a committee that would make better decisions for you given your commitments than you would yourself, you should, from a prudential standpoint, accept it. Many philosophers have argued otherwise. For them, self-governance is valuable regardless of whether we desire it; in order to lead a fully good life, we must live life on our own terms, even if a committee could do a better job of it ([Crisp, 1997](#); [Griffin, 1986](#); [Wall, 1998](#)). I have a very strong intuition that these philosophers are right. For that reason, I think that we all ought to care about resisting temptation. However, anyone who does not share my intuition should still grant that agents who have an interest in self-governance ought to care about resisting temptation.

I have argued that we must have a way of resisting temptation if we want to be self-governing. In what ways can we resist temptation? According to [Sartre \(1956\)](#), in order for the gambler to resist the temptation to gamble, he must "recreate" his thoughts of financial ruin and of disappointing his family as "experienced fear" (p. 69). This suggests that knowing the reasons that led us to form a resolution isn't enough for the resolution to be effective; we must also feel the *force* of those reasons. But that can't be right. Although feeling the force of the reasons that led us to form a resolution might make it easier to be *motivated* to follow through on the resolution, there are plenty of ways we see our resolutions through even when the force of their underlying reasons are inaccessible to us.<sup>32</sup> For instance, we can introduce external pressures to help us stick to our desired course

---

<sup>32</sup> George [Ainslie \(2001\)](#) elaborates on these methods and others for overcoming temptation.



of action. A professor of mine once told me that she would buy herself a pair of earrings and leave them wrapped on the mantle, not to be opened until she finished the paper she was working on; the promise of a new pair of earrings helped her avoid temptation to deviate from her writing schedule. Consider, too, the story of Odysseus, who respected his resolution to resist the bewitching song of the Sirens only by tying himself to the mast of his ship and having his crew plug their ears with beeswax. Another way we get ourselves to act on reasons whose force is inaccessible to us is by diverting our attention away from sources of temptation, as someone seeking to curb a gambling addiction might avoid the game table altogether.

As effective as these methods for resisting temptation are, we have another that works *directly* to fight off contrary inclinations: the ability to form resolutions.<sup>33</sup> Resolutions aren't always enough to get us to stick to a course of action, but as [Holton \(2009, p. 10\)](#) points out, very often they are. People frequently wake up on cold, dark mornings to go for runs and kick bad habits using no other mechanism than a resolution to overcome their contrary desires. Resolutions are important; they are a power we have to execute our plans in the face of temptation.

### 3 Resolutions and Promissory Reasons

In order to be self-governing, we must act in accordance with our commitments, even when we are tempted to do otherwise. I have argued that forming resolutions is one way we can get ourselves to stick to our plans in the face of temptation. My interest in what follows is not in the descriptive question of *how* resolutions do this, or what psychological processes make resolutions effective. Rather, I am interested in a normative question. Given that resolutions work, why is it rational to act on them even when, faced with temptation, our judgments about what we have most reason to do all things considered change? On my view, resolutions are reason-giving: when an agent resolves to  $\phi$ , she incurs an additional normative reason to  $\phi$ . The reasons that resolutions give us can "tip the deliberative scales."

---

<sup>33</sup> Compare this to Bratman's (1987) idea that intentions are "*conduct-controlling* pro-attitudes" that we are "disposed to retain without reconsideration" (p. 20, emphasis added). There is a sense in which resolutions directly control our conduct to shield our plans from temptation in a way that the methods of resisting temptation discussed previously do not.

When this happens, but we nevertheless change our minds about what we have all-things-considered reason to do, we fail to be responsive to our reasons for action. This is why, on my view, it can be irrational to fail to follow through on our resolutions even when we change our minds about what we ought to do.

Where do the reasons we incur from forming resolutions come from? In this section, I argue that the reasons we incur from forming resolutions are similar in important ways to the reasons we incur from making *promises*.

\*\*\*

A person promises to  $\phi$  when she communicates an intention to be under an obligation to  $\phi$ . Although there is disagreement over the source of the normativity of promises—some think that promises are binding because they generate an expectation that the promisor will perform the promised act (Scanlon, 1990); others attribute the normativity of promises to the interests we have in forming intimate relations with others (Shiffrin, 2008), upholding our end the contract we enter into in virtue of being members of a society (Hobbes, 1894), or in sometimes having authority over how others act (Owens, 2006)—there is agreement that promises *bind*. When you break a promise to someone, you wrong them.

Resolutions appear to be similar to self-promises in important ways. Compare the statement "I resolve to watch less TV" to the statement "I promise myself I will watch less TV." The invocation of a promise in the latter statement may make it appear as though there is something *moral* rather than merely prudential on the line. However, in what follows, I hope to show that there are no significant differences between the statements.

