UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Self-Articulation and Self-Formation

Permalink

https://escholarship.org/uc/item/8bj5726h

ISBN

9798193289562

Author

Tiran, Noam

Publication Date

2026-10-12

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Self-Articulation and Self-Formation

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Philosophy

by

Noam Tiran

Committee in charge:

Professor Matt Fulkerson, Co-Chair
Professor Dana Nelkin, Co-Chair
Professor Christine Harris
Professor Nick Riggle
Professor Clinton Tolley
Professor Manuel Vargas

2026

Copyright

Noam Tiran, 2026

All rights reserved.

The Dissertation of Noam Tiran is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2026

DEDICATION

To Avigail, with love.

TABLE OF CONTENTS

DISSERTATION APPROVAL PAGE iii

DEDICATION iv

TABLE OF CONTENTS v

LIST OF FIGURES viii

ACKNOWLEDGEMENTS ix

VITA xiii

ABSTRACT OF THE DISSERTATION xiv

INTRODUCTION 1

Chapter 1. SELF-ARTICULATION AS SELF FORMATION 6

1. Introduction 6

2. Existential Self-Ascriptions at Work 9

2.1. Existential Self-Ascriptions in Therapy 10

2.2. Existential Self-Ascriptions and Radicalization 13

3. Change and Constraint 17

4. Description and Determination 21

5. Affinities and Differences 28

6. Conclusion 32

Chapter 2. AGENTIAL UNITY, THE SENSE OF IMPROT, AND THE NECESSITY OF GIVING A DAMN 34

1. Introduction 34

2. Two Problems 35

3. Relegation and Reflection 37

3.1. Relegation 38

3.2. Reflection 41

4. The Sense of Import 43

5. Synchronic Integration....47
6. Diachronic Stability and Caring....49
7. Caring and the Sense of Import....55
8. Conclusion....58
Chapter 3. THE IMPISHNESS OF AFFECT....60
1. Introduction....60
2. The Interpretation-Regulation Connection....64
3. The Knowledge Account....67
4. Theoretical and Deliberative Modes of Awareness....72
5. The Sense of Import....75
6. From Emotions to Commitments....80
7. From Commitments to Regulation....82
8. Conclusion....87
Chapter 4. ENGAGEMENT, RESONANCE, AND THE AESTHETICS OF ETHOS....89
1. Introduction....89
2. Authenticity....91
3. Ethos....96
4. Connection....100
5. Games....103
6. Rides....110
7. Woods and Gardens....114
8. Is This Some Kind of Joke?....117
Chapter 5. THE ETHICS (AND AESTHETICS) OF UPTAKE REFUSALS....121
1. Introduction....121
2. Uptake: Some Preliminaries....122

3. Two Kinds of Uptake Refusal....124
4. Covert Exercitives....126
5. The Dangers of Uptake Refusals....129
6. Evaluative Underdetermination....132
7. Dialogical Self-Interpretation....135
8. Refusals Done Right....139
9. Engagement and Resonance....147
9.1. Engagement....147
9.2. Resonance....151
10. Conclusion....153
EPILOGUE....155
REFERENCES....157

LIST OF FIGURES

Figure 1. Cares are referentially unified.. 77
Figure 2. The sense of import is referentially disunified. 78
Figure 3. The sense of import is multiply interpretable. 80

ACKNOWLEDGEMENTS

It takes a village to write a dissertation. Luckily, the one I inhabited in the past five years has been particularly hospitable, welcoming, and generous. Any intellectual value that may be attributed to this project is a reflection (perhaps a dim reflection) of the value I gained from the communities to which I belonged during my graduate studies.

My family is a never-ending source of strength and inspiration. My mother, Michal Dekel, is an exemplar of care and resilience. Everything I do, I do with an eye toward making her proud. My grandparents, Yael and Noach Barski, are paragons of persistence, dignity, and resolve. I can say with the utmost confidence that they don't make 'em like that anymore. My siblings, Mai Dekel, Aviv Dekel, Lanuel Tiran-Tishbi, and Jonathan Tiran-Tishbi, demonstrate the wildly different ways in which humans can be talented, kind, and just plain amazing. I'm honored to try to keep up with them. My three non-biological brothers, Adam Etzion, Itamar Mushkin, and Jonathan Mushkin are the best—and worst (in the best way)—fellow travelers a person can ask for. Together, there's nothing we can't spectacularly fail at.

The dear friends I have made in San Diego have kept me sane—at times quite literally—during graduate school. Marya Dyshel, Avi Libster, Johni Ben-Shabbat, Dafna Mark Ben-Shabbat, Tomer Rousso, Maytal Yifrach, Adi Zuker and Sasha Zuker were the best home away from home I could have asked for. I am also fortunate to have had the precious comradery of my cohort members, Daniel Cohen, Tyler Farmer, Bosco Garcia, Julia Greig, Alice Koga, Faron Ray, and Karina Ortiz; and I am beyond grateful for the support and friendship of my other peers and colleagues in and outside of UC San Diego's philosophy department—Eli Benjamin Israel, Zack Brants, Haggeo Cadenas, Sam Delorme, Rylan Garwood, JC Gonzalez, David Heering, Annie Kantar, Amir Liron, Huadian Liu, Ying Liu, Joe Masotti, Gabe Nyberg, Subin Park, Sam Ridge,

Emiliano Salomón, Andy Sin, Tyler Sproule, Bas Tönissen, and many others—who were my intellectual base camp.

Most people can only dream of the kind of counsel, guidance, and encouragement I have received from my academic mentors during my training at UC San Diego. I am humbled to have Professors Matt Fulkerson and Dana Nelkin as my dissertation co-chairs, and I am lucky that they are as kind and generous as they are terrifyingly smart. Matt's research is a paradigm of engaging, intelligent, empirically-informed scholarship; and Dana has shown me in numerous ways what it means to do systematic, nuanced, and responsible (no pun intended) philosophy. I will forever strive to meet the astonishingly high bar that they set. Furthermore, Matt and Dana were just as eager to help me find my own philosophical voice as they were to uphold me to the highest philosophical standards (I am afraid that on both counts I still have much work to do). I hope to one day be able to approximate for my students the kind of caring mentorship that they have extended to me.

The rest of my committee members are a bona fide dream team. Professor Nick Riggle has provided me with countless examples of what truly original philosophy can be, when done right; of the different ways in which philosophers should attune themselves to both rigor and beauty. The conversations I've had with Nick over coffee were some of the most philosophically exhilarating dialogues I've ever had the pleasure of participating in. Professor Clinton Tolley constantly reminded me of how intellectual breadth and profundity should never exclude each other; and of the extent to which charitable, precise, and careful reading can be just as demanding as writing. I will forever be grateful to him for encouraging (read: enabling) me to foray into intellectual domains that lie beyond the reaches of the surgical light of analytic philosophy. If Professor Tolley is my inspiration for exploring diverse philosophical traditions, Professor Manuel Vargas is my

role model for building bridges between such traditions and analytic philosophy. I have benefitted tremendously from both his intelligence and his humanism. Finally, I was extremely fortunate that Professor Christine Harris had agreed to join me on this weird, wild ride. Her presence on my committee, her erudition and kindness, have kept me honest and responsible with respect to the empirical aspects of this project. Our conversations in and outside the classroom were prime examples of how engaging with scientists can contribute to—rather than detract from—one's enchantment regarding the human condition.

Besides my committee, I am indebted to numerous other teachers, guides, and advisors. A woefully incomplete list includes Yuval Dolev, my first philosophical mentor, who shaped my philosophical orientation in ways that he will never fully realize; Hilla Jacobson and Preston Werner, my MA advisors, who initiated me into this bizarre cult with care and precision; Ayelet Shavit, who taught me what engaged academia can and should look like; Ori Beck, who is the best philosophical big brother one can hope to have; members of San Diego's marvelous intellectual ecosystem—including Reuven Brandt, David Brink, Cami Coepke, Jonathan Cohen, Amelia Glaser, Harvey Goldman, Tina Gremel, Aaron Gross, Deborah Hertz, Patricia Marechal, Kerry McKenzie, Sam Rickless, Gila Sher, and Caroline Wells—and many other wonderful people who provided feedback, comments, and companionship along the journey, such as Arnon Cahen, Ellen Fridland, Keren Gorodeisky, Zac Irving, Chris Kochevar, Amy Levine, Tyler Olsson, Kylie Shahar, Shlomi Sher, and Vida Yao. On an especially reflective note, I'd be remiss not to mention my brief, yet deeply meaningful interaction with Jonathan Lear, right before his untimely passing. I'll always cherish both his work, and our short—*too short*—correspondence.

Finally, I wish to express my deepest gratitude and admiration for the love of my life, Avigail Eitan. I don't know what I would have been without her, and I'm glad I don't have to find out. This dissertation is dedicated to her.

Also thanks to Effie, San Diego's number one dog.

VITA

2018 Bachelor of Arts in Philosophy, Shalem College

2021 Master of Arts in Philosophy, The Hebrew University of Jerusalem

2026 Doctor of Philosophy in Philosophy, University of California San Diego

ABSTRACT OF THE DISSERTATION

Self-Articulation and Self-Formation

by

Noam Tiran

Doctor of Philosophy in Philosophy

University of California San Diego, 2026

Professor Matt Fulkerson, Co-Chair
Professor Dana Nelkin, Co-Chair

This dissertation argues that interpreting our own values, cares, concerns, and other elements of our identities can significantly transform our evaluative, motivational, and affective lives. More specifically, I develop an empirically informed account of how applying lexical concepts to our evaluative-motivational perspectives (paradigmatically, in the context of sincere, intimate dialogue) can reshape those very perspectives, often bringing them into closer alignment with the contents those concepts ascribe to them. In short, I show that self-interpretation is, in the usual case, a mode of self-formation. In fact, I argue, it is a particularly *valuable* mode of changing—and holding on to—who we are.

INTRODUCTION

This dissertation can be read as a peculiar, extended footnote to several comments made by the great 20th century Jewish philosopher, theologian, and sage, Martin Buber. Throughout his writings, Buber attempts to characterize the nature and significance of a unique relation that may be established between oneself and other entities, which he calls "dialogue." While the most salient paradigms of dialogue involve two human persons, Buber is adamant that one may stand in dialogue with works of art, with nature, with one's own will or decisions, and even with machines. The capacity for dialogue is central to our humanity, but it is also, in part, what allows us to come into intimate contact with the more-than-human world.

One of dialogue's most striking features, which Buber persistently insists on, is that it is a relation of *actualization* or *realization* (*verwirklichung*). In dialogue, both *relata* become somehow more *real*. Their identities or essences draw nearer to fruition, bloom further toward being. In the artistic process, for example, "a human being confronts a form [*gestalt*] that wants to become a work through him" (2023: 9). Interestingly, Buber suggest that when confronted with such a form, "I cannot experience nor describe [it], I can only actualizae it" (10). Others, then, force a choice on the self: she must either actualize them, or take advantage of them. We are truly free only when and insofar as we choose the former. A free person, Buber says, "listens to that which grows, to the way of Being in the world, not in order to be carried along by it but rather in order to actualize it in the manner in which it, needing him, wants to be actualized by him" (50). Dialogue, then, is a reciprocal relation into which entities must enter in order to form each other—and themselves—into what they "want" or "need" to become. Famously, this relation stands at the heart of Buber's moral psychology, ethical vision, political agenda, and theological framework.

I am captivated by Buber's remarks. They strike me as both inescapably true and delightfully cryptic. This project is an attempt to spell them out in one way that may be accessible to analytic moral psychologists, philosophers of mind, and ethicists. However, I should note that this attempt is—perhaps fortunately—incomplete. I do not purport to draw conclusions about the ultimate nature of reality, or the divine, or even, for that matter, human personhood. Rather, I wish to show how a Buber-inspired view might have application in a specific, albeit salient domain: the formation and transformation of people's practical and ethical identities. I wish to show how *something like* dialogue—sincerely and vulnerably talking to each other (and, derivatively, also to ourselves)—can form and transform us in deep and pervasive ways.

Each of the chapters in this dissertation is more-or-less a standalone piece, yet together, I hope, they tell a coherent story. **Chapter 1** begins by providing evidence for the thought that acts of self-ascribing values, reasons, cares, and concerns—acts which I will call "existential self-ascriptions" (or ESAs)—are capable of altering their utterer's evaluative-motivational outlook. On the other hand, as I explain, such self-ascriptions also appear to be constrained by the speaker's preexisting emotional realities. The juxtaposition of these two claims yields an air of paradox: How can ESAs both be constrained by and alter who we are? I resolve this puzzle by suggesting that articulating our cares and values can be a matter of making them more determinate. Certain self-ascriptions shape who we are by making our determin*able* affective, motivational, and evaluative dispositions into determin*ate* modes of caring and valuing.

Chapter 2 develops the main theoretical machinery that I utilize to show how our evaluative-motivational lives can be indeterminate in ways that are relevant to the picture proposed in chapter 1. I argue that, at least for creatures like us, *agency itself* requires that we cultivate a set of care-like dispositions, or a "minimal mode of mattering." This minimal mode of mattering is

necessary to initiate action, as well as to sustain and guide it. Furthermore, for our actions to issue from explicit *practical deliberation*, this minimal mode of mattering must be *stable*. It is not sufficient that some things matter to us – that we take some things rather than others to be important in some minimal sense—but that we find *the same things* important across *the same contexts*. Agency in general requires a pattern of mattering, and a distinctively *human* kind of agency requires a reliable, regular pattern of mattering. I call this pattern *the sense of import* (or SoI). I argue that, like our caring and valuing attitudes, the SoI is evaluative and motivating, but unlike these attitudes, it is *referentially disunified*. It is not intentionally directed at one specific object, but at different features of our action contexts. As such, the SoI is *multiply interpretable.* There are different interpretations that we can legitimately ascribe to it.

Chapter 3 uses these insights to show how interpreting our SoIs can change our evaluative outlooks. The chapter is framed as a response to the following question: A large body of empirical research suggests that a person's capacity to interpret, label, or articulate her own affective states in specific, fine-grained ways is closely connected to her ability to regulate them effectively. What might explain this connection between emotional interpretation and regulation? I argue that the connection is best explained by the claim that many of our affective states and processes—primarily the states and processes that comprise our SoIs—are *impish*: acts of voluntarily interpreting, labeling, or articulating them can directly transform them. Specifically, self-interpretation transforms an agent's referentially disunified SoI by *unifying* it. By directing the attitudes that comprise it toward the same intentional object(s). As I argue, this is possible because self-interpretation involves, in the usual case, making provisional commitments to care about certain things. I explain how this story can account for the connection between emotional interpretation and regulation in a way that is superior to the obvious competing account.