Historically, philosophers have been dubious of promises we make to ourselves. In particular, they have doubted that self-promises have the normativity or binding force that interpersonal promises obviously have. In an interpersonal promise, a promisor (the party bound by the promise) can only be released from her obligation to fulfill the promise by the promisee (the party to whom the promisor is bound). However, because in a self-promise the promisor and promisee are one and the same, it appears as though the promisor can release herself from an obligation to fulfill a promise *at will*. Therefore, it may seem as though the promisor was never bound by the promise at all. Thomas Hobbes

(1894) articulates this objection in *Leviathan*, when he discusses why it is impossible for a sovereign to be bound by laws when the laws are up to him:

Having power to make and repeale Lawes, he may when he pleaseth, free himself from that subjection, by repealing those Lawes that trouble him, and making of new; and consequently he was free before. For he is free, that can be free when he will: Nor is it possible for any person to be bound to himself; because he that can bind, can release; and therefore he that is bound to himself onely, is not bound (p. 124).

Call this the *objection from self-release*.

Despite historical doubts about the normativity of self-promises, in recent years, the philosophical consensus has shifted; there is now widespread agreement that self-promises and duties to oneself more generally are possible despite the possibility of self-release (e.g., [Dannenberg, 2015](#); [Fruh, 2014](#); [Muñoz, 2020](#); [Oakley, 2017](#); [Schaab, 2021](#); [Schofield, 2021](#)). Connie Rosati (2011) offers a representative response to the objection from self-release. Rosati argues that self-promises are possible because we can meaningfully distinguish between *releasing* oneself from a self-promise and *breaching* a self-promise. A promisee releases a promisor from her obligation to fulfill a promise when the promisee recognizes herself as having communicated a *genuine change of mind* to the promisor about the promise. (Note that the reasons for the promisee's change of mind need not be *good* reasons; we release people from promises all the time for inconsequential reasons, and on the basis of reasons that we even come to regret.) According to Rosati, there's no reason to think that promissory release operates any differently when it comes to self-promises. There is an important difference between genuinely changing your mind about a promise you made to yourself and acting against that promise. When you fail to fulfill a promise to yourself, you might recognize that you are compromising your values, letting yourself down, or acting self-destructively, even though it was in your power to release yourself from the promise.

There are striking similarities between resolutions and promises that explain why we use resolutions and self-promises interchangeably. Consider first the similarities between the *functions* of promises and of resolutions. Promises play an important role in our social lives; they allow us to form *expectations* about others' future conduct and, relatedly, provide us with *assurance* that others will

act a certain way. If I promise I'll pick you up from the airport, then you can reasonably expect that I'll be there when your flight lands and plan the rest of your day accordingly.<sup>34</sup> Analogously, an agent forms a resolution when she wants assurance that she will follow a course of action, and her resolution generates an expectation about her future conduct. Suppose, knowing that you have writing to do tonight, you are trying to decide whether to watch any TV at all because you suspect that you'll be tempted to watch more than one episode. In resolving to restrict yourself to one episode, you assure yourself that you will watch only one episode, and that your decision to watch any TV at all is sensible. Moreover, resolving to watch only one episode generates an expectation that you will watch only one episode. Having made the resolution, you can reasonably plan ahead for nighttime writing, perhaps by planning to spend less time writing the next day. If you hadn't resolved to limit your TV consumption, it might seem foolish or overly optimistic to expect yourself to get any writing done in the evening. It might, of course, be foolish and overly optimistic for you to expect yourself to get any writing done *even if* you resolve to limit yourself to one episode of TV. If you have a track record of failing to follow through on your resolutions, you might not have reason to trust yourself to follow through on this one. But the same thing is true of promises. If I promise I'll pick you up from the airport, but I have a bad track record of keeping my promises, it might be foolish and overly optimistic of you to form any expectation that I will actually pick you up when your flight lands.

Because they create expectations and assurances about how others will act, promises facilitate *interpersonal coordination* and allow us to reap the benefits of social cooperation. Promises give us a basis for trusting others. Analogously, the expectations and assurances generated by resolutions facilitate *intrapersonal coordination* and allow us to enjoy the benefits of planning agency.

One way to illustrate this similarity between promises and resolutions is to consider them as solutions to prisoner's dilemmas. In the standard prisoners' dilemma (illustrated in the payoff matrix below), two players have the option of

---

<sup>34</sup> As I explain later in this section, some philosophers have argued that the expectations and assurances generated by promises aren't merely useful functions of promises; they ground the *normativity* of promises. But even if you disagree with the claim that expectations and assurances ground the normativity of promises, it's hard to deny that promises play the role of facilitating interpersonal coordination.

either defecting against or cooperating with their opponent. Each player does better by defecting no matter what their opponent does; defecting is therefore a dominant strategy. But if both players cooperate, they each do better than they would have if they both defected; cooperation therefore yields a Pareto optimal outcome.