The last two chapters of the dissertation draw some of the value-theoretical implications of the framework that has been developed in the first three chapters. **Chapter 4** serves as a coda of sorts, and is formulated as a reply to a recent theme in the transformative experiences literature. Specifically, to a growing trend of critiquing the fundamental motivations and commitments that undergird it. Both Yao and Dover, for instance, argue that the anxieties that animate discussions regarding the rationality of transformative decisions only arise due to the acceptance of an overly individualistic ethos of *authenticity*. While I agree that some form of this ethos is indeed prevalent among theorists of transformative decision-making, I argue that it does more than generate certain characteristic anxieties. Specifically, I claim, the ethos of authenticity attunes its adherents to a distinctive kind of aesthetic value, *engagement value*, which inheres in actively shaping one's agency in a struggle to overcome challenges. At the same time, I maintain, Dover's and Yao's critiques are motivated by a competing, or complementary ethos, which I dub *the ethos of connection.* While the ethos of authenticity prizes *engagement value*, the ethos of connection highlights what I call *resonance value*: the value of being shaped by others and by one's environment.

If what I say in chapter 4 is true, then a natural question to ask is whether processes of self-formation and self-maintenance must force a choice between authenticity and connection—and the distinctive values that they enable. **Chapter 5**, in part, attempts to provide an optimistic, if rudimentary, answer. The chapter is pitched as a response to a highly pervasive sentiment in contemporary social philosophy of language. Many authors point to the pernicious effects that denying a speaker's self-ascribed identity components might have for that speaker. In many cases, refusing to give the intended uptake to such self-ascriptions can lead to palpable and substantive harms to speakers' autonomy, dignity, material conditions, and even bodily integrity. Given the

massive harmful potential of such uptake refusals, it is perfectly understandable that their positive aspects have been largely ignored—and, in some cases, ruled out categorically. While not denying this harmful potential, the chapter draws on the framework developed in chapter 1-3, to argue that sometimes, refusing to give uptake to a certain uttered ESA can help the speaker shape their evaluative outlook in beneficial ways. Given that uptake refusals can be both highly beneficial in some cases, and positively destructive in others, teasing apart the conditions under which they are indeed permissible—and, perhaps, obligatory—is a pressing question in the ethics of dialogue. I offer a few clarifications regarding what these conditions might include. Finally, the chapter concludes by proposing that, surprisingly, the mere possibility of uptake refusals makes uttering ESAs in contexts of sincere, intimate dialogue a particularly valuable way for someone to engage in identity formation and maintenance. As I argue, uttering ESAs in sincere, vulnerable dialogue is a way to realize *both* engagement and resonance value. As such, dialogue is an especially valuable mode of self-formation and transformation.

Chapter 1. SELF-ARTICULATION AS SELF FORMATION

> *No, you clearly don't know who you're talking to, so let me clue you in. I am not in danger, Skyler. I am the danger. A guy opens his door and gets shot, and you think that of me? No. I am the one who knocks!*
>
> *Walter White, Season 4, Episode 6 of Breaking Bad*

1. Introduction

Notwithstanding his protestations to the contrary, Walter White is, in fact, in danger. At this particular moment in the narrative arc of *Breaking Bad* he is cornered, vulnerable, hunted, and exerts very little control over his health, his family life, or, for that matter, his meth-cooking endeavors. His wife, Skylar, is characteristically both concerned for and underestimating—even infantilizing—him. He is desperate for money, recognition, agency… But this pivotal scene does not merely depict the hollow table-pounding of a terminally ill man-child who is in over his head. No. At this moment, Walter White channels, embraces, his criminal alter-ego, Heisenberg. With this speech, he *summons himself* into the drug-lord persona he has been hesitantly cultivating for three-and-a-half seasons. His words are transformative. They work to accelerate his transition (his rise? his fall?) into Heisenberg.

There is a long philosophical tradition, going at least as far back as Plato, which insists on the transformative power of dialogue. In some circumstances, it seems intuitively plausible that self-*interpretation*, and *formulation* can be a kind of self-*formation*. Examples of this rough idea at work abound. Some forms of talk-therapy are explicit in assigning an important therapeutic role to formulating and reformulating "self-statements;" Alcoholics Anonymous' twelve-step program famously ritualizes utterances of the formula: "my name is _____ and I am an alcoholic;" and Arno Michaelis, a de-radicalization activist and former neo-Nazi reports in a recent interview that his first step in recruiting teenagers into his white supremacist group was always "to get this kid

to identify as white" (Risafi, 2025). Our self-interpretations can be both edifying and corrupting. It is said that God created the world through speech, we mortals use speech to create and recreate our *selves*.

In this chapter, I will be concerned with a specific form of potentially transformative speech, which I will call "existential self-ascriptions." These are utterances in which the speaker ascribes certain motives, cares, concerns, values, or identity components to herself. "I value X," I care about Y," "I hate cowards," "I love you." I will take seriously the thought that expressing such statements can lead to genuine change in their speakers, and I will aim to explain why this is so. My paradigms for this inquiry are utterances of existential self-ascriptions in the context of sincere, vulnerable dialogue with our intimates. My goal is to make progress towards developing a model for how existential self-ascriptions operate in conversation to change who we are: what happens when we engage in existential self-ascriptions in dialogue.[1]

In section 2, I survey some evidence to the effect that existential-self ascriptions can be tremendously formative, for good or for ill. As we will see, among other things, such ascriptions play a therapeutic role in clinical mental health settings; but they may also be pivotal in the corruption of character. Making our self-interpretations *manifest*, as it were, by speaking or writing them is a particularly effective way of changing our motivational-evaluative makeup. Such self-interpretations are a way of *working out* which things matter to us, of effecting changes to our cares and concerns, broadly construed. Call this *the talking point*.

[1] This focus does not preclude the idea that cases of *monological* self ascriptions—talking to ourselves *about* ourselves—harbor transformative power as well. While I cannot discuss this issue at length here, I argue in another work that the transformative success of such cases is largely due to the fact that they make use of capacities and resources that are primarily dialogical. Transformative monological self-ascriptions, that is, are importantly derivative from *dialogical* self-ascriptions.

Section 3 presents a puzzle. Self-ascribing certain identity components is, presumably, a way of *changing* our cares and concerns. But quite obviously, not all self-ascriptions are created equal. Our interpretations cannot be acts of unadulterated spontaneous creativity. There must be *some constraints* on these interpretations. It seems that for self-ascriptions to work their transformative power, speakers must *resonate* with them. Subjects must take these ascriptions as tracking something significant about their *selves*. Minimally, this is because some of the fodder for our existential self-ascriptions, the material of which they are interpretations, is our emotional reactions. In turn, emotions are *evaluatively significant*. They present situations as bearing a certain import*:* as being good or bad for their subjects. But if this is the case, and if our existential self-ascriptions are interpretations of our emotional reactions, then it seems like the former don't *change* our values, but simply rearticulate, or affirm them. In other words, our self-articulations seem to routinely achieve an impossible goal: they both *alter* our cares and concerns, *and* are constrained by them.

In section 4, I propose that this puzzle arises only under the assumption that self-articulation consists simply in the detection and subsequent reporting of our mental states. That when we self-ascribe reasons, motives, and cares based on our affective feelings, we merely *label* a feeling that is waiting to be brought to the foreground of our conscious experience. Call this *the extraction model* of self-ascriptions. Against the extraction model, I argue that there is a crucial difference between articul*able* and articul*ated* feelings. According to the view I favor, there are multiple ways of specifying and acknowledging the same unarticulated feeling—and hence multiple, more *determinate* emotions to which it can be transformed. Finally, in section 5, I explicate the view further by contrasting it with two adjacent views regarding agents' acts of shaping the contours of their agency.

2. Existential Self-Ascriptions at Work

In this section, I survey a few examples of the kind of identity work that existential self-ascriptions make available to their speakers. As I show, a person's acts of verbally articulating her identity components may cause drastic changes to her evaluative-motivational makeup. Under certain circumstances at least, uttering existential self-ascriptions may alter speakers' cares, concerns, motives, and reasons for action. Call this *the talking point.* Crucially, as we shall see, such alterations may be either beneficial or harmful for their speakers—morally or prudentially.

A quick note: for the purposes of this chapter, I remain agnostic regarding whether acts of overt, verbal articulation of our values, cares, concerns, and their likes are necessary for existential self-ascriptions' putatively transformative force. I am happy to allow that (perhaps) merely applying a lexical concept to our feeling states, or covertly taking an evaluative stance towards some of our values and motives is sufficient to changing some of our identity components. Still, the evidence that I will harness to argue for the talking point involves exclusively cases where agents verbally articulate or express their existential self-ascriptions. So, at least for the moment, I am happy to limit the scope of my claims accordingly.

Now, before I try to impress you with the empirical literature, I want to bring out a few perfectly ordinary, putative examples of the phenomenon I am interested in. Consider these testimonies, courtesy of Redbook fashion magazine (Hilton Andersen, 2017):

> One night we were doing the dumb 'you hang up,' 'no *you* hang up' thing when I finally decided I really was tired and I'd hang up. So I said, 'Fine! I'm hanging up. Goodnight, I love you!' We both gasped and I said, 'Dammit, did I just say that out loud?' He was laughing and asking me to say it again. I started cursing, and I probably said 'damn it' and 'fuck' about 10 times until I realized what I had said was true.
>
> We were on our way to a haunted house when he leaned over and whispered that he thought he was falling in love with me. I blurted it out in return and realized I'd found my home. Fifteen years later, no regrets.

> [T]he very first time we said 'I love you,' it was our third date, which happened to be on Valentine's Day, and that cemented it for me. Four weeks later, we were engaged, and we got married two months after that.

According to these individuals, their existential self-ascriptions mark a pivotal moment in their relationships. Uttering these words seems like their way of *cementing,* as the third person aptly puts it, their relation to their loved one. Those individuals, we may suspect, have *settled the question* of whether they love their partner by *saying* that they do. On the other hand, these moments also have the character of a realization—perhaps even an epiphany. Engaging in existential self-ascriptions *clarified* to these individuals what they were feeling. I will come back to this tension in the next section, but for now, I want us to appreciate how natural it is to think—with, presumably, our informants—that their words altered something deep and important about themselves. I hope that these cases remind you, dear reader, of a time when it seemed to *you* as well that your own acts of self-ascription served to establish or fix some of *your own* identity components. Perhaps you felt (or thought you felt) the transformative power of your own words when you, too, said "I love you;" or when you confessed to your best friend that you think you might be queer; or when you muttered hesitantly (although, if you're reading this dissertation, this is less likely) "maybe academia isn't for me." If any of these possibilities strike you as remotely familiar, then I take it that you tentatively accept *the talking point*. I now turn to a few more cases that, hopefully, will make this acceptance less tentative.

2.1. Existential Self-Ascriptions in Therapy

Perhaps most obviously, the idea that talking about oneself can change important aspects of one's psychology is a *premise* of practically any kind of psychotherapeutic approach to mental health. More specifically, however, a number of clinical modalities explicitly argue for the benefits of articulating, expressing, and interpreting one's cares, concerns, and values.

For instance, a few models within cognitive behavioral therapy (CBT)—most predominantly, so-called *third wave CBT* approaches such as acceptance and commitment therapy (ACT)—place a high premium on "values work," that is, the process of assisting clients in clarifying and orientating towards their values. As Dahl and colleagues explain, it is "the process of *developing values statements* that guide client behavior" (Dahl et al., 2009: 10), or, alternatively, of "*developing* or *strengthening* values or creating a sense of purpose in chosen life dimensions" (63. My emphases). ACT protocols incorporate a wealth of values work techniques, which include structured questionnaires and surveys (Lundgren et al., 2012; Wilson et al., 2010), therapy cards sort activities (Veage et al., 2014), guided imagination and recollection exercises (Hayes et al., 2011; Wilson, 2009), as well as unstructured conversations. Importantly, values work is often construed as iterative, and therapists are encouraged to revisit clients' articulations of their values throughout the therapeutic process.

One oft-cited rationale for assigning values work a central role in the progression of the clinical process within such approaches is its affective, evaluative, and motivational benefits. Among other things, articulating clients' value systems seems to be strongly correlated with adherence to treatment plans for weight-loss (Forman et al., 2009), and for trichotillomania (Woods et al., 2006); with elevated motivation for engaging in habitual physical activity (Lillis & Bond, 2019); with increased pain tolerance (Branstetter-Rost et al., 2009; Páez-Blarrina et al., 2008); with higher self-efficacy during opioid use disorder treatment (Ameral & Palm Reed, 2021); and with reduced levels of cortisol during exposure to social stressors (Gregg et al., 2014). As a recent manual nicely summarizes it, "bringing values into the picture early can connect clients to a source of motivation that can help them through the difficult process of making important life changes. Meaning motivates" (LeJeune & Luoma, 2019: 52).

Another psychotherapeutic approach that incorporates the articulation of self-ascriptions as an integral part of its toolbox is narrative therapy. Generally, narrative therapy interventions aim to guide clients in examining, interpreting, and re-telling entrenched "life stories" in beneficial ways. As its chief developers, White and Epston explain, "since the stories that persons have about their lives determine both the ascription of meaning to experience and the selection of those aspects of experience that are to be given expression, these stories are constitutive or shaping of persons' lives (White & Epston, 1990: 40). According to narrative practitioners, by recasting dominant "self-narratives," clients are able to engage in the "'re-authoring'…of lives and relationships" (42). This re-authoring, in turn, is explicitly theorized as contributive to identity formation and transformation (Denborough, 2014; DeSocio, 2005; Ricks et al., 2014; White, 2011).

Practically all of narrative therapy's unique intervention techniques can be seen as geared toward forming and expressing constructive existential self-ascriptions and towards disengaging from harmful ones. Many narrative therapy exercises are designed to help clients "externalize" or cease from identifying with their problems; to attempt different ways of narrating important life events; to re-interpret and restate their group memberships and social affiliations; to design and participate in self-oriented ceremonies; and to partake in self-addressed letter-writing, to name a few examples (for reviews, see Etchison & Kleist, 2000; Ricks et al., 2014; White, 2024).

Recent empirical work lends support to narrative therapy's claims for efficacy in a broad range of domains. Narrative therapy treatments have been shown to improve symptoms for major depressive disorder (Vromans & and Schweitzer, 2011), and somatic disorders (Hu et al., 2024); and to increase hope and other positive emotions in subjects with depression (Seo et al., 2015). Furthermore, narrative therapy has been found effective in assisting emotional and social skill development in children (Beaudoin et al., 2016), and there is initial evidence that it is associated

with, among other things, symptom reduction and positive identity shifts in individuals with eating disorders (Conti et al., 2022).

Naturally, a comprehensive review of all therapeutic approaches that explicitly encourage patients to verbalize, or otherwise manifest their self-interpretations in the forms of existential ascriptions is far beyond my scope. Apart from ACT and narrative therapy, one can find explicit protocols for such processes in logotherapy (Frankl, 2014), gestalt therapy (Perls et al., 1994), expressive writing (Pennebaker, 1997), focusing-oriented psychotherapy (Gendlin, 2012), and a number of existential therapy approaches (Craig et al., 2019). I hope the point is clear enough. Existential self-ascriptions seem to have a kind of edifying transformative power. Self-interpretation in the context of a therapeutic dialogue can lead to dramatic positive changes to our motivational, evaluative, and affective internal landscapes.

2.2. Existential Self-Ascriptions and Radicalization

Expectedly, not everything is sunshine and rainbows. The (putative) transformative power of speech may be harnessed to facilitate improvements in mental health, but we can also talk ourselves into the mouths of harmful ideologies, moral depravity, and agential failures. One kind of process in which the harmful potential of existential self-ascriptions is vigorously expressed is political radicalization. As I will show, *the talking point* applies in this sphere as well. Talking about ourselves in certain ways may change our evaluative-motivational makeup in a way that is troubling from both a moral and an agential perspective.

First, though, I want to acknowledge an important caveat, poignantly advanced by Cassam (2018). In general, it is ill-advised to think that there is *one* generic, intelligibly specifiable process of radicalization. Humans are astonishingly complex particulars, and any model of radicalization is, plausibly, bound to leave some individual radicalization processes unaccounted for. I will rely on a specific research program in the psychology of radicalization, but I do not think that the

specific model advanced by this program can subsume (at least not in a highly informative way) the entire array of political radicalization processes. A limited scope is sufficient for my purposes. I want to demonstrate how, *sometimes*, existential self-ascriptions can make us prudentially and morally worse off.

Notably, I am more concerned here with what might be called *psychological* rather than *behavioral* radicalization. The former is the process of developing a certain, *extremist* cognitive, affective, motivational, and evaluative psychological makeup, whereas the latter involves a turn to overt political violence. While the two processes often overlap, neither is necessary nor sufficient for the other. Some people may commit acts of political violence without having a stereotypical radicalized psychological makeup, and others may have an extremist mindset yet may never commit illicit extremist activities. As Sageman succinctly notes, "very few people talking about violence actually go on to use it" (2017:90).

With these qualifications in mind, I want to consider Kruglanski and colleagues' highly influential ideas regarding the social psychology of political radicalization. According to Kruglanski's group, an extremist psychology—both in the political domain, and in other areas—is characterized by a distinctive kind of *motivational imbalance*.

Humans have a variety of different, often conflicting needs, desires, and goals, each of which places constraints on what we take to be viable behavioral pathways for the attainment of others. My desire for a cordial work environment, for instance, constrains the ways in which I communicate with my colleagues to ask for a quieter office. Often, however, priorities change and one goal comes to dominate our occurrent motivational landscape: "a parent may cancel a professional obligation on learning that her or his child is ill; a physician may interrupt a family vacation to attend to an emergency; and a police officer might forego an uninterrupted night's

sleep to hurry to a crime scene" (Kruglanski et al., 2019: 38). In such cases, constraints that are placed on potential actions aimed at the attainment of the prioritized, dominant goal are released: "behavior that formerly was barred from consideration now constitutes a legitimate option for the pursuit of [the goal] and is likely to be selected" (39). The motivational economy of psychologically extreme individuals tends to be governed by a single, dominant goal, which both inhibits the pursuit of alternative goals, and disposes the individual to engage in behaviors that would be barred from consideration when other motives are activated (Bélanger et al., 2013). Psychological extremists are (relatively) monomaniacs. They are ruled by a single commitment, or a narrow set of typically congruous commitments.

Over the course of many studies, Kruglanski et al. demonstrate how certain priming interventions can lead to the kinds of motivational singlemindedness that, on a broad scale, is characteristic of particularly political forms of extremism. More specifically, many of the experimental manipulations that the group utilized involve subjects explicitly *construing themselves* in certain ways.

In one study, US college students wrote an essay relating an instance of either personal success or personal failure, followed by an assessment of the strength of their national identification as Americans. As the researchers predicted, people who were primed by relating a past failure reported a significantly stronger sense of collective identification. These findings lend initial support for the team's hypothesis that people who interpret their lives as involving failure, humiliation, or loss of personal significance would be prone to embrace a more collectivistic identity, other things being equal (Orehek & Kruglanski, 2018).

In a different study (Orehek et al., 2014, study 1), participants were primed by using either singular first-person pronouns (e.g., "I," or "me"), or collective pronouns ("we" or "us"). As it

turns out, subjects in the collective pronouns condition demonstrated reduced death anxiety (Kruglanski et al., 2019). Other experiments in the series (Orehek et al., 2014, studies 2 and 5) demonstrated stronger approach and/or lesser avoidance of death under collectivistic priming, using different manipulations. In general, results from the study indicate that "compared to an independent [individualistic] self-construal, an interdependent [collectivistic] one attenuates individuals' anxiety about death (Experiments 1-3), fosters a greater willingness to sacrifice the self for a cause (Experiments 3 and 4), prompts lesser avoidance of death (Experiment 5), and greater willingness to approach death (Experiment 5)" (Orehek et al., 2014: 271).

In another experiment, individuals were either assigned to a condition designed to trigger feelings of insignificance, or to a neutral, control condition. Under the "loss of significance" condition, individuals were instructed to "think back to a time when they were humiliated and to describe the details of the situation." The group found that the manipulated condition "increased individuals' political extremism and that this relation was statistically mediated by individuals' [need for closure]" (Kruglanski et al., 2019: 108; cf. Webber et al., 2018).

I think that there is a case to be made here about the importance of certain self-ascriptions in the development of extremist motivational imbalances of the kind Kruglanski and his colleagues are interested in. Talking about oneself as a part of a group reinforces one's collective identity—and collectivist motivations; describing oneself as a failure leads one to emphasize motives for closure and redemption; and speaking about one's affiliation with an oppressed minority may change one's motivational-evaluative makeup in a way that alters what one may perceive as legitimate actions towards liberation. To re-emphasize, such evaluative-motivational changes are neither sufficient nor necessary for behavioral extremism, and sometimes they are conducive to admirable prosocial behavior. But they may also tip agents' motivational scales in ways that make

them more prone to take morally troubling actions, and lead to monomaniacally narrow—and hence impoverished—lives.

3. Change and Constraint

The lesson I wish to draw from the previous section is that *the talking point* has wide application. Expressing our self-interpretations can induce changes to our evaluative and motivational makeup, for good or ill. This, I now turn to argue, leads to a puzzle. On the one hand, existential self-ascriptions are transformative: at least sometimes, they seem quite effective at changing our values, cares, and concerns in many ways. On the other hand, as we will see momentarily, it seems like they are *constrained* by what we care about, find valuable, and take to be important aspects of our identity.

Sincere existential self-ascriptions express our self-interpretations. As such, they seem to be assessable for *veridicality*. Our self-interpretations (and the ascriptions expressing them) are beholden to normative constraints that safeguard their propriety. Simply put, not everything goes. Our interpretations of ourselves ought to be *credible*. They must respect what Schechtman calls "the reality constraint." That is, they must acknowledge "the most basic features of reality" (Schechtman, 1996: 119), including facts about our historical, psychological, social, and material realities. Self-ascriptions that fail to respect these requirements for coherence with reality are delusional, or self-deceptive, or simply insincere.

Apart from this abstract point about what self-interpretations *are*, it seems that in order to *effectively* induce motivational or evaluative transformation, our existential self-ascriptions should follow the reality constraint. It is important that self-ascriptions will track something real about their utterer. A self-interpretation that is too disconnected from reality cannot support the kinds of inter- and, for that matter, intra-personal interactions that are crucial for the maintenance of a stable

evaluative and practical identity (cf. Lindemann, 2014). If your self-interpretations consist of too blatant or too many pieces of bullshit, they simply won't be convincing: they will result in too many dissonances, too much friction with reality.[2] A commitment to upholding such self-interpretations come-what-may is precisely the stuff of delusions and other pathologies of agency. Hence, at least in non-pathological cases, self-ascriptions that are too at odds with reality lack the requisite force to effect the kind of personal change that I've been at pains to establish that they *can* and often *do* effect.

This observation—and the air of paradox that follows in its wake—is reflected, for instance, in a telling duality that characterizes thinking about values-work in the psychotherapeutic approaches we have surveyed in the previous section. On the one hand, all of these approaches encourage clients to view their values, cares, and concerns as something that they *already* possess, and that should be *discovered* or *revealed* through the therapeutic process. ACT therapists, for instance, are encouraged to look for *indicators* of a client's existing values (LeJeune & Luoma, 2019: 75-6; Luoma et al., 2017), and in many ACT exercises, clients are instructed to describe cases that involve living a life that is consistent with the values that they currently possess. Importantly, such exercises are theorized as aimed to *clarify* the client's values (see, e.g., Wilson & Sandoz, 2008: 98). Similarly, narrative therapy is theorized as aiming at identifying, or revealing components of people's lived experiences that "inevitably [fall] outside the dominant stories about the lives and relationships of [these] persons. Those aspects of lived experience …. provide a rich and fertile source for the generation, or re-generation, of alternative stories" (White & Epston,

[2] The reality constraint can trace, I think, two aspects of what Goldie (2003; cf. Mackenzie, 2008) calls "successful self-narratives," namely, *coherence* and *meaningfulness*. Our self-interpretations must be coherent in that they "[represent] a series of events … suggesting causal connections between them" (307), and meaningful in that they explain how our actions, and action-guiding emotions could make sense to us (or, at least, have made sense to us at the time of their occurrence).

1990: 15). As one review suggests, "[u]sing narrative therapy techniques, a counselor assists clients in separating from their problem-saturated stories to *find the hidden story* of the way people wish to live their lives" (Ricks et al., 2014: 100, my emphasis).

On the other hand, simultaneously, these approaches treat clients' values as not only *malleable*, but as *up to the client* in important respects. For example, CBT approaches like ACT construe cares, concerns, and values as *freely chosen*, or *constructed*. Clients are encouraged to "try on values" (Dahl et al., 2009: 167), and values-work is often interpreted as a kind of iterative, developmental process, involving trial and error and decision making (Dahl et al., 2005: 61; LeJeune & Luoma, 2019: 10; Wilson, 2009: 64). The same is true for narrative approaches to therapy and their emphasis on the client's authorship of and responsibility for her practical identity (e.g., White & Epston, 1990: 65, 82-83). Values, then, seem to be both firmly anchored in subjects' psychological reality, *and* highly plastic.

This air of paradox around existential self-ascriptions can be recognized in more sinister contexts as well. As we have seen, some self-ascriptions can contribute to their utterers' radicalization: they can, for instance, turn them into more motivationally single-minded, and more willing to sacrifice and engage in violence for the sake of a political cause. However, presumably, these self-ascriptions should also track something *psychologically real* about their utterers: their identity-constituting membership in an oppressed group, a failure they endured, a humiliation they suffered, and so forth. In such cases too, it seems that to be moved or transformed by our self-ascriptions, we must *resonate* with them. We must feel as though they track something true—and perhaps also deeply significant—about us.

In fact, we might think that the paradox is even more thorough. At the very least, our interpretations of ourselves must incorporate salient aspects of our material, social, and

psychological realities. But our psychological reality very much includes our affective responses. A realistic self-ascription must acknowledge and make sense of the emotions that we feel. In fact, our emotional reactions to our situations are a key aspect of what our self-interpretations are interpretation *of*. Convincingly construing a scenario as humiliating or exalting often requires that we *experience* humiliation or exaltation; and realistically suggesting that I value intimacy or adventure requires that I *feel* something towards scenarios that embody or compromise those values.

And herein lies the problem. If our affective feelings constrain our self-ascriptions, then it seems like most of the interpretive work is done *before* anything has been uttered, before *we* have had a chance to make sense of our emotions. This is because emotions are themselves *evaluative*. They present things as bearing a certain significance for us, as *mattering* to us in different ways. Emotional reactions involve experiencing things as good, or bad, or otherwise evaluatively important, and hence they comprise at least attenuated ways of *caring* about or of *valuing* those things. So it seems like existential self-ascriptions *can't* really alter our values or cares. At best, they are ways of re-articulating those values or cares. In other words, existential self-ascriptions can't possibly be doing the work that I've spent the previous section arguing that they do.[3]

To summarize, our existential self-ascriptions seem to routinely achieve an impossible goal. On the one hand, at least sometimes, they appear to be deeply transformative of our values, cares, and concerns. On the other hand, the same self-ascriptions seem to be highly constrained by our emotional reactions, and hence by what we *already* value, or care about, or are concerned with.

[3] This thought is not limited to so-called "appraisal theories" of emotion (e.g., Ellsworth, 1991; Ellsworth & Scherer, 2003; Frijda, 2009; Lazarus, 1991; Moors et al., 2013) For instance, according to Ekman (1992), emotions involve automatic appraisal processes; Tappolet suggests that emotions are perceptions of evaluative properties (2016); Goldie argues that emotions are "evaluative feelings" (2000); Helm defends a view of emotions as "felt evaluations" (2001); and Prinz argues that emotions are "embodied appraisals" (2003).

How can such ascriptions both *alter* our cares and concerns, *and* be constrained by them? Call this the Puzzle of Change and Constraint.

4. Description and Determination

I suggest that the Puzzle of Change and Constraint arises only under a certain presupposition about what the articulation of our values, cares, and concerns consists in. Namely, that articulation is mere *detection* and subsequent *description*. That it is a matter of *uncovering*, or calling attention to some pre-formed psychological reality that is "waiting for some sort of liberation" (as Nehamas describes what he calls a "vulgar" Freudian idea, 1983: 391). Call this presupposition *the extraction model* of expression. According to the extraction model, when a subject is making an existential self-ascription, what she does is track a specific value or concern that is embodied by, among other things, her affective reactions, and then communicate it to her audience. On this picture, it seems weird indeed to think of our self-ascriptions as having any kind of motivational or evaluative transformative force. They just *rehash*, *reveal*, or *re-present* a pre-existing inner motivational-evaluative landscape.

But there is an alternative way of understanding what we do when we bring our communicative capacities to bear on our affective, evaluative, and motivational experiences. On this model, expression—and more specifically, articulation—is an act of *determination*. Here, I take inspiration from Charles Taylor, who writes:

> [A]rticulations are not simply descriptions, if we mean by this characterizations of a fully independent object, that is, an object which is altered neither in what it is, nor in the degree or manner of its evidence to us by the description … on the contrary, articulations are attempts to formulate what is initially inchoate, or confused, or badly formulated. But this kind of formation or reformulation does not leave its object unchanged. *To give a certain articulation is to shape our sense of what we desire or what we hold important in a certain way*. (Taylor, 1985, my emphasis)

We can make good sense of Taylor's suggestion if we think of many of our pre-articulated affective states as having *determinable objects*, and as incorporating *determinable psychological*

attitudes, in a manner somewhat akin to the way we think of, say, colors. Crimson, scarlet, and vermilion are different ways of being red. Analogously, feeling, say, insecure, or jealous, or aggrieved are different ways of experiencing a situation as threatening; and valuing one's heritage, caring about the material needs of one's community members, and striving to aggrandize one's culture are (arguably) different ways of being concerned with one's in-group.[4]

My suggestion is that often, the patterns of emotional and attentional reactions that comprise the materials for our self-interpretations can be made more determinate through these self-interpretations. By self-ascribing reasons, values, cares, and concerns on the basis of our emotional reactions, we determine those reactions. We don't only specify, but we *render more specific* what, exactly, they are about, and what kinds of attitudes they are. Call this *the determination model* of expression.[5]

Toy example: I should really return that book I borrowed from you already, but as I approach your office, I find you engaged in conversation with one of our colleagues. I lower my posture, softly knock on the door, and gently place the book on your desk without setting foot in your office. Then, not uttering a word, I wave goodbye to both of you and quickly remove myself from your sight. My behavior is clearly expressive of my evaluative perspective with respect to this situation, but presumably, this perspective may be credibly interpreted in a number of ways.