	Cooperate	Defect
Cooperate	3, 3	0, 5
Defect	5, 0	1, 1

### The prisoner's dilemma

The challenge of the prisoner's dilemma is to figure out how to get players to cooperate and achieve the Pareto optimal outcome despite the fact that defecting is a dominant strategy. A potential solution to this challenge is for the players to *promise* to cooperate, to assure each other that they intend to band together in order to secure the best overall outcome. There is, of course, a question of how such a promise could be credible, especially in a one-off game. If the players don't know each other before the game and won't play against each other again in the future, then they appear to have little reason to follow through on any promise to cooperate that they make. However, most real-life prisoner's dilemmas are *iterated* versions of the game. Because we can expect to encounter our opponents again and again, we have incentive to follow through on our promises to cooperate so our opponents will continue to trust us in the future. Robert [Axelrod \(1980\)](#) has shown that winning strategies in iterated prisoner's dilemma tournaments are *trustworthy*: they earn the trust of their opponents and do not betray that trust by defecting for their own advantage. But even in one-off prisoner's dilemmas, there is (I hope) *moral* incentive to follow through on any promises we make to cooperate. In making a promise to cooperate, an agent who cares about morality *changes the payoffs in the game*. Since defecting would involve breaking a promise, the cost of defecting increases.

In addition to prisoner's dilemmas between two parties, we can recognize *intrapersonal* prisoner's dilemmas, where the two players are versions of oneself at different points in time and with different motivational states. George [Ainslie \(2013\)](#) has suggested that recovering from addiction can be understood as an intertemporal version of an iterated prisoner's dilemma. The present-addict faces

a choice between relapsing or abstaining. If he relapses now, he fulfills his present desire for a high, but imposes a cost on his future self, who must start from scratch tomorrow. If he abstains now, but relapses tomorrow, then he might as well have chosen to relapse now, so his present self incurs a cost. Relapsing therefore appears to be a dominant strategy. However, if the present- and future-addict both abstain, they attain a Pareto optimal outcome: recovery. Moreover, because the addict must engage *repeatedly* in this intertemporal bargaining, every current defection diminishes her credibility and jeopardizes cooperation in the future.

Just as promises provide a potential solution to the interpersonal prisoner's dilemma, *resolutions* provide a potential solution to the intrapersonal prisoner's dilemma. When we respect our resolutions, we build up credibility that we will respect our resolutions in the future, giving our future selves reason to engage in intertemporal bargaining with our present selves. Just as there are benefits to cooperating with others, there are benefits cooperating with *ourselves* by respecting our resolutions. When we cooperates with ourselves, it's easier for us to attain outcomes that are better for ourselves overall, even if they might not appear to be best for us at the moment of action due to the preference and judgment shifts caused by temptation.

I have argued that promises play the role in our interpersonal lives that resolutions play in our intrapersonal lives. I now discuss a second striking similarity between promises and resolutions: the same considerations that can be used to ground the normativity of promises can be used to ground the normativity of resolutions *for the same reasons*. My aim is not to take a stand on which account of the normativity of promises and resolutions is correct, nor to survey every view of the normativity of promises in the literature and show how they also explain the normativity of resolutions. Rather, my aim is to examine a sample of popular theories of promising—Thomas Scanlon's expectation theory and David Owens' authority interest theory—and to show that if one of these successfully grounds the normativity of promises, it also grounds the normativity of resolutions.

I previously showed that promises generate expectations about how others will act. For some philosophers, the value of the expectations created by promises explains why promises bind. [Scanlon \(1990\)](#), for instance, argues that negligently causing someone to expect that you will  $\phi$  and then failing to  $\phi$  amounts to deception and is therefore morally wrong. There is also a case to be made that the

value of the expectations that resolutions generate explains why resolutions bind. An agent who negligently causes herself to expect that she will  $\phi$  and then fails to  $\phi$  engages in *self-deception*. Although the wrong of self-deception may not be the same as that of the deception of others (c.f. Baron, 1988), it is hard to deny that there is something objectionable about self-deception, especially if an agent seeks to be self-governing. Like promises, the normativity of resolutions may be grounded in the wrongness of deception, whatever that wrong amounts to.

On another view of promising advanced by David Owens (2006), promises derive their normativity from our *authority interest*: our desire, in some situations, to have authority over what others do. To understand authority interest, it's helpful to think about the difference between communicating an intention to  $\phi$  and promising to  $\phi$ . Suppose I tell you that I intend to pick you up from the airport, but I don't promise that I will. According to Owens, I refrain from making a promise because I don't want to be *bound* to that course of action. If something else comes up, I want to be free to alter my plans without wronging you. I want, in other words, to retain authority over my actions. If had I promised you I would pick you up from the airport, I would have ceded my authority to you, granting you the power to require me to give you a ride. There are many reasons why I might want in some circumstances to grant you this authority, deriving from your interests and needs as well as my own. Promises exist and derive their power, Owens argues, to serve our authority interest.

It's very plausible that *resolutions* exist and derive their power to serve our authority interest *over ourselves*. If resolutions allow us to overcome the threat of temptation to our self-governance, then they clearly serve our authority interest over ourselves. (Here I diverge slightly from Owens' analysis. In his discussion of authority interest, Owens asks us to consider an akratic agent who judges that she ought to give up smoking but knows perfectly well that she is unlikely to exercise self-control (p. 70). Owens argues that she might nevertheless have an interest in retaining her *right* to exercise self-control. She might, for instance, reasonably refuse to cede her decision-making authority to someone else, even if she would be better off doing so. I agree with Owens that the akratic agent may reasonably wish to retain her right to decide what she shall do even if she knows that she is unlikely to act in accordance with her better judgment. But I think that when an agent is *unable* to resist temptation, she fails to exercise authority

over her actions even if she retains her right to exercise self-control. Our authority interest is closely tied to our ability to resist, and therefore to our ability to form and respect resolutions.)