[4] The analogy with colors isn't perfect. First, while this is a contentious matter, some theorists think that instantiated colors are *maximally determinate*. There are no instantiations of red simpliciter, but only of (maximally determinate) kinds of red. This, as I will clarify momentarily, is not the case with our pre-articulated evaluative reactions. We may have evaluative reactions that are not maximally determinate with respect either to their intentional objects or to the kinds of attitude that they consist in. Secondly, different instantiated evaluative attitudes are not mutually exclusive in the same way that different instantiated colors are. The same thing cannot be both vermilion and cyan, for example, but one can both care about the material needs of one's community members *and* strive to aggrandize one's culture.
[5] I should note that existential self-ascriptions may have other, *complementary* effects to the determination processes I am interested in here. For instance, perhaps sometimes, existential self-ascriptions might boost the motivational force of a pre-formed attitude, making it more efficacious (I thank Dana Nelkin for this thought). The point is not that determination is the *only* relevant effect that self-ascriptions might have, but that this effect is of particular interest in the context of diffusing the Puzzle of Change and Constraint.

It may be an expression of respect towards your ongoing conversation, it might reflect my submissive stance towards you, it may highlight my disapproval of intrusive behaviors in general, or it may betray my insecurity. On the determination model, it is perfectly possible that in the moment of action, I am guided by a pattern of emotional and attentional responses, and that this pattern is not maximally determined with respect to the specific values or concerns it embodies. The different interpretations I have enumerated are determin*ates* of my action-guiding, evaluative perspective in the moment of action. But we should keep in mind that, as noted already, not everything goes. Some interpretations of my pre-articulated evaluative perspective are wrong, or delusional, or unconvincing. It is highly dubious to suggest that my behavior is an expression of my desire for dominance, for instance. And such an interpretation is dubious precisely *because* such a desire or concern is *not* a determinate of my determinable evaluative perspective, as it is manifested in my action-guiding emotional reactions.

Recognizing the plausibility of the determination model requires making explicit what valuing or caring consists in. For our present purposes, the specific differences between valuing and caring, as well as the details of extant account of these attitudes are not crucial. What *is* significant is that to value, care for, or be concerned with something involves a reliable pattern of cognitive, but also *motivational*, and *affective* dispositions (Callard, 2018; Helm, 2001; Jaworska, 2007b; Jaworska & Wonderly, 2017; Kolodny, 2003; Scheffler, 2010; Seidman, 2009; Shoemaker, 2003; Wallace, 2018). To value or care about something does not only mean treating it as a source of reasons for action, but also to be emotionally *attuned* or *vulnerable* to it in certain ways. It involves, that is, being disposed to undergo certain affective states in response to the weals and woes that befall the valued or cared-for object; as well as in response to alterations in one's relationship to it. Valuing one's kin, for example, consists, in part, in being disposed to feel

indignation at their oppression, pride at their triumphs, and anxiety at the prospect of isolation from them.

As I said, in certain cases, we may be prone to undergo complex emotional reactions that are not sufficiently unified or specific to render them instances of full-blown valuing or caring. They may, that is, not express the full pattern of emotional and attentional responses that we would be inclined to classify as *bona fide* caring or valuing attitudes with respect to a fully determinate object.

On the other hand, these reactions may still reflect an inchoate sense of "what we hold important" (to use Taylor's words). Under such circumstances, our affective makeups may be *multiply interpretable.* They may have several legitimate interpretations, and the act of settling on, or of uttering one of those interpretations may help *determine* those makeups. That is, it may unify such bundles of responses into something that resembles more closely full-blown valuing or caring attitudes. While I cannot argue for this suggestion at length here (I offer an extensive defense of it in future work), it might be helpful to gloss over two ways in which it might be made more precise.

For one, a subject may undergo an affective episode that does not have a highly specific *object*. Moods, for instance, are often construed as representing *underdetermined* objects (Rossi, 2021); or a subject's *total environment* (Mitchell, 2019); or the *world-as-a-whole* (Solomon, 1993; see also Crane, 1998; Goldie, 2000; Kriegel, 2019); or *"things" indiscriminately* (Kriegel, 2019). But even if we set aside vexed questions regarding the ontology of moods, it is highly plausible that other affective states have objects that are non-specific enough to admit to multiple interpretations. To give just one example, a person may rage at the humiliation of a member of their in-group, and this rage may be legitimately interpreted as targeting the humiliation of as

specific *individual*, a representative of a *minority group*, or *social injustice* in general (or all of the above).[6]

Secondly, a person may be disposed to experience bundles of affective reactions to a situation that are *compatible* with each other yet not *referentially unified* enough to comprise a valuing attitude towards one specific intentional object. Before a job interview, I may experience a complex of slight—perhaps barely noticeable—affective responses to my predicament: a sense of insecurity about the font of my CV, irritation at the noises in the waiting room, discomfort about my overly-presumptuous clothes, and a feeling of profound vulnerability. Together, while these attitudes are directed at different (and partly minute) aspects of my situation, they may be interpreted as implicitly disclosing an overarching value or concern. Importantly, this value or concern might be *underdetermined*, and hence multiply interpretable, in the sense I am trying to convey. For instance, it may be legitimately interpreted as a concern for my sense of self-worth, or for my livelihood, or for the fate of my professional aspirations. In turn, in this context, settling on *one interpretation*, or *self-ascribing* a determinate value or concern, may lend unity *prospectively* to the bundle of attitudes that I am instantiating at the moment. For instance, it may prime me to instantiate those attitudes—and a host of other affective responses—in different contexts that I take as having a bearing on the object that I have interpreted as being the target of my affective responses. In other words, it may move me closer to *actually valuing* that object.

I should stress again that this is merely a sketch of the determination model of expression (some of the much-needed details are filled in the next chapters of this dissertation). But as a further proof-of-concept, note that it can help us make sense of the thought that existential self-

[6] This idea is comparable to—although not quite the same as—Velleman's understanding of Freudian drives, as "highly labile psychic energies, which have only a vague direction in themselves but can be invested in specific activities" (Velleman, 2020a: 418).

ascriptions can be both beholden to and transformative of our evaluative-motivational attitudes. On the one hand, our affective reactions can be thought of as pre-articulated, determinable (or under-determined), inchoate or rudimentary modes of finding important certain aspects of our situations. Our existential self-ascriptions are often interpretations of such pre-articulated evaluations, and—as we've seen—they must respect those evaluations to be effective. We cannot convincingly interpret ourselves as caring about something we do not find important, even in an attenuated sense. This condition is satisfied when we self-ascribe reasons, values, cares, and concerns that are genuinely determin*ates* of our determin*able* pre-articulated evaluations, and it is breached otherwise. On the other hand, existential self-ascriptions can be transformative in the sense that they render our pre-articulated modes of caring more determinate. They are ways of constructing sharper, more specific, communicable motives.

It won't be a gross mistake, I hope, to characterize the range of phenomena I am trying to describe as a *generalized form* of what Callard calls "aspiration" (2018, 2020), that is, "the rational process by which we work to care about (or love, or value, or desire…) something new" (2018: 4). In aspirational activities,

> [t]he aspirant sees that she does not have the values that she would like to have, and therefore seeks to move herself toward a better valuational condition. She senses that there is more out there to value than she currently values, and she strives to come to see what she cannot yet get fully into view (2018: 5).

Like the processes of self-interpretation that we have been examining, aspirational activities involve "the aspirant's value-discovery [which] is at the same time her self-creation" (14). However, while aspiration is purposive and intentional in a substantive way, the kind of self-transformation that is (sometimes) implicated in the articulation of existential self-ascriptions need not be. Callard remarks: "On an aspirational account of self-creation, the creator does not determine, choose, or shape the created self; rather, she looks up to, imitates, and seeks to become

the created self" (13). What our discussion so far suggests is that many cases of value-determination are *not* aspirational in this regard. When I seek to make sense of the inchoate, action-guiding sense of what mattered to me in a certain situation, I do not necessarily "look towards" a self that I wish to become. In fact, I might not have in view any kind of value that I desire to realize or to get a better grasp on. Furthermore, I might not appreciate the event of self-interpretation as an opportunity for value-determination. Instead, I may simply try to reach for *any* legitimate interpretation that will allow me to make sense of my condition.

Perhaps the most controversial aspect of Callard's view involves her claim that to account for the rationality of aspirational activities, we must acknowledge the existence of a distinctive kind of reasons which she calls "proleptic." Such reasons have a "proximate face that reflects the kinds of things that appeal to the person [the aspirant] is now and a distal one that reflects the character and motivation of the person she is trying to be" (73). They embody the aspirant's "inchoate, anticipatory, and indirect grasp of some good she is trying to know better" (72). I want to remain agnostic about whether proleptic reasons are truly necessary to the understanding of aspiration,[7] and instead wish to point out that we can think of aspiration as one specific case of value-determination. For example, both the aspirant and the person who existentially self-ascribes values, cares, and concerns (they are often the same person), have a vague sense of what matters to them, of what they find important (this may be the "distal face" of the aspirant's proleptic reason). After all, the aspirant, like the self-interpreters I've been discussing, is not completely alienated from the values she is rendering more determinate during the aspirational process. As Callard admits, "for the transformative options to get a grip on us, the transformative process must *already be under way*. Thus we needn't imagine that the agent evaluating them stands fully outside

[7] See, e.g., Katsafanas (2021) for some worries.

of what she evaluates" (60, emphasis original). Both aspirational activities and existential self-ascriptions reduce the vagueness that characterizes their agents' action-guiding values, by making these values more determinate.

5. Affinities and Differences

The view I am developing bears affinities to several views regarding the ways in which our motives, values, and agential psychological structures are shaped by our self-interpretations. In this section, I wish to consider two of these views—J. David Velleman's and John Doris's—in a way that may help bring my positive proposal into sharper relief.

First, on Velleman's account of self-narration, agents are partly motivated by a desire to act in ways that *make sense to them*, and this means that they have pro tanto motivating reasons to see to it that they act in intelligible ways given, among other things, what they say about themselves. Narrating our lives, according to Velleman, is a way of saddling ourselves to a particular path of action. It is a way of *pressuring ourselves* to act so that our lives will fit the stories we tell about them. To use Velleman's example, saying "I am leaving" is typically not issuing a prediction about my future behavior, but rather a way of prompting myself to leave (Velleman, 2020: 291). In general, Velleman thinks that whatever makes the most sense for us to do is determined in part by what he calls our *self-conception*. This is a description "under which your actions and reactions make sense to you in causal explanatory terms" (Velleman, 2009: 16, fn.). So, one attractive thought about existential self-ascriptions is that they are ways of structuring our self-conceptions. By articulating which things we care about, we in fact *commit* to caring about those things.

Now, on this level of specification, the view does not tell us what—if anything—constrains our self-interpretations. An agent's self-narration must be faithful to her biography, but our

biographies themselves admit of many conflicting interpretations. Still, as we have seen, some self-interpretations (in the form of existential self-ascriptions) seem to be more effective, more credible, more commissive than others. Velleman can be read as responding to this worry by suggesting that the key lies in our emotional responses: "a description of events qualifies as a story in virtue of its power to initiate and resolve an emotional sequence" (199). A *good* self-interpretation, then, renders a sequence of events intelligible by completing an emotional cadence, set in motion by a specific affective reaction to our situation.

This story, however, seems unable to dissolve the Puzzle of Change and Constraint. For your self-interpretation to be effective, it must resolve an emotional *tension* that you experience. But such emotional tensions, on Velleman's view, are resolvable in predetermined, highly predictable ways. As he says, "The cadence that makes for a story is that of the arousal and resolution of affect, a pattern that is biologically programmed" (195). Or "a sequence of emotions has beginnings and endings, because emotions naturally sort themselves into *ticks* and *tocks*. Emotions naturally qualify as initiatory or conclusory" (197). These remarks suggest that, once you feel an emotional reaction that calls for interpretation, the range of eligible interpretations for this reaction is highly narrow. Specific emotions have pre-determined patterns of resolution, and issuing a successful interpretation means identifying the specific initiatory emotion, and the specific narrative that makes intelligible its "biologically programmed" course of development. But if this is correct, our interpretations don't *really* change anything important about us! Specifically, they don't change the values, cares, and concerns that are involved in the initial emotional tension that we experience to begin with. All our self-ascriptions must do (all they can do, perhaps) is correctly *identify* the emotional tensions, and accurately *describe* them, in a way that would resolve the tension.

There is much in Velleman's position I am sympathetic to,[8] but its inability to solve the Puzzle of Change and Constraint seems to be due to an acceptance of something like the extraction model: the thought that our interpretations are tracking something that was already there all along, fully formed, patiently waiting to be brought to the surface of our consciousness. My suggestion (my friendly amendment, if you will) is to loosen slightly the way Velleman views the affective reactions that call for our self-interpretations in the first place. The evaluative attitudes that we strive to make intelligible with our existential self-ascriptions are *determinable*. They do not come to us as parts of pre-programmed patterns. Our interpretations can *determine*, or render more specific, what Velleman calls "initiatory emotions." And such acts of determination influence the dynamic trajectories that our emotional, evaluative patterns proceed to take.

Doris too suggests that an important way of prospectively aligning our agency with our values involves collaborative dialogue about our biographies. He agrees with Velleman that "people are motivated to maintain a correspondence between story and self" (Doris, 2015: 143), and that "one's biography helps tie one to the mast, and secures behavior expressing one's values, even in the face of unfavorable circumstances or unstable inclination" (144). Insofar as our biographies are comprised, in part, of our existential self-ascriptions, I am onboard with the mast point.

On Doris's account, much of our prospectively efficacious biographies consist of confabulations. As he puts it, "[t]he influence of biography has as much (or more) to do with

[8] Although I will not defend it here, my position is different in some important respects. First, I think that, insofar as we are pressured to be intelligible, this pressure is primarily other-directed rather than self-directed. We have a standing desire to make sense to *others*, which might conflict with other foundational standing desires (here I am in agreement with Walden, n.d.). Second, I think that, to the extent that we have a self-conception, it is typically quite fragmented and irregular (on this point, I think, I am in agreement with Doris, 2002, 2015). Third, I think that self-ascriptions of the kind I am concerned with determine our cares not by changing our conception of ourselves, that is, not by altering a higher order, metacognitive state, but *directly*, by changing our cares, values, and reasons themselves (here, I am told, I am in agreement with Hegel).

subjective personal salience than objective historical accuracy, and an inaccurate biography may sometimes be more motivationally compelling than an accurate one" (145). In our terminology, the idea is that often, our existential self-ascriptions may amount to agents faking it 'til they're making it. Motivational force is more important, more transformative, and often stands in direct conflict with narrative accuracy. In fact, ignorance, as he says, "is the life-blood of agency" (129). On the other hand, while "you become your bullshit … your bullshit better become you" (144). Doris agrees that our self-narrations cannot be purely fantastical, but it is unclear which constraints, he thinks, should be imposed on our self-interpretations. Especially given that "self-ignorance often functions to effect self-direction, and its *absence* can be an impediment to agency" (144). If Velleman's story seems to place too many constraints on our self-interpretations, Doris's seems to place far too few.