I have argued that resolutions play the same role in our intrapersonal lives that promises play in our interpersonal lives, and that the normativity of resolutions can plausibly be explained in the same way as the normativity of promises. This should lead us to think that resolutions exert their normativity in the same way promises do: one incurs a reason to  $\phi$  by resolving to  $\phi$  in the same way that one incurs a reason to  $\psi$  by promising to  $\psi$ .

What kinds of reasons does a person incur when she makes a promise? First, she incurs a reason to perform the promised act; if I promise to pick you up from the airport, I have a reason to be at the airport when your flight lands. Second, if she fails to perform the promised act, she incurs a reason to mitigate the effects of that failure; if for some reason I can't make it to the airport, I have reason to apologize to you, and depending on why I failed to fulfill my promise, I might also have reason to help you find another ride home. Analogously, when a person forms a resolution, she incurs a reason to perform the resolved act; if you resolve to limit yourself to one episode of TV tonight, you have reason to watch no more than one episode. And if a person fails to perform the resolved act, she incurs a reason to mitigate the effects of that failure; if you fail to fulfill your resolution and you end up watching more than one episode, making it impossible for you to get writing done in the evening, you incur a reason to mitigate the effects of your actions by, say, doing extra writing tomorrow.

If resolutions give us reasons for action, then our ability to form them is a *normative power*: a power, that is, we have over our normative circumstances. Other phenomena regarded as normative powers include promising, consent, and forgiveness. When we make promises, we give ourselves reasons to perform the promised acts. When we give consent, we waive certain rights we have, making it permissible for others to perform acts that would otherwise violate our rights. When we forgive, we relieve others of their debts to us. Alongside these phenomena, resolutions are a power we have to create reasons for action—to change the normative landscape—through sheer acts of will.

In arguing that resolutions are reason-giving, my point is not that we ought to be resolute *no matter what*. When you resolve to limit yourself to one episode



of TV per evening except on days that are truly different from other days (days, perhaps, that have been especially stressful, or where you have less work to get done than usual), there's nothing wrong with watching more than one episode on days that truly are different from other days. Without the capacity to resist temptation, we would be slaves to our whims, unable to unify our agency and execute plans. But denying ourselves *any* flexibility in the pursuit of our goals isn't ideal, either.

#### 4 The Bootstrapping Problem

Broadly speaking, there are two ways of understanding the normativity of resolutions. On one approach, our capacity to make resolutions is a power we have to create normative reasons by our willing. Call this the *reasons view*. On a second approach, we are rationally required to follow through on our resolutions as a matter of means-end coherence. On this approach, although we have reasons to become the kinds of people who are disposed to follow through on resolutions, resolutions themselves do not give us reasons. Call this the *requirements view*.

There are important differences between the reasons and requirements views, between thinking that we have *reasons* to follow through on our resolutions and that we are *rationally required* to follow through on our resolutions. Reasons have weight in practical reasoning, but are slack. They can be overruled by other, weightier reasons, but they always stay on the scene. Here is an example to illustrate how the two views work. If resolutions are reason-giving, then when you resolve to limit yourself to one episode of TV so you can get writing done in the evening, you incur a reason to limit yourself to one episode of TV so you can get writing done in the evening. But new information might come to light that overrides your previously all-things-considered reason to watch only one episode. If a friend calls you and asks you to binge-watch a TV show with her to take her mind off a recent breakup, you incur a reason to watch more than one episode of TV that overrides the reason you incurred from your resolution. By contrast, requirements have no weight, and are strict; you either satisfy them or you don't, and they cannot be outweighed by other requirements. If we are required to follow through on resolutions as a matter of means-end coherence, then if you resolve to limit yourself to one episode of TV so you can get writing

done in the evening, you must adopt the means to achieve that resolution and refrain from watching a second episode. But if your recently single friend calls you and asks to binge watch TV, you can abandon your resolution and adopt a new end of watching more than one episode, for which you are now rationally required to adopt the means.

The view I have offered in this paper is a *reasons view*: I have argued that resolutions are reason-giving. However, philosophers tend to be skeptical of the claim that mental attitudes (including intentions and resolutions) give us reasons because of concerns about *bootstrapping*. If mental attitudes gave us reasons, the worry goes, then we could bootstrap any action into rationality simply by acquiring the relevant mental attitude. John Broome (2001) summarizes the bootstrapping problem in the following passage:

The view that intentions [and therefore resolutions] are reasons is implausible. If you have no reason to do something, it is implausible that you can give yourself a reason just by forming the intention of doing it. How could you create a reason for yourself out of nothing? Suppose, say, that you have no reason either for or against doing some act, and you happen to decide to do it. So now, if intentions are reasons, you have a reason to do it. Since you have no contrary reason not to do it, the balance of reasons is in favour of your doing it. But [...] [i]t is implausible that just deciding to do something can make it the case that you ought to do it, when previously that was not the case (p. 87).