As I've said, I am inclined to think that accuracy and efficacy are not as separate as Doris might make them appear. In fact, I think that typically, self-ascriptions are effective *because and insofar* as they are sufficiently accurate. This is not a conceptual necessity, of course, but more often than not, if we are to be *convinced* by our own biographies, they better be constrained by what we recognize as our material, relational, and—crucially—affective-motivational realities.

I have said that while Velleman's account struggles to make intelligible the transformative powers of our existential self-ascriptions, Doris's seems to struggle with articulating the ways in which they are constrained. I have also registered my suspicion that Velleman's account faces this difficulty because of a tacit adoption of something like the extraction model. I now want to suggest that, perhaps surprisingly, Doris's sharp (too sharp to my liking) distinction between transformative efficacy and descriptive accuracy stems from a similar acceptance of the extraction model.

One of Doris's key points in his positive account of how self-ignorance contributes to our agency involves *rationalizations*, that is, "serviceable explanations for [agents'] behavior, even when information, including information about their own psychological states, is seriously inaccurate and/or incomplete" (141). Now, according to Doris,

> a central form of rationalization presents behavior as an exercise of agency: *I chose to do this. I meant to do that. I had my reasons*. By presenting behaviors as exercises of agency, rationalizations depict their bearer as appropriately subject to participant reactive attitudes and a fit partner for reciprocal social exchange. (142)

I don't deny that we often do engage in rationalizations of the kind Doris describes, but I'm inclined to think that many cases he might construe as "healthy confabulations," are cases in which our evaluative attitudes are simply determin*able*. Doris seems to suggest that many cases of agential reason-giving involve *incorrect* ascriptions. I suspect that a lot of these cases involve acts of determining a previously indeterminate evaluative pattern. By engaging in games of giving and asking for reasons, agents don't only (or mostly) confabulate their motives, cares, and concerns, but rather decide—within constraints—*what* their motives, cares, and concerns *are*. Doris summarizes: "where self-awareness fails, and 'I don't know' won't do, rationalization ensues, disciplined not by some bedrock reality, but by acute sensitivity to social cues" (143). On the determination model that I've been advancing, "bedrock reality" is often more flexible than Doris seems to think, and it may be readily bent to accommodate social cues.

6. Conclusion

In the heat of the moment, in the flow of action, we are often moved by an inchoate, determinable pattern of affective responses that may legitimately be interpreted and articulated in different ways. I have tried to suggest that such acts of articulation—when and if they occur—may helps us *fix*, or *render specific* the cares, concerns, and commitments that are embodied by such affective patterns. However, I also want to insist that, at least in non-pathological cases, for the

articulation to work, it cannot be divorced from our psychological reality. Our self-interpretations must be appropriate, and this means that they must be constrained by our concrete situation and by our actual affective responses to it. Our determinable affective responses are multiply-interpretable, but not radically so. This *determination model* of expression can explain why existential self-ascriptions seem to be both transformative and revelatory; how they appear to change what we care about, but also to reflect genuine events of self-discovery.

What I have *not* done in this essay is provide a detailed analysis of the transformative process that includes existential self-ascriptions. What makes existential self-ascriptions succeed or fail in inducing changes to our motivational-evaluative reality? Do they require a certain uptake by their speaker or by their hearers? And is there a more nuanced (and empirically adequate) description of the inchoate, determinable affective patterns that guide our actions in-the-moment, and that our existential self-ascriptions serve to articulate (some of these questions are answered in what follows). What this chapter was aimed to achieve is, primarily, the reader's openness to the idea that acts of self-interpretation can be transformative without being false. That self-formulation can be a kind of self-formation *without* being merely confabulatory.

Chapter 2. AGENTIAL UNITY, THE SENSE OF IMPROT, AND THE NECESSITY OF GIVING A DAMN

1. Introduction

Agency theorists have pointed out several tight connections between our capacity for caring and many key agential phenomena. It has been argued, among other things, that caring is necessary for free will (Shoemaker, 2003), for identification, or 'internality' (Jaworska, 2007b), for the unique moral status that beings such as ourselves possess (Jaworska, 2007a), and for our capacity to value (Seidman, 2009). This chapter expands on this trend by making two claims. First, I argue that, at least for creatures like us, *agency itself* requires that we cultivate a set of care-like dispositions, or a "minimal mode of mattering". Such a mode of mattering is necessary to initiate action, as well as to sustain and guide it. In simple terms, you cannot act if you don't give a damn. Second, in order to have full blown *cares* with respect to some object, this mode of mattering must be *stable*. It is not sufficient that some things matter to us—that we take some things rather than others to be important in some attenuated sense—but that we find *the same things* important in *the same contexts*. Human agency requires a pattern of mattering, and human *cares* require a reliable, regular pattern of mattering. I call this pattern *the sense of import*.

In sections 2-4 I defend the first claim of this chapter, according to which action requires a minimal mode of mattering. I do so by first discussing two problems that any agent must face in order to act: the problems of action selection and action specification. In section 3, I explain why creatures like us cannot solve these problems either by relying exclusively on the operation of automatic, sub-personal mechanisms or on reflective, practical deliberation. Instead, as section 4 explains, we must develop a capacity that is reasons-responsive, fast, and crucially, intelligible from the personal-level perspective. I call this capacity, whatever it may be, our *sense of import*

and I argue that in human agents, the sense of import consists of systematic dispositions to undergo quick affective and attentional episodes in light of our practical circumstances. Sections 5 and 6 defend the second claim of this chapter, that the pattern of dispositions that constitutes the sense of import must be *synchronically integrated* and *diachronically stable* in specific senses, which I articulate and defend. Lastly, section 7 argues that the sense of import is similar to—but also distinct from—full blown cares and concerns in ways that merit regarding it as a genuine, yet minimal mode of mattering. If you will, there is a wide region in our affective, motivational, and valuational lives between our cares proper and our mere behavioral impulses. This region is inhabited by our sense of import.

2. Two Problems

Walking into your kitchen, you lay your eyes on a delicious-looking slice of cake, enticingly placed on the counter. What do you do? Some courses of action come to mind: you may, for instance, reach out to grasp the cake, you may step towards your refrigerator door with the purpose of opening it and putting the cake in the fridge for later, you may ignore the cake and keep looking for your keys, and so on. No matter how you end up acting, clearly you must do *something*. Call this *the problem of action selection*.

In general, in any practical context, organisms are faced with an indefinite number of actions that they may perform (or abstain from performing). No practical context uniquely determines a single course of action, and hence the problem. In Wu's construal, an agent must choose a mapping between one member of the set of perceptual inputs she is being presented with and a potential action (Wu, 2011, 2023). It is up to the agent to settle on a specific action type that she is about to perform in response to her situation, on pain of inaction, atrophy, and—if I may be dramatic—death. This is the motivating thought behind illustrations such as Buridan's ass: torn

between two identical, equidistant bales of hay, the ass has no ground to favor one bale of hay over the other, and ultimately perishes of hunger.

To save ourselves from this grim fate, we must order our motivational landscape in a way that favors one action over all other possibilities available to us in a context. This means that, absent external assistance, we must develop a minimally unified structure for our agency. We must respond to the "raw necessity of eliminating conflict among [our] various motives" (Korsgaard, 1989: 110) by ordering our wills such that we settle on a specific purpose to pursue, and on a specific action to perform in pursuit of that purpose. We must corral different motives, desires, and inclinations and ensure that we aim at one of those (or a set of compatible motives) rather than another or none at all. Selecting an action type to perform is, *ipso facto*, unifying our wills as agents. Action selection, then, requires what I shall call, a *unity of motives*.[9]

Another kind of problem that faces all agents requires a second kind of unity. Suppose you manage to settle on a specific action to perform with respect to the cake that, alas, still haunts your field of vision. Imagine you succumb to temptation and decide to grab the slice of cake and eat it before anybody finds out. How are you to do this? you can reach for it with your left hand or your right hand, you may do so slowly or quickly, with intensity and vigor or with attention and care. Generally speaking, actions of the same type may be performed with, among other things, diverging force, speed, and bodily organs. Once an organism successfully picks out a single action type to perform, it must figure out a specific *way* of preforming it. It must specify the relevant

[9] Talk about "the unity of agency" is sometimes ambiguous between a unity among our *goals* or *ends* on the one hand, and a unity among our *intentions* or *motives* to perform some actions in pursuit of those goals on the other. This is understandable since settling on an end rationally necessitates settling on an action to perform in order to achieve this end. But often our ends do not uniquely favor an action to perform for the sake of their attainment. To reduce this ambiguity, I name the specific agential unity that is required to solve the selection problem "the unity of motives". But for my purposes here, "motive" should not be heard as a conscious, or deliberative, or even necessarily propositional attitude. Following Wu (2023, esp. ch. 3), I use "intention" and "motive" interchangeably and I take them to denote an agent's representation of a solution to the selection problem that serves to influence or bias the agent's action selection process (cf. Burnston, 2017).

parameters of its movement, and it must do so continuously, across the duration of the action. In every single moment wherein an organism acts, it has multiple trajectories of movement available to it that could constitute the action chosen for performance by the organism. This is the problem of *action specification*: it is not enough that we choose what action type to perform, we must, by necessity, decide *how* to perform it.

To see the significance of this problem, note that Buridan's ass wouldn't be salvaged from its predicament even if it were to choose to approach one bale of hay rather than the other. It would still have to decide which *course* its action will take. You cannot just approach a bale of hay *simpliciter,* but in some specific, particular way. Success in selecting an action to perform but failure in choosing the specific movements that shall constitute that action still leads to a dead end—and a dead donkey. Action (please excuse me) is one big ass problem after another.

Hence, the problem of action specification presents a second—and often neglected sense—in which our agency must be unified: when we act, we must rally our wills behind a single motive, but we must also settle on a single, unified way of executing the prescriptions of this motive. In addition to a minimal unity of motives, action requires a unity among what agents take as the potential ways of *implementing* their motives. Agents—or something about them—must determine the identity of the specific movements that are to constitute the actions they perform towards the realization of their intentions. Call this the *unity of manners*. Action requires at a minimum that our agency be doubly unified: that we enjoy a unity of motives and of manners.

3. Relegation and Reflection

How can a creature develop this twofold agential unity? How can agents settle on a specific action to perform, and on a specific way of performing it? How can they organize their

motivational makeup to exhibit both a minimal measure of consistency and coherence with respect to motives, and with respect to the manner with which their motives are to be realized?

3.1. Relegation

One quick and easy way of solving these problems is by relegating them to sub-agential mechanisms (Dennett, 1969; see Drayson, 2012, 2014 for excellent discussions of the personal/sub-personal distinction). Instead of treating action selection and specification as problems to be solved by agents-as-wholes, these problems can be worked out by proper parts of those agents. Agents may be primed to act according to highly specific, preestablished motor programs in a way that is "rapid, immediate, and non-reflective in nature" (Ridderinkhof et al., 2011: 179) by specific environmental cues. One potential benefit of such a strategy is that it allows action selection and specification to be settled *simultaneously*, instead of serially by first deciding what to do and then settling on a specific way to do it. Indeed, human actions can be prepotentiated in response to perceptual stimuli without being mediated by any conscious intentions on the part of the perceiving agent. An influential fMRI study by Chao and Martin (2000), for example, found that merely attending to pictures of graspable tools can elicit strong activations in the premotor and posterior parietal cortices, which are associated with hand movements such as grasping, and similar results were obtained by Proverbio et al., in an event related potential (ERP) study (2011. See also Grèzes et al., 2003; Grèzes & Decety, 2002; Tucker & Ellis, 1998, 2004).

The idea—or the caricature of the idea—is that, when agents are presented with a choice situation, most if not all of the work is already done for them. Certain environmental cues initiate highly specific motor programs, with a predetermined range of action parameters in a way that obviates the agent having to decide for herself. If a prespecified program is potentiated in a strong enough manner, it gets executed and both problems—of action selection and specification—are resolved in one swoop. Notice that this story allows for highly—perhaps maximally—unified

agency in the sense that is relevant to our purposes. A creature that it applies to simply *cannot* have a divided will: the environment plus some contingencies of its nervous system uniquely settle what such a creature would be doing and how.

Now, while this is surely a part of the story regarding human behavior (and perhaps the whole of it for some of our less sophisticated fellow creatures), it is unsatisfactory as a story about human *action*. To put it bluntly, by treating action selection and specification as problems that are to be solved exclusively by sub-personal mechanisms, outside the purview of agent-level processes, such a story assimilates all actions to mere behaviors, on a par with breathing, blinking, and blushing. Consider, as an example of an intelligent, yet markedly sub-personal process, the dilation of a person's pupils. As Frankfurt contends, this process is *guided* in the sense that "there are mechanisms which guide its course. But the occurrence of this movement does not mark the performance of an action by the person; his pupils dilate, but he does not dilate them" (Frankfurt, 1988: 159). In the present vocabulary, a specific case of pupil dilation is perfectly intelligible as a solution to selection and specification problems: roughly, what is to happen—and how—given changes in lighting conditions. But such a process is not intelligible as an *action* precisely because actions are personal-level phenomena—they are agential achievements if ever there were any. Since those achievements must involve the working out of solutions to specification and selection problems, there must be some agent-level states and processes—states and processes that are attributable to agents-as-wholes—that are responsible for the two kinds of unity we have identified.

This might sound too fast. Can we not think that at least a *part* of the achievement is due to sub-agential mechanisms? For instance, perhaps a solution to the selection problem must be arrived at by some personal-level state or process for a piece of behavior to count as an action, yet,

after such a solution is reached, *action specification* can be fully relegated to sub-agential mechanisms that determine the manner in which the action unfolds over time. I do not deny that many action parameters are determined by such mechanisms (nobody, to the best of my knowledge, denies *that*), but I do wish to insist that at least a good portion of action specification on-the-fly must be done by agent-level processes. That is, both selection and specification must be attributable in part to subject-level phenomena.

To see why, think about the difference between some sub-personal and personal-level effects of our intentions. Consider, for example, intentionally preparing a response to a perceived affront. Assume you form an intention to confront a rude person, who regularly blocks your elderly neighbor's accessible parking space. This intention might lead you to concentrate your attention on the perpetrator, walk towards them, and prepare certain verbal behaviors, all of which are properly attributable to *you*. On the other hand, this intention might have causal effects on your sympathetic nervous system, which may result in narrowing blood vessels, increased heart rate, and changes to your breathing patterns. While being caused by your intention, these latter responses, as opposed to the former, are not your *actions*—indeed, they are paradigmatic non-actions.