There are two kinds of cases where resolutions appear to lead to problematic bootstrapping. The first is *Buridan's donkey cases*, cases where a person has the exact same reasons to pursue multiple courses of action. Suppose a donkey finds himself halfway between two bales of hay. He has no reason to prefer one to the other. But he needs to eat, so he decides to head toward the one on his right. Suppose next that a gust of wind blows the bale of hay on his right away, and he is now closer to the bale of hay on his left. If his resolution to the bale of hay to his right gave him a normative reason to head toward it, then it appears as though he still ought to walk toward it even though he would be better off pursuing the one to his left. He has problematically bootstrapped his pursuit of the bale of hay to the right into rationality. The second kind of case where resolutions appear to lead to problematic bootstrapping involves resolutions to perform immoral or self-destructive actions. For instance, if morality demands requires that we give to

charity, it seems implausible that we could make it rational *not* to give to charity simply by resolving and thereby giving ourselves a reason to do so.

The requirements view easily sidesteps the bootstrapping challenge. To see why this is the case, we can draw a distinction between narrow and wide scope versions of means-end coherence:

NARROW SCOPE: If you adopt the ends, then rationality requires of you that you adopt the necessary means.

WIDE SCOPE: Rationality requires of you that if you adopt the ends, then you adopt the necessary means.

Because means-end coherence is taken to have a *wide scope*, you can choose to fulfill the requirement either by adopting the necessary means *or by giving up the end*. No bootstrapping occurs; if you abandon an end and adopt a new one, you are no longer rationally required to take the means to your previous end.

If I am right that the reasons resolutions give us are akin to promissory reasons, then the reasons view also has a response to the bootstrapping objection; we can respond to these objections by thinking about why promises are not susceptible to them. After all, on all accounts, promises give rise to reasons: when you promise to  $\phi$ , you incur a normative reason to  $\phi$ . Therefore, if promises don't bootstrap problematically, resolutions shouldn't either.

Consider, first, Buridan's donkey cases. Does the donkey's resolution to pursue the bale of hay on the right give him reason to continue pursuing it even if it blows away? We can explain why it does not by thinking about how changing circumstances render promises (and the reasons they generate) irrelevant. Suppose I promise to pick you up from the airport when your plane lands on Tuesday. I now have a reason to drive to the airport on Tuesday. But if your flight is cancelled or you decide to take a cab home instead, I no longer have a reason to go to the airport on Tuesday; the changing circumstances have rendered my promise irrelevant. Similarly, even if resolutions are reason-giving, changing circumstances can render the reasons our resolutions give us irrelevant. Buridan's donkey resolves to walk toward the bale of hay to his right, but if that bale of hay is blown away, his resolution becomes irrelevant. He now has most reason to walk toward the bale of hay to his left, since it's closer.

The power of resolutions to tip the scales in favor of one course of action

when we would otherwise have no reason to prefer one over the other is especially important when we make what Ruth Chang (2017) calls *hard choices*. In a hard choice, an agent must decide between alternatives where every alternative is better in some relevant respects, yet none of them is better than the other overall (p. 1). Choices between careers are often hard in this sense; two careers might be in the "same neighborhood of 'goodness as a career for you,'" but manifest very different qualities of what makes a career good for you (p. 15). Unlike Buridan's donkey cases, in a hard choice, an agent has very different reasons to pursue one course of action over another. But similar to Buridan's donkey cases, in hard choices, an agent's reasons for pursuing one course of action over another "run out." According to Chang, when our reasons run out when we make hard choices, we must rely on our normative power to create *will-based reasons* for ourselves to choose one alternative over the other—resolutions provide the will-based reasons Chang is looking for. In hard choices, a resolution may bootstrap, but it does not do so problematically.

Let's now turn to the second kind of case where resolutions appear to lead to problematic bootstrapping: cases involving immoral or self-destructive resolutions. Do immoral *promises* bind? If I promise someone I will rob a store, do I now have an obligation to rob the store, despite the demands of morality? The answer to this question is controversial. Joseph Raz (1977) argues that immoral promises *do* bind. For him, promises give rise to obligations, and fulfilling obligations will sometimes require agents to perform acts that should not be performed on the balance of reasons (p. 224). Therefore, a person who makes an immoral promise will always end up doing something wrong: either she will fail to fulfill her obligation to perform the promised act, or she will perform the immoral act that she promised to. Holly Smith (1997) disagrees. On her view, it would be troubling if the power to promise allowed us to dictate the moral status of actions, that we could give ourselves any reason at all to perform actions that are immoral. If resolutions give us reasons that are akin to promissory reasons, then we can respond to concerns about immoral and self-destructive bootstrapping by appealing to either of these arguments. We can follow Raz and argue that resolutions of this kind bootstrap, but not problematically, or we can follow Smith and argue that these resolutions do not bootstrap at all.

I suspect that we can appeal to a comparison between resolutions and

promises to respond to other scenarios where resolutions-based reasons appear to lead to unacceptable bootstrapping. That promises are not susceptible to problematic bootstrapping suggests that resolutions shouldn't be, either.

## References

- Ainslie, G. (2001). *Breakdown of will*. Cambridge: Cambridge University Press.
- Ainslie, G. (2013). Intertemporal bargaining in addiction. *Front Psychiatry, 4*, 63.
- Ainslie, G., & Haslam, N. (1992). Hyperbolic discounting. In G. Loewenstein & J. Elster (Eds.), *Choice over time* (pp. 57–92). New York: Russell Sage Foundation.
- Archer, A. (2017). Integrity and the value of an integrated self. *Journal of Value Inquiry, 51*(3), 435–454.
- Arpaly, N., & Schroeder, T. (1999). Praise, blame and the whole self. *Philosophical Studies, 93*(2), 161–188.
- Axelrod, R. (1980). Effective choice in the prisoner's dilemma. *Journal of Conflict Resolution, 24*(1), 3–25.
- Baron, M. (1988). What's wrong with self-deception? In B. P. McLaughlin & A. O. Rorty (Eds.), *Perspectives on self-deception* (pp. 431–449). Berkeley: University of California Press.
- Bergamin, J. (2017). Being-in-the-flow: Expert coping as beyond both thought and automaticity. *Phenomenology and the Cognitive Sciences, 16*, 403–424.
- Bermúdez, J. P. (2017). Do we reflect while performing skillful actions? automaticity, control, and the perils of distraction. *Philosophical Psychology, 30*(7), 896–924.
- Bok, H. (2008). Acting without choosing. *Noûs, 30*(2), 174–196.
- Borkovec, T., & Sharpless, B. (2004). Generalized anxiety disorder: Bringing cognitive-behavioral therapy in the valued present. In S. C. Hayes, V. M. Follette, & M. M. Linehan (Eds.), *Mindfulness and acceptance: Expanding the cognitive-behavioral tradition* (pp. 209–242). New York: The Guilford Press.
- Botton, S. (2012). Conversations with writers braver than me: Cheryl strayed. *The Rumpus*. Retrieved from <https://therumpus.net/2012/07/09/conversations-with-writers-braver-than-me-13-cheryl-strayed/>.
- Bratman, M. E. (1983). Taking plans seriously. *Social Theory and Practice, 9*(2-3), 271–287.
- Bratman, M. E. (1987). *Intentions, plans, and practical reason*. Cambridge: Cambridge University Press.
- Bratman, M. E. (2000). Reflection, planning, and temporally extended agency. *The Philosophical Review, 109*(1), 35–61.
- Bratman, M. E. (2007). *Structures of agency: Essays*. New York: Oxford University Press.

- Brison, S. (1996). Outliving oneself: Trauma, memory, and personal identity. In D. T. Meyers (Ed.), *Feminists rethink the self*. Boulder: Westview Press.
- Broome, J. (2001). Are intentions reasons? and how should we cope with incommensurable values? In C. W. Morris & A. Ripstein (Eds.), *Practical rationality and preference: Essays for David Gauthier* (pp. 98–120). New York: Cambridge University Press.
- Browning, J. C. (2000). Shiloh witness. In *Deep in our hearts* (pp. 37–84). Athens: University of Georgia Press.
- Buss, S. (2006). Needs (someone else's), projects (my own), and reasons. *The Journal of Philosophy*, 103(8), 373–402.
- Buss, S. (2012). Autonomous action: Self-determination in the passive mode. *Ethics*, 122(4), 647–691.
- Buss, S. (2020). Some musings about the limits of an ethics that can be applied — a response to a question about courage and convictions that confronted the author when she woke up on November 9, 2016. *Journal of Applied Philosophy*, 37(1), 1–33.
- Buss, S. (2021). *What's so great about being a self-governing agent?* Unpublished manuscript.
- Calhoun, C. (2015). Standing for something. In *Moral aims: Essays on the importance of getting it right and practicing morality with others* (pp. 123–153). New York: Oxford University Press.
- Callard, A. (2018). *Aspiration: The agency of becoming*. New York: Oxford University Press.
- Chan, W. T. (1955). The evolution of the Confucian concept *Jên*. *Philosophy East and West*, 4(4), 295–319.
- Chang, R. (2017). Hard choices. *Journal of the American Philosophical Association*, 3(1), 1–21.
- Chin, J., & Vasarhelyi, E. C. (2018). *Free solo*. National Geographic Documentary Films.
- Confucius. (2019). The analects. In D. Sturgeon (Ed.), *Chinese text project: A dynamic digital library of premodern Chinese*. Retrieved from <https://ctext.org/analects/yong-ye/zh?en=on>.
- Cottingham, J. (2010). Integrity and fragmentation. *Journal of Applied Philosophy*, 27(1), 2–14.
- Crisp, R. (1997). *Mill on utilitarianism*. New York: Routledge.
- Dannenberg, J. (2015). Promising ourselves, promising others. *The Journal of Ethics*, 19(2), 159–183.
- Deraniyagala, S. (2013). *Wave*. New York: Alfred A. Knopf.
- Dostoyevsky, F. (2003). *Crime and punishment* (D. McDuff, Trans.). London: Penguin.
- Feinberg, J. (1986). *Harm to self: The moral limits of the criminal law*. New York: Oxford University Press.