Thinking about action specification processes as fully sub-personal processes that take intentions as their mere causal antecedents (as the story on offer recommends) amounts to treating them as analogous to patently non-agential happenings. But this amounts to treating *actions* as non-agential happenings. To be properly agential, specification processes must be at least partially responsive to the rational, psychological makeup of their subjects. As Fridland writes, "fine-grained, automatic motor processes instantiated in motor skills are not simply causally connected to intentional states but, rather, continue to be semantically sensitive and responsive to personal-

level goals throughout execution" (Fridland, 2017: 1557; see also, e.g., Christensen et al., 2016; Pacherie & Mylopoulos, 2021).[10]

3.2. Reflection

Luckily, there are, presumably, other ways of unifying our agency. One such way goes, in a sense, in the opposite direction to the previously considered strategy, by harnessing the full extent of what many view as humans' most definitive agential capacities: our reflective and deliberative powers. Taking a step back from our situation and evaluating it allows agents to structure their agency in a way that is reasons-responsive, and that ensures that their responses to practical contexts are in line with their goals, motives, and values—indeed, with their *personhood*. As Korsgaard says,

> When you deliberate about what to do and then do it, what you are doing is organizing your appetite, reason, and spirit, into the unified system that yields an action that can be attributed to you as a person. Deliberative action pulls the parts of the soul together into a unified system. (2008: 119)

However, as we are well-aware, rational reflection and practical deliberation are slow, expensive, and dishearteningly serial. Our finite human resources make it utterly unrealistic to stop and deliberate about all our actions at the time of performance. In Bratman's words, "we need ways to allow deliberation and rational reflection to influence action beyond the present" (Bratman, 1987: 2). Famously, Bratman suggests that plans are the means by which we respond to this need. In forming plans through rational deliberation, we settle on courses of action without the need to

[10] Importantly, staunch anti-intellectualists too do not deny that even the most fluent, non-reflective cases of expert action are guided by personal-level phenomena. Dreyfus, for instance, approvingly cites Merlau-Ponty, who writes: "For the [soccer] player in action the soccer field is not an "object." It is pervaded by lines of force … the player becomes one with it … At this moment consciousness is nothing but the dialectic of milieu and action" (2002: 168-9). Dreyfus suggests that the consciousness of the skilled player—as well as our own consciousness, when we skillfully cope with our environment—amounts to an "absorption in a familiar field of relevant affordances" (2013: 17). Dreyfus's anti-intellectualism, then, is motivated precisely by phenomenological analysis of our conscious (and personal-level) structures.

stand back and reflect on these courses of action from moment to moment. Through deliberation and practical reasoning at a time,

> we settle on complex—and, typically, partial and hierarchically structured—future-directed plans of action, and these play basic roles in support of the organization and coordination of our activities over time … plans introduce characteristic forms of practical reasoning: plans provide a somewhat stable background framework that needs to be filled in appropriately with specifications of means and the like. (Bratman, 2007: 26)

Planning allows us to maintain agential unity over time and across extended action sequences, in a way that does not require us to determine in advance all the particularities of our action. At any moment, our capacity for reflection can kick in and help us "step back, assess our motivation, and take a stand" (Bratman, 2007:28). On this model, then, the problems of action selection and specification may be solved partly thanks to prior planning, and partly thanks to reflective acts at the time of action.

Just as with the previous strategy, surely deliberation must be a part of the story about human agency, but it cannot be the whole of it. An attempt to reduce the processes by which we unify our agency into reflective planning and deliberation fails to appreciate both the levels of detail that must go into action specification, *and* the kinds of evaluation and assessment that agents must engage in during an action's performance. That is, first, reflection cannot determine all the minute movements that constitute our actions; second, action specification requires sensitivity to reasons and rapid response to those reasons in a way that bypasses the need to deliberate.

Notably, I take neither Bratman nor Korsgaard to be committed to the idea that deliberation is our *exclusive* way of solving the selection and specification problems. Both of them seem to me to appreciate the fact that, while deliberation might be necessary for the kind of agency that humans enjoy, it is by no means sufficient for it. Thinking of Buridan's ass, Bratman, for example, revealingly says: "I conjecture that we have an ability to decide in the face of equidesirability" (1987, 11). I am interested in fleshing out this conjecture. *What is it* about humans that allows us

to select and specify our actions in rapid, intelligent ways that take account of the deliverances of prior deliberation and planning?

To drive home the point regarding the insufficieny of both deliberative agency and sub-personal mechanisms for action selection and specification, think about nonverbal dyadic interactions, for example. You are attending a workshop in a foreign country. At the reception, one of the workshop's participants approaches you. You make an initial movement towards offering them your hand, and they lean forward, turning their face to the right for a kiss on the cheek. Both of you have a split-second to avoid an awkward mismatch between your greeting behaviors, and engaging in deliberative activity is out of the question. Your stock motor program for greeting situations is failing you, and deliberation is too slow to pick up the reigns. Let us assume that you manage to turn your face to the left in an unacknowledged feat of social agility, sparing both you and your colleague the impending embarrassment. There must have been *something* about your psychological makeup that enabled you to react to the normative pressures of that situation in a non-deliberative, yet intelligent manner. What could that be?

4. The Sense of Import

Action requires solving both the selection and the specification problems by unifying both our intentions to act, and the specific movements we elect to realize those intentions. One way of solving these problems is by having an inventory of motor programs that get automatically triggered by situational cues, but exclusive reliance on those would render our behaviors non-agential. Deliberation might help in a wide range of cases, but it is too slow and expensive to be deployed in many contexts. We need a capacity that is both (defeasibly) reasons-responsive and able to provide quick solutions to the selection and specification problems.

Moreover, this capacity ought to be reasons-responsive in quite a robust sense. Whatever features this capacity takes as its inputs, it must compare, contrast, and evaluate those features in some kind of principled way. It must employ a set of stable standards to assess which of these features should influence our actions and to what degree. Settling on a course of action that includes turning your cheek for a collegial kiss requires that you take in some factors as important and others as irrelevant. Your colleague's bodily movement, your own posture, and the price of violating a social norm are highly important in the context we are considering, whereas the conversations in the background, the position of the coffee urn, and your doubts about the previous talk's thesis are beside the point at that moment. You need to sort the normative wheat from the chaff and this requires the ability to gauge how relevant considerations hang together to recommend a specific action. The price of violating a social norm means you must change your behavior, your own bodily posture makes it impossible to take a step back, your colleague's approaching cheek means you must turn your face to the other side so as not to accidentally French them, and so on.

Let us introduce some terminology. Following Taylor (especially the essays in part I of his 1985. See also Helm, 2001: 32), I will use the term "import" to designate the property some things have of *mattering* to a subject. The things that carry import for us are the things that we find significant or relevant, the matters to which we are non-indifferent. Things of great import may lead us to feel excitement and exhilaration, guilt and glee, pleasure and pain, and are the objects of our cares and concerns.[11] Now, presumably, in every situation, different things may carry import

[11] To be clear, I do not intend to suggest that objects of worth have an ontologically distinct property, *import*, that stands over and above other properties in virtue of which they are of concern to us. "Import" is a general axiological term that I will employ to capture the vast variety of ways in which some things matter to some subjects. Certain things may carry import for us because they are *dangerous* while others may do so because they are the objects of our love.

for an agent. In every situation, different considerations will be of worth to her. Let us designate the set of all imports to a subject that a situation encapsulates as that situation's *landscape of import*.

In these terms, we can say that our capacity to act requires that we shall (i) be sensitive to our landscape of import at a given situation, and (ii) be able to recognize how our landscape of import motivates a specific action by assessing how the things that matter to us—the bearers of import—will be affected by our actions. In other words, how our putative actions will influence our landscape of import. Call this capacity our *sense of import.* Human action necessitates having a sense of import.

What constitutes the sense of import in creatures like us? I claim that it is our dispositions for undergoing systematic patterns of attentional and emotional episodes in light of our situations.[12] It is a near consensus in the scientific and philosophical literatures on affect that emotional processes involve appraisals or evaluations.[13] Affects are widely believed either to result from "the appraisal of events with respect to their implications for well-being or for the satisfaction of goals motives, or concerns" (Frijda, 1993: 357-8),[14] or to incorporate as part of their components some evaluations of an organism's situational factors.[15] These evaluations are not isolated.

[12] Just as with deliberation, the idea that having a sense of import is a necessity for action in creatures like us is not to be confused with the claim that it a sufficient condition for human action. I see no reason why the sense of import, in the general sense that I am developing here, should be a uniquely human phenomenon (although human senses of import might be importantly different from non-human senses of import, for instance because of the way they interact with our more distinctive, reflective capacities. I touch on this possibility in the next section as well as, more fully, in a work in progress).

[13] To be clear, this does not mean that, necessarily, emotions or a specific component thereof such as valence just *are* appraisals.

[14] Note that according to some theories of emotion, the relation between appraisal and emotion is not just one of necessary causal antecedence. Emotions are often thought to involve the appraisal of situations as parts of their constituents—perhaps even their essential constituent. This is the thought behind many variations of *appraisal theories* of emotion (see, e.g., Ellsworth, 1991; Ellsworth & Scherer, 2003; Frijda, 2009; Lazarus, 1991; Moors et al., 2013. One philosophical analysis of emotions that draws heavily on appraisal theories is Nussbaum, 2003) . I do not take a stand on this issue here.

[15] This thought is by no means limited to so-called appraisal theories. For instance, Ekman (1992) suggests that emotions involve automatic appraisal processes; Tappolet argues that emotions are perceptions of evaluative properties (Tappolet, 2016); Goldie suggests that emotions are "evaluative feelings" (Goldie, 2000); Helm defends a view of emotions as "felt evaluations" (Helm, 2001); Prinz argues that emotions are "embodied appraisals" (Prinz, 2003); and even theorists such as Colombetti, who shun some of the arguably conceptual, inferential, or computational

Emotional episodes are widely considered to be constituted by, minimally, two dimensions: valence and arousal. The former is, approximately, the felt "pleasantness" or "unpleasantness" associated with an emotional episode, and the latter is—again, approximately—"[t]he subjective level of alertness, activation or energy elicited by an experience" (Samanez-Larkin & Knutson, 2015: 281). Now, valence signals are thought to be mediated by a single neurophysiological network, which can serve the role of "common currency" in processes of rapid decision making (Levy & Glimcher, 2012). As Carruthers summarizes: "Valence produced by many different properties of a thing or event can be summed and subtracted to produce an overall evaluative response, and such responses can be compared to enable us to choose among options that would otherwise appear incommensurable" (Carruthers, 2018). This means that emotional responses can be sensitive to our landscape of import and integrate information from diverse features of our situation to motivate fast, flexible responses. Rapid, emotional episodes, then, can be reasons-responsive yet fast enough to satisfy the requirements posed for our sense of import.

Notice that I said that our sense of import is constituted by our disposition to undergo *emotional episodes* or *reactions* rather than to feel *emotions*, a distinction I borrow from Goldie (2000). An emotional episode "is a particular pattern of thoughts, feelings, bodily states, involuntary facial expressions, predispositions to act, perceptual selectivity, direction of attention and imagination, etc. which amounts to a person being, for example, angry, or fearful, or ecstatic at a particular time" (Jaworska, 2007b: 552). In contrast, a full-blown emotion is an enduring state that consists, in part but not entirely, in emotional episodes and dispositions to experience them. It includes "various past episodes of emotional experience, as well as various sorts of disposition to

connotations of the notion of "appraisal" agree that emotions are essential parts of an organisms' process of sense-making, which is the "evaluation of [a] systems surrounding," whereby its surroundings acquires meaning and get transformed "into an *umwelt*" (Colombetti, 2017: 447).

think, feel, and act, all of which can dynamically interweave and interact" (Goldie, 2000: 11). Our sense of import is the set of our dispositions to undergo attentive and emotional episodes rather than full-blown emotions.

I have said that our sense of import must be sensitive to our landscape of import in a way that enables us to gauge how potential actions will shape and influence it, and I claimed that this is a necessary condition for our agential unity: it is a requirement that arises in the face of our need to unify our wills behind a single intention or motive, and behind a single set of movements that we settle on as constituents of the action that characterizes the intention's content. In humans, the sense of import is constituted by our dispositions to undergo rapid emotional and attentional episodes in a systematic way. What does this systematicity amount to?

5. Synchronic Integration

We can elaborate on the putative way in which our sense of import must be systematic by highlighting two features that, I argue, our landscape of import must exhibit: *synchronic integration* and *diachronic stability*. In this section, I explain the former and then, in the next section, I clarify the latter and argue that it is a necessary precondition for the development of full-blown cares.

Our dispositions to undergo certain emotional and attentional episodes must be synchronically integrated. That is, in usual circumstances, it is not enough that we are disposed to undergo *one* emotional episode at a time. We must be capable of having at least a few emotional reactions to our situations, and all of these must participate in the unification of our motives and manners in a principled way. It is easy to see why when we ask what emotional reactions react *to*. What are their eliciting circumstances? Theorists are in wide agreement that specific emotional episodes are typically triggered by what an agent takes (e.g., believes, imagines, perceives) to be

specific agent-environment features that carry a certain significance—in our terminology, import—to her. Typically, for instance, fear reactions are responses to (real or imagined) dangers, anger reactions are responses to (real or imagined) affronts, compassion responses are triggered by another's (real or imagined) suffering, and so on.[16]

In other words, types of emotional episodes are elicited by corresponding types of considerations in favor or against different actions and different ways of performing those actions. That is, different affective responses are responses to different *reasons* for and against different actions. Now, being capable of undergoing one affective reaction at a time means being able to respond affectively to one reason (or at least, to a highly circumscribed set of reasons) at a time. But such a limited emotional capacity renders impossible the necessary intelligence in our reasons-responsiveness which we have identified earlier. For an agent to maintain flexibility in a performance, she must be sensitive to a *variety* of reasons: for example, for altering, aborting, or proceeding with her action's current trajectory. There are many imports – some of those conflicting – that bear on your split-second decision to turn your cheek for a collegial kiss. The prospect of being taken as engaging in inappropriate behavior might favor taking a step back; the norms of collegiality might favor staying put; and the cost of accidentally doing something even more awkward (what if you overcorrect and kiss your colleague's neck?) might favor keeping your original trajectory, despite the cost of mild embarrassment. I'm not arguing that you are sensitive to all these imports, but I am trying to illustrate that, to be minimally flexible, you must be sensitive to *some of them*. You must be disposed to undergo more than one emotional reaction at a time.

[16] Let me stress: although many take these circumstances, or what Lazarus calls "core relational themes" to be the intentional objects of affective reactions, this is not what I have in mind. As I said, we can be neutral with respect to the representational content of an emotion, and still acknowledge that it is triggered by a specific kind of import.

Moreover, the diverse emotional reactions that you must be disposed to undergoing should be *integrated* or *ordered* according to what you ultimately take to bear greater or lesser import in your situation.[17] When your emotional reactions pull you in different directions with equal motivational forces, you find yourself in Buridan's ass' position, saddled by equidesirable action paths. If our sense of import is to solve the selection and specification problems, it must consist in motivationally proportional responses to a landscape of import that is not uniform. Some of our affective motivational reactions to things that bear import for us must move us to act in more powerful ways than others.[18] This is what I mean by saying that your sense of import—your pattern of dispositions to attend and emote in response to your landscape of import—must be synchronically integrated.[19]

6. Diachronic Stability and Caring

This and the following sections defend two, final claims. First, in addition to its synchronic integration, our sense of import must be *diachronically stable*. Our dispositions of emoting and attending in response to our situations must be regular in the sense that similar situations will elicit similar senses of import. Having radically different emotional and attentional responses to similar practical situations leads to a specific kind of psychological deficit: a failure to form full blown

[17] I take Railton to make a similar point when he argues that our emotional reactions operate as a kind of "summary outputs" of our evaluative system. As he says: "The broad affective system might be nature's first version of the sort of representation-mediated, multimodal, unified prospective guidance of thought and action once thought to be peculiar to conscious deliberation and choice" (Railton, 2017: 340)

[18] Notice that a motivational target can be recommended by more than *one* emotional reaction. The same situation can elicit both fear and anger responses, for instance, and both of these reactions can motivate a fight response. This is entirely consistent with what I am saying. I think such motivational aggregation is a highly prevalent phenomenon that plays a crucial role in ensuring that our sense of import is synchronically integrated (for more on motivational aggregation and integration, consult (Fulkerson, 2023, and n.d.)

[19] It is important to emphasize that I do *not* mean here that all our emotional reactions must be consonant with each other. Our affective-motivational sets may and do push and pull us in different directions. What is crucial is that such dissonances are *resolved* in some minimal sense. That conflicting motives are integrated, rank-ordered, suppressed, and so on, such that our sense of import, as a unified phenomenological structure, gives us some minimal inkling regarding what to do and how. I thank Matt Fulkerson for requests for clarification on this point.

cares. Second, despite its affinities to *bona fide* caring attitudes, the sense of import is crucially different, and, in a sense, it is a more *minimal mode of mattering* than cares proper.

Let us start with a few diagnostic remarks about caring attitudes. First, according to an influential class of accounts of cares (Jaworska, 2007a, 2007b; Jaworska & Wonderly, 2017; Seidman, 2009, 2016; Shoemaker, 2003; cf. Frankfurt, 1998, 1999; Helm, 2001), such attitudes consist in a structured complex of first and foremost emotional—but also more purely cognitive, and conative—dispositions that are centered around *a specific object*.[20] Caring about an object is a matter of manifesting a kind of emotional vulnerability (Jaworska & Wonderly, 2017) to the fate of this object. Typically, for instance, it involves feelings of happiness when the object of our care fares well and sadness when it is harmed; hope that it will flourish and anxiety in the face of a potential threat to it; as well as motivations to act on behalf of the cared-for object and, often, positive evaluative judgements regarding it.

Second, cares seem to occupy a pivotal role in configuring a person's moral and practical identity. Frankfurt, for instance, identifies what a person cares about with whatever it is that their life is *devoted* to (1998). In the same vein, Seidman identifies our cares with our valuing attitudes, and he suggests that the objects of our cares and concerns are important for our well-being as agents (2009:285). Moreover, Jaworska claims that cares have a unique moral significance for agents for a few reasons, one of which being that desires that are based on cares have a much tighter connection to an agents' foundational interests than simple appetites, and they endow her with a unique moral standing, which is not shared by creatures incapable of caring. Hence, frustrating a person's care-based desire is a much more severe harm against them than frustrating a mere appetite (2007b: 533).

[20] Helm explicitly rejects the cognitive-conative distinction, but on his view, the attitudes that are involved in caring have two directions of fit.

Third, according to Helm's account of caring, which is endorsed to a large degree by Jaworska (2007a, 2007b), caring involves a "projectible, rational pattern of felt evaluations and evaluative judgements, such that one is motivated not only to feel and judge, but also to act accordingly" (Helm, 2009: 75). The *rationality* of this pattern consists, among other things, in the fact that the emotional dispositions constitutive of caring are related to each other through specific rational coherence requirements. Helm mentions two such requirements: first, "tonal commitments" connect opposingly-valenced affective states. For example, if I genuinely care about my dog, then the contentment I feel when I see her lounging peacefully on the couch commits me to feeling nervous upon learning that she stole a chocolate muffin from the kitchen counter. Second, "transitional commitments" connect forward- and backward-looking emotions. My fear that something bad will happen to my dog after she snatched said chocolate muffin rationally commits me to, say, feeling relieved when I manage to get it out of her mouth. Generally, if I am afraid in the face of danger to the cared-for object, then a failure to feel relief when the danger has passed is a failure of rationality, according to Helm. Too many such failures will corrupt the emotional pattern to such an extent that it will no longer be intelligible as a genuine instance of caring.[21]

Lastly, it is almost a consensus that our caring attitudes must involve, minimally, attentional and emotional dispositions. For instance, according to Seidman (2016), cares encompass dispositions to be guided by *reasons,* by considerations pertaining to the cared-for object. However, any given situation confronts the agent with a vast – perhaps indefinite – number of potential reasons to be guided by. "The world is full of valuable persons, goals, projects, and so on that a person could, rationally unexceptionably, care about" (2796). The objects of our cares

[21] but see Seidman (2016) for a critique of the idea that the emotions involved in caring impose rational requirements on their subjects.

are privileged, in part, in the sense that we come to allow reasons pertaining to them to guide us. Consider an extension of an example by Seidman's: Mia can be guided by considerations that pertain to her Ming vase, to her Afghan rug, to her cactus, etc... The fact that she cares about her vase and not about these other objects means that she has come to be moved by the weals and woes that befall her vase rather than her rug or her cactus. The story about what has led her to care about the vase specifically "is a story about how the vase has captured her *attention*, so that she notices, and responds to, the reasons for joy, fear, relief, and so on that *it* gives her, rather than the analogous reasons that some no less beautiful work of art, or no less valuable object of another kind, gives to her" (Seidman, 2016: 2797, emphases original).

As we have just seen, cares (to stress their difference from the sense of import, let us call such modes of mattering *cares proper*) are clusters of affective, cognitive, and conative dispositions that are referentially unified in that they have a single intentional object. These clusters are an important component of human agents' lives insofar as they guide our actions across multiple contexts and over time.[22] Why think that a diachronically stable sense of import is a necessary precondition for the kind of systematic, rationally integrated, emotional vulnerability that characterizes our cares proper? To answer the question, it is useful to imagine what an agent who lacks a diachronically stable sense of import with respect to a given practical domain would be like. That is, an agent who undergoes radically different (albeit, perhaps synchronically integrated) action-guiding emotional and attentional reactions every time she finds herself in a certain practical domain.

[22] While I do not take this line of thinking as necessarily hostile to situationist approaches to moral psychology (Doris, 2002; Miller, 2013), I do not wish to belabor this point here, especially because I am quite convinced by recent critiques of situationism. Westra, for example, forcefully argues that situationists tend to have in mind an inappropriate concept of context-stable traits, and, perhaps more importantly, that much of the empirical results that support situationist conclusions are highly ecologically suspect (see Westra, 2020, 2022).

Consider, for example, a person—call her Dina—who experiences such diverging reactions whenever, say, she goes out with her colleagues for a drink. One time, Dina's attention is completely captured by her conversationalists, on another occasion, she is mesmerized by the bartender, and on a different outing, she cannot take her eyes off the bar's other patrons. Similarly, in one case, her friends' attempts at small talk may irritate her, another time, the exact same remarks might spark intense involvement, at other times still, their mundane chatter could trigger feelings of despair at her life, which is slowly wasting away, and at different instances, she may be completely paralyzed by feelings of social anxiety.

Given what we know about her, can we say that diverging Dina *cares* in the full, previously explicated, sense of the term about any of the potential intentional objects of her diverse attentional and emotional attitudes? I am tempted to say no. For Dina to *care proper* about the bartender, or the other patrons, or her colleagues, or, for that matter, her slowly diminishing time on this earth, her emotional and attentional reactions must be at least minimally stable. She should take interest *consistently* in these objects, and her interest should manifest in her patterns of emotion and attention. But by our stipulation, she does *not* have such stable patterns. She is incapable of *taking interest* in a consistent manner. That is, she is incapable of finding the same things as bearers of import in the same practical domain.

Am I making mountains out of molehills? We all have our good days and our bad days, our careless moments and our catastrophizing hours, our breaking points and our euphoric periods. Is Dina really that different? It depends. What entitles us to say that we *care* about, say, our ideals, or pets, or loved ones in the face of the tremendous variation that characterizes our emotional lives is precisely our consistency. Even if I am occasionally inattentive to, or irritated by my partner, the fact that I exhibit a *generally stable*, predictable pattern of affective, cognitive, and conative

reactions to her is what entitles me to claim that I truly care about her. Too much inconsistency corrupts the pattern into something unidentifiable as a caring attitude proper.

I conjecture that what makes us think that Dina might still care about, say, her colleagues, is that we envision her as having a more-or-less stable pattern of emotional and attentional reactions towards them in contexts that *do not involve* going out for drinks. Dina may be unpredictable around bars, but perhaps she *does* exhibit consistent emotional and attentional reactions in other contexts: in the office, or when she carpools with her friends to work, or at company parties. If this is the case, then Dina might very well care, in the full blown, *bona fide* sense about her colleagues, but this highlights two crucial points.

First, Dina's case shows that, if an agent's capacity to care is to remain intact, a breakdown in consistency in one domain of action must occur against a background of diachronic stability in a majority of relevant domains. *Even if* our sense of import in one—or few—domains of action is unstable, our capacity to care proper is premised on an overall relative diachronic stability in our emotional and attentional reactions in most types of practical context.

Second, while Dina might be said to care about her colleagues in that case, I would argue that her emotional and attentional instability in a specific type of practical context *casts a negative light* on her caring attitudes towards her friends in general. Dina's inability to display the relevant reactions specifically when she goes out for drinks is *rationally criticizable*, and it is criticizable precisely because it does not cohere with the rest of her psychological demeanor. If Dina *is* generally consistent in her emotional and attentional responses to her colleagues, we would like to know what is it about bars in particular that messes with her capacity to display the appropriate reactions. Indeed, it makes sense for her friends to question the depth of her care about them if she is unable to find them important whenever she enters a bar.

In conclusion, our capacity to care seems to be predicated on a background of relative diachronic stability in our sense of import. While our caring attitudes can endure some instability in a limited range of practical contexts, too much instability destroys the pattern. Moreover, even instability that is confined to a particular domain of action seems to reflect negatively on the pattern as a whole.

7. Caring and the Sense of Import

Harry Frankfurt says:

> [T]he notion of what a person cares about … coincides in part with the notion of something with reference to which the person guides himself in what he does with his life and in his conduct. It is not to be presumed, of course, that whenever a person's life displays over a period of time some more or less stable attitudinal or behavioral disposition, this reflects what the person cares about during that time. After all, patterns of interest or of response may be manifestations only of habits or of involuntary regularities of some other kind; and it is also possible for them to develop merely by chance. (1998: 82)

Before concluding the chapter, I wish to offer a friendly addition to this remark. Some attitudinal dispositions may reflect what a person finds important in relevant circumstances, yet fall short being of cares proper. Moreover, these patterns of interest reflect what we find important in a situation *despite the fact* that such patterns of interest and response may be manifestations of habits or involuntary regularities, or that they develop in large part—if not "merely"—by chance. Senses of import are precisely such attitudinal dispositions. There is a vast region in our affective-motivational topography between mere impulses to act and full-blown cares, and this region is occupied by the sense of import. To bring this into sharper relief, I want to consider the ways in which our senses of import do not constitute cares proper.

First, recall that it is almost a consensus in the caring literature to view cares as somehow deeply connected to internality. It is common to hear claims to the effect that cares proper delineate what a person's life is devoted to, what she truly values, or that they are among the kinds of psychological attitudes that are most truly *her own*. But these features don't seem to hold

categorically in cases of considerations that our senses of import are sensitive to. I may be an aggressive driver, and my behavior might be guided by a diachronically stable and synchronically integrated pattern of emotional and attentional dispositions. I may, for instance, undergo an anger episode whenever somebody cuts me off on the road. But this does not mean that my life is *devoted to* preventing these occurrences from happening. Similarly, it is unclear that I must *value* whatever it is that is slighted when I am cut off (maybe it's my all-too-fragile masculine ego). And it certainly does not seem true that by frustrating my desire to honk at the car in front of me for not driving immediately after the light has changed to green, you would be frustrating a deeply internal desire of mine. In fact, you might be doing me a favor precisely because I am alienated from my sense of import with respect to driving.

Second, while caring proper is constitutively a complex of primarily emotional dispositions, it also involves paradigmatically cognitive dispositions. In humans, caring attitudes typically include dispositions to *believe* certain things about their objects. Specifically, according to Seidman (2009; see also Jaworska, 2007b, fn. 98), caring often involves dispositions to believe that considerations pertaining to the object of her care are practical reasons for her. In Seidman's example, there is something deeply confused about an agent who "claims to care, say, about another person, and yet does not see the fact that this person is in danger as any reason to try to help her, does not see the fact that this person is in distress as any reason to try to comfort her, and so on" (288). Caring disposes us to believe that, for instance, the woes that inflict the objects of our cares are practical reasons to rush to their aid. This, however, does not seem to be the case with our sense of import. While I might be disposed to feel anger after being cut-off, I may not be disposed to *believe* that being cut off is a practical reason for me to act in any specific way.[23]

[23] This might be a bit more complicated. We can plausibly assume that often, when I am being cut off by someone on the road I speed up and try to catch them and cut them back. We can also imagine that, if an alarmed passenger were

Third, our habitual senses of import do not depend for their existence on the satisfaction of rational coherence requirements of the sort Helm identifies as crucial for caring attitudes. Recall that, according to Helm, for a pattern of emotional dispositions to be intelligible as a caring attitude proper the different emotions that it comprises must be related to each other in certain systematic ways. I must be worried when it is in danger, relief when it is out of harm's way, sadness when misfortune befalls it, and joy when it is thriving, and I must experience these emotions in ways that are more of less proportional to each other. Too many fissures in these rational coherence links between my emotions will corrupt the pattern and render it unintelligible as a care proper.

This is not the case with our sense of import. To use the previous example, I might find myself in the grips of an anger episode whenever I drive without this state being mitigated by any specific event that would count as an appeasement to an offense, and I might find my anger rapidly subsiding without me doing anything in particular to amend any infraction. This does not mean that my sense of import with respect to driving cannot be assessed and criticized according to its rationality—you may justifiably think that my anger is irrational. My point is that while cares proper are not intelligible as such without the satisfaction of sufficient rational coherence requirements, this is not the case for our senses of import. It makes perfect sense to think that our patterns of attending and affectively responding to similar contexts can be consistent, and quite irrational. In simple terms, it is intelligible to think that, say, being cut off on the road *matters* to me, at least in the minimal sense of consistently capturing my attention, hijacking some my

to ask me what I was doing I would have "snapped out" of my enraged attempt to chase down the perpetrator. In such a case, it is not unreasonable to think that I would say something like "I'm sorry! I was so angry with that asshole!" It is plausible to interpret such a comment as indicating that I take the perpetrator's behavior to be a *motivating* reason for my conduct. But we should not interpret it as suggesting that I am trying to *justify* my behavior. Yet this is precisely the way in which Seidman understands the doxastic dispositions that are involved in caring. They are dispositions to take considerations that pertain to the objects of cares as *justifying* practical reasons (see his 2009, fn. 4).

emotions, and guiding some of my behaviors in circumscribed ways, while still not think that this mode of mattering rises to the level of a care proper.