- Ferrero, L. (2010). Decisions, diachronic autonomy & the division of deliberative labor. *Philosophers' Imprint*, 10(2), 1–23.
- Foot, P. (2002). *Virtues and vices and other essays in moral philosophy*. Oxford: Oxford University Press.
- Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68(1), 5–20.
- Frankfurt, H. G. (1988a). Identification and wholeheartedness. In *The importance of what we care about* (pp. 159–176). Cambridge: Cambridge University Press.
- Frankfurt, H. G. (1988b). Rationality and the unthinkable. In *The importance of what we care about*. Cambridge: Cambridge University Press.
- Frankfurt, H. G. (1999a). Autonomy, necessity, and love. In *Necessity, volition, and love* (pp. 129–141). Cambridge: Cambridge University Press.
- Frankfurt, H. G. (1999b). On the necessity of ideals. In *Necessity, volition, and love* (pp. 108–116). Cambridge: Cambridge University Press.
- Frankfurt, H. G. (2002). Reply to J. David Velleman. In S. Buss & L. Overton (Eds.), *Contours of agency: Essays on themes from Harry Frankfurt* (pp. 124–128). Cambridge: MIT Press.
- Fruh, K. (2014). The power to promise oneself. *The Southern Journal of Philosophy*, 52(1), 61–85.
- Ganeri, J. (2021). *Virtual subjects, fugitive selves: Fernando Pessoa and his philosophy*. New York: Oxford University Press.
- Gilbert, E. (2006). *Eat, pray, love: One woman's search for everything across Italy, India, and Indonesia*. New York: Viking.
- Griffin, J. (1986). *Well-being: Its meaning, measurement and moral importance*. Oxford: Clarendon.
- Hammett, D. (1930). *The maltese falcon*. New York: Alfred A. Knopf.
- Hattrup, K. N. (2018, jul). *Sainthood process begins for young iraqi priest killed in 2007*. Retrieved from <https://aleteia.org/2018/07/05/sainthood-process-begins-for-young-iraqi-priest-killed-in-2007/>.
- Hengel, E. (2017). Publishing while female: Women held to higher standards? evidence from peer review. *Cambridge Working Paper Economics*.
- Hobbes, T. (1894). *Leviathan*. London: Routledge.
- Holton, R. (2009). *Willing, wanting, waiting*. New York: Oxford University Press.
- Hornbacher, M. (1997). *Wasted: A memoir of anorexia and bulimia*. New York: Harper Collins.
- Hurka, T. (2006). Games and the good. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 40(1), 217–264.
- Jacobs, J. (2003). Some tensions between autonomy and self-governance. *Social Philosophy and Policy*, 2, 221–244.
- Kant, I. (2017). *The metaphysics of morals* (L. Denis, Ed. & M. Gregor, Trans.). Cambridge: Cambridge University Press.



- Khader, S. J. (2011). *Adaptive preferences and women's empowerment*. New York: Oxford University Press.
- Kolodny, N. (2013). Introduction. In Samuel Scheffler, *Death and the afterlife* (pp. 3-14). New York: Oxford University Press.
- Korsgaard, C. M. (1996). *The sources of normativity*. Cambridge: Cambridge University Press.
- Korsgaard, C. M. (2009). *Self-constitution: Agency, identity and integrity*. Oxford: Oxford University Press.
- Kyle, B. G. (2017). Courage, cowardice, and Maher's misstep. *Canadian Journal of Philosophy*, 47(4), 565-587.
- Logan, G. D. (2018). Automatic control: How experts act without thinking. *Psychological Review*, 125(4), 453-485.
- Lugones, M. (1987). Playfulness, 'world'-traveling, and loving perception. *Hypatia*, 2(2), 3-19.
- Lugones, M. (1990). Hispaneando y lesbiando: On Sarah Hoagland's lesbian ethics. *Hypatia*, 5(3), 138-146.
- Lugones, M. (1991). On the logic of pluralist feminism. In C. Card (Ed.), *Feminist ethics* (pp. 35-44). Lawrence: University of Kansas Press.
- McFall, L. (1987). Integrity. *Ethics*, 98(1), 5-20.
- Milgram, S. (1974). *Obedience to authority*. New York: Harper & Row.
- Mills, C. (2018). Artistic integrity. *The Journal of Aesthetics and Art Criticism*, 76(1), 9-20.
- Mullin, A. (1995). Selves, diverse and divided: Can feminists have diversity without multiplicity? *Hypatia*, 10(4), 1-31.
- Muñoz, D. (2020). The paradox of duties to oneself. *Australasian Journal of Philosophy*, 98(4), 1-12.
- Narayan, U. (2002). Minds of their own: Choices, autonomy, cultural practices and other women. In L. M. Antony & C. E. Witt (Eds.), *A mind of one's own: Feminist essays on reason and objectivity* (pp. 418-432). Boulder: Westview.
- Nguyen, C. T. (2019). Games and the art of agency. *The Philosophical Review*, 128(4), 423-462.
- Oakley, T. (2017). How to release oneself from an obligation: Good news for duties to oneself. *Australasian Journal of Philosophy*, 95(1), 70-80.
- Orwell, G. (1949). 1984. New York: Harcourt.
- Oshana, M. (2006). *Personal autonomy in society*. Aldershot: Ashgate Publishing.
- Owens, D. (2006). A simple theory of promising. *Philosophical Review*, 115(1), 51-77.
- Parfit, D. (1984). *Reasons and persons*. New York: Oxford University Press.
- Paul, L. A. (2014). *Transformative experience*. Oxford: Oxford University Press.
- Pianalto, M. (2012). Integrity and struggle. *Philosophia*, 40(2), 319-336.
- Plato. (1901). *The apology of socrates* (D. F. Nevill, Trans.). London: F. E. Robinson & Co.

- Rachels, J. (1999). *The elements of moral philosophy*. Boston: McGraw-Hill.
- Rawls, J. (1982). Social unity and primary goods. In A. Sen & B. Williams (Eds.), *Utilitarianism and beyond* (pp. 159–185). Cambridge: Cambridge University Press.
- Raz, J. (1977). Promises and obligations. In P. Hacker & J. Raz (Eds.), *Law, morality, and society: Essays in honour of h.l.a. hart*. Oxford: Clarendon Press.
- Roberts, J. L. (1996). *The importance of oskar schindler*. San Diego: Lucent Books.
- Rosati, C. S. (2011). The importance of self-promises. In H. Scheinman (Ed.), *Understanding promises and agreements* (pp. 124–155). New York: Oxford University Press.
- Sartre, J.-P. (1956). *Being and nothingness* (H. E. Barnes, Trans.). New York: Washington Square Press.
- Sartre, J.-P. (2007). *Existentialism is a humanism* (J. Kukla, Ed. & C. Macomber, Trans.). New Haven: Yale University Press.
- Scanlon, T. (1990). Promises and practices. *Philosophy & Public Affairs*, 19(3), 199–226.
- Schaab, J. D. (2021). On the supposed incoherence of obligations to oneself. *Australasian Journal of Philosophy*, 99(1), 175–189.
- Schofield, P. (2021). *Duty to self: Moral, political, and legal self-relation*. Oxford: Oxford University Press.
- Setiya, K. (2017). *Midlife: A philosophical guide*. Princeton: Princeton University Press.
- Shiffrin, S. V. (2008). Promising, intimate relationships, and conventionalism. *Philosophical Review*, 117(4), 481–524.
- Shoemaker, D. W. (2003). Caring, identification, and agency. *Ethics*, 114, 88–118.
- Smith, H. M. (1997). A paradox of promising. *The Philosophical Review*, 106(2), 153–196.
- Sripada, C. (2016). Self-expression: a deep self theory of moral responsibility. *Philosophical Studies*, 173(5), 1203–1232.
- Stark, C. (1860). *Memoir and official correspondences of Gen. John Stark*. Concord: G. Parker Lyon.
- Stoljar, N. (2018). Feminist perspectives on autonomy. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*.
- Strayed, C. (2012). *Wild: From lost to found on the Pacific Crest Trail*. New York: Alfred A. Knopf.
- Styron, W. (1979). *Sophie's choice*. New York: Random House.
- Taylor, G. (1981). Integrity. *Aristotelian Society Supplementary Volume*, 55(1), 143–159.
- Valdman, M. (2010). Outsourcing self-government. *Ethics*, 120, 761–790.
- Velleman, J. D. (1992). What happens when someone acts? *Mind*, 101(403), 461–481.
- Velleman, J. D. (2002). Identification and identity. In S. Buss & L. Overton (Eds.),

- Contours of agency: Essays on themes from Harry Frankfurt* (pp. 91–123). Cambridge: MIT Press.
- Verplanken, B., Roy, D., & Whitmarsh, L. (2018). Cracks in the wall: Habit discontinuities as vehicles for behavior change. In B. Verplanken (Ed.), *The psychology of habit: Theory, mechanisms, change, and contexts* (pp. 189–205). Cham: Springer.
- Wall, S. (1998). *Liberalism, perfectionism and restraint*. New York: Cambridge University Press.
- Watson, G. (1975). Free agency. *The Journal of Philosophy*, 72(2), 205–220.
- Wertheimer, A. (1987). *Coercion*. Princeton: Princeton University Press.
- Williams, B. (1973). The Makropulos case: Reflections on the tedium of immortality. In *Problems of the self: Philosophical papers 1956-1972* (pp. 82–100). Cambridge: Cambridge University Press.
- Williams, B. (1981). Persons, character and morality. In *Moral luck: Philosophical papers 1973-1980* (pp. 1–19). Cambridge: Cambridge University Press.
- Wolf, S. (1982). Moral saints. *The Journal of Philosophy*, 79(8), 419–439.