Finally, we saw that our caring attitudes are unified in the sense that they are directed at a single intentional object. This marks another crucial difference between them and our sense of import. While the latter are *unified* insofar as they are indexed to, or centered around specific practical contexts, they are not referentially related to any one specific object. My sense of import with respect to driving contexts is composed of dispositions to attend and respond to many different objects: aggressive drivers, traffic lights, pedestrians, and so on. These dispositions are unified in virtue of the fact that they pertain to a certain practical domain, but this *contextual* unity is different from the *referential* unity that the emotional dispositions that compose our cares proper enjoy.[24]

To conclude, while our senses of import are sensitive to reasons in a robust sense, and are composed of at least some of the same kinds of attitudes that cares proper consist in, they do not amount to cares proper. Our sense of import is a distinct mode of mattering. It is an attenuated mode of having our landscapes of import impress themselves on us.

8. Conclusion

Human agency requires that we develop a sense of import, which is a minimal mode of mattering, akin to—but importantly different from—our caring attitudes. This minimal mode of mattering occupies a part of our affective and evaluative lives, situated between our full-blown concerns, projects, and values on the one hand; and our mere behavioral impulses on the other. Moreover, our sense of import somehow constrains our cares proper. Too much instability in our

[24] Contextual unity is not necessarily "weaker" than referential unity in the sense that one implies the other but not vice versa. If my attitudes enjoy referential unity then they are about the same object. if they enjoy contextual unity then they will be entertained in the same contexts. it is tempting to say that one unity is more diachronic and the other is more synchronic, but I don't know if that is true.

sense of import with respect to too many practical domains corrupts our emotional and attentional makeup to such a degree that we cannot develop cares proper. Given this dependency, questions concerning the positive relation between the sense of import and our cares proper naturally present themselves. Are senses of import merely *constrain* our cares, or do our cares proper related to senses of import in a tighter way? For instance, are cares constitutively, or developmentally, or, for that matter, conceptually linked to our senses of import?

Let me venture a highly speculative hypothesis: our cares and concerns at are at least partly determined by our sense of import in that they *emerge* from it. Once our sense of import around a practice reaches a certain level of differentiation, sophistication, and articulation it is indistinguishable from—indeed it is identical to—cares about certain matters in that practice. Once senses of import take on a certain *form*, they become cares and concerns proper. The sense of import is *the matter of cares.*

Chapter 3. THE IMPISHNESS OF AFFECT

> *[W]hen presently afterwards the little Man came in and asked, 'Now, your Majesty, what is my name?' at first [the Queen] asked:*
> *'Is your name Tom?'*
> *'No.'*
> *'Is it Dick?'*
> *'No.'*
> *'Is it, by chance, Rumpelstiltskin?'*
> *'The devil told you that! The devil told you that!' shrieked the little Man; and in his rage stamped his right foot into the ground so deep that he sank up to his waist. Then, in his passion, he seized his left leg with both hands, and tore himself asunder in the middle.*
>
> *Grimm & Grimm (1909)*

1. Introduction

The queen's mere invocation of Rumpelstiltskin's name marks a moment of literal rupture in the fairytale's narrative. Upon hearing it, Rumpelstiltskin launches into an act of terrifying (and entirely not age-appropriate) self-dissolution, allowing the queen, we might legitimately suppose, to live happily ever after. Speaking the *true name* of some magical creatures can alter, bind, and even destroy them.

In honor of Rumpelstiltskin, call a psychological state, process, or attitude *impish* if, roughly, *saying* what it is *changes* it, or *makes* it so. If its subject's act of *naming* it enables her to *conjure* or *control* it. In this paper, I will defend the idea that a large class of our *affective* states and processes are impish in this sense. That is, I will argue that, sometimes, talking about those psychological entities is a powerful way of altering them, and I will develop a model for how such alteration might occur.

I will get clearer on what I take impishness to amount to in a moment. But I want, first, to remind you of the pervasiveness in (some corners of) our culture of the thought that labelling or talking about our emotions can change them. Self-help manuals, wellness tutorials, and new-age texts abound with advice to *affirm* our beliefs and values, and to *manifest* the emotions and traits

we desire by ritualized acts of self-ascribing them. As New York Times bestselling author, Brené Brown, puts it,

> Language doesn't just communicate emotions; it shapes them. So when I use a word to describe how I'm feeling, my body will often follow my language. If I use hyperbole and go from saying, "I'm stressed" to "I'm overwhelmed," often I make that a self-fulfilling prophecy. (Oprah Daily, 2022)

Rough ideas about impishness seem to be part and parcel of contemporary folk psychology (at least for those members of the folk who go on meditation retreats and own personalized yoga mats), but they also seem like a particularly *dubious* part of it. While they are not full-blown affirmations of the paranormal, they appear to carry more than a whiff of magical—or just good old-fashioned wishful—thinking.

In this chapter, I aim to vindicate, at least to an extent, this part of our folk psychology. I'll argue that you *can* shape the contours of your emotions by talking about them. My argument for the impishness of affect takes the form of a response to a question: there is converging empirical evidence to the effect that a heightened capacity for interpreting, articulating, and reflecting on our affectively-valenced states and processes is at least a highly reliable *indicator* of people's capacity to effectively regulate them. Call this the *Interpretation-Regulation Connection*. I will ask what explains the Connection, and I will argue that the best way of accounting for it involves thinking about our affective states and processes as impish: their subjects' acts of interpreting them are (sometimes) sufficient to change them to accord with the interpretation. I will then develop a model for how labelling our emotions can sometime achieve this effect.

In Section 2, I survey evidence for the Interpretation-Regulation Connection, and in section 3, I present an explanation for the Connection that competes with the view I will recommend, and that has been recently advanced by Stichter and Fridland (2025). The alternative explanation—which I call the Knowledge Account—suggests that interpretive acts directed at our emotions are

not sufficient for changing them. Instead, such acts can yield important *knowledge* regarding our cares, concerns, values, and goals. This account, I argue, is unsatisfactory, since some forms of gaining knowledge with respect to our evaluative outlooks can *alienate* us from these outlooks, and hence exacerbate our condition.

In section 4, I analyze the Knowledge Account's failure to explain the Interpretation-Regulation Connection by drawing on Moran's (2001) distinction between theoretical and deliberative modes of gaining self-knowledge. While the former essentially involves evidence-based inference, the latter involves making certain commitments. The problem with the Knowledge Account, I claim, is that it takes the theoretical mode as typical of how we become aware of our affective reactions' significance for us. Instead, I argue that we should think about acts of self-interpretation as *commissive.* At least in the usual case, when we self-interpret, we provisionally commit to caring about certain things. Making such provisional commitments means both determining what our affective reactions *stand for*, or are *about* (and hence changing them in a way that fits the content of our interpretation), *and* rendering them internal to our evaluative-motivational perspective. This proposal can elegantly account for the Interpretation-Regulation Connection, as well as the intimate relationship between a person's heightened propensity for self-interpretation and a better awareness of her cares and concerns. Then, in sections 5-7, I draw on earlier work to fill in the details of this suggestion, to provide a model for what our affective reactions are like such that interpreting them can amount to changing them.

Before starting, I want to clarify the highly circumscribed sense in which I think interpreting our emotions can change them. Here, then, is a more nuanced way of spelling out what I take impishness to amount to:

> IMPISH. A psychological state, process, or attitude is impish if its *subject's acts* of applying a lexical concept (or a string of lexical concepts) to it is sufficient to *causally* change it,

often in a manner that makes it more akin to the content ascribed to it by the lexical concept (or string of lexical concepts).

A few remarks. First, notice that the *application of a lexical concept* (or concepts) can occur overtly, in, say, dialogue or in writing; or covertly in inner speech. In fact, you might think that subjects can apply lexical concepts to their attitudes even without internal monologue. My characterization is supposed to apply to all of these cases.[25]

Second, the modifier *causally* is important to distinguish claims about a mental entity's impishness from a general thesis regarding a conceptual connection between facts and descriptions. For instance, you may deny a broadly Davidsonian thesis according to which the individuation of mental states is grounded in a specific description of those states and still think that some psychological entities are such that overtly interpreting them can cause them to change. That a certain state is impish does not mean that there isn't *a fact of the matter* with respect to its identity independently of the interpretative act. Rather, it suggests that such facts can change in virtue of the interpretation.

Thirdly, the emphasis on *subjective acts* is important to distinguish cases of impishness from other (presumed) cases of sub-personal *cognitive penetration*.[26] Some studies purport to show that, when the object of a subject's perceptual state is *labeled*, the content or phenomenal character of the state can change in accordance with the label. For example, in a famous study, participants who were presented with a test odor that was labeled either as "body odor" or as "cheddar cheese," were found to rate the odor as significantly less pleasant when the former label was used (De Araujo et al., 2005); and in another study, subjects who were presented with a racially ambiguous

[25] While I think that the etiological grounds for the impishness of our affective attitudes ultimately lie in the development of our interpersonal, dialogical capabilities, I will not touch on this claim in this paper (and it certainly does not function as an assumption here).

[26] Roughly, a non-cognitive state is cognitively penetrable, if it can be altered by its subject's beliefs or desires, while certain facts about her, such as her sensory inputs, the state of her sense organs, and the orientation of her spatial attention are held fixed (Macpherson, 2012; Silins, 2016).

face judged that face to be darker when it was labeled "black" than when it was labeled "white" (Levin & Banaji, 2006). While these results might provide evidence for the idea that perception can be influenced by a subject's cognitive states, they don't provide direct evidence for the idea that they are impish, since they do not involve the personal-level interpretation of these perceptual states by the subjects themselves.[27]

Lastly, the term *acts* is meant to bypass debates about what should count as full-blown actions. I am happy to allow that some agent-level processes of applying a lexical concept to a psychological attitude (for brevity, I will interchangeably use the terms *labelling* or *interpreting* for such acts) can be habitual or unintentional, and I am happy to call such processes 'actions'. If, however, you are reluctant to use the term to refer to those agential activities, then hopefully you'd be more comfortable with the terminology I'm offering.

All right. Enough with the ground clearing.

2. The Interpretation-Regulation Connection

Let us start by surveying some of the evidence for the presumed connection between interpreting our emotions and regulating them (the following survey as well as the discussion in the next section are deeply indebted to Stichter and Fridland's 2025 work of mapping and drawing connections between these phenomena). One such line of evidence can be gleaned from the literature on emotion differentiation, or emotional granularity (henceforth EG. Barrett et al., 2001).[28] Roughly, EG is an individual's scalar capacity to draw fine-grained distinctions between her occurrent affective states. Individuals who score high on EG metrics tend to label their emotions "in a differentiated, specific, and context-dependent way," (Erbas et al., 2019: 258)

[27] For an excellent treatment of the personal/subpersonal distinction, consult (Drayson, 2012, 2014).
[28] I will follow common practice and use the term "emotion differentiation" as interchangeable with "emotional granularity" (although see Smidt & Suvak, 2015 for a distinction in the way these constructs may be operationalized).

while individuals low on EG tend to use broad, general, non-specific vocabulary to describe their affective experiences.

Several studies indicate that people with high EG are better at regulating their emotions than people with low EG (see Smidt & Suvak, 2015; and Thompson et al., 2021 for reviews). For instance, empirical evidence suggests that the former are less likely to respond aggressively when offended (Pond et al., 2012); and to use alcohol to cope with negative emotions (Kashdan et al., 2010); and among people with borderline personality disorder, high EG has been found to safeguard against self-injury (Zaki et al., 2013). Furthermore, better capacities for emotion differentiation are related to lower levels of neuroticism and to higher levels of self-esteem (Erbas et al., 2014), which, in turn, predict better emotion regulation. On the other hand, low EG is correlated with a variety of mental conditions associated with deficits in emotion regulation, like major depressive disorder (Demiralp et al., 2012), schizophrenia and schizoaffective disorders (Kring et al., 2003), borderline personality disorder (Suvak et al., 2011), and autism (Erbas et al., 2013).

More evidence for the Interpretation-Regulation Connection can be drawn from studies regarding *affect labeling* (henceforth AL)—the act of naming and describing one's occurrent affective feelings, or the emotional content of a scene. There is a rich line of results which suggest that "putting our feelings into words" is correlated with many beneficial regulatory effects. For instance, AL is correlated with reduced distress in response to negative visual images (Burklund et al., 2014; Lieberman et al., 2011); reduced skin conductance and voice pitch in the face of such images (Matejka et al., 2013); and a decrease in people's self-perceived negative reactions to aversive stimuli (Mendolia & Kleck, 1993). Furthermore, AL was found to correlate with increased effectiveness of exposure treatment for individuals with public speaking anxiety (Niles

et al., 2015), and with arachnophobia (Kircanski et al., 2012); and with improved test scores for students who wrote about their exam-related anxieties prior to taking a test (Ramirez & Beilock, 2011). In addition to its ameliorative effects, AL is linked with increased intensity of positive emotions (Vlasenko et al., 2021). Indeed, AL was recently theorized as a process of implicit emotion regulation (Torre & Lieberman, 2018).

Finally, studies of alexithymia consistently point to the idea that an *in*capacity to talk about, interpret, and differentiate between affective states is related to disruptions in emotion regulation. Alexithymia is often characterized as "an impaired ability to be aware of, explicitly identify, and describe one's feelings" (Hogeveen & Grafman, 2021: 48). Alexithymia ratings can be construed as falling on a spectrum from low to high, where sufficiently low ratings are considered pathological. Relevantly for our purposes, it is widely agreed that alexithymic individuals suffer from deficits in their capacities for emotional processing and emotion regulation (Hsing et al., 2013; Laloyaux et al., 2015; Preece et al., 2022, 2023). In fact, alexithymia is associated with a large number of affective disorders (G. J. Taylor et al., 1997), including major depressive disorder, panic disorder, generalized anxiety disorder, and post-traumatic stress disorder (Frewen et al., 2008; Leweke et al., 2011; Onur et al., 2013); and it is considered a transdiagnostic risk factor for many of these conditions (Honkalampi et al., 2007; Preece et al., 2022. For reviews, see Hogeveen & Grafman, 2021; and Zamariola et al., 2018).

Taken together, these sources of evidence point to the idea that an agent's capacity to articulate her emotions is somehow tightly related to her capacity to *control* them. A heightened ability to bring our linguistic or conceptual resources to bear on our emotional lives is at least indicative of a heightened ability for emotional regulation. Is this connection *only* indicative? Is it merely a byproduct of such an ability? An affirmative response to this question could be based on

the idea that nuanced emotional differentiation skills, a high tendency for affect labelling, and/or low alexithymia can be traced back causally to a refined capacity for *emotional awareness*. People with such dispositions are better at *detecting* their own occurrent emotions, and that is why they are better at *describing* or *reporting about* them. On this story, acts of interpreting, articulating, or conceptualizing our affective states and feelings are not sufficient for changing them. They might be a part of coming to realize what we have been truly feeling, but *the feelings themselves* have been there all along. I think this account is inadequate, but to give it a full hearing, I want to consider a recent sophisticated and perceptive version of it.

3. The Knowledge Account

In a recently-published Cambridge Element (2025), Stichter and Fridland develop and propose a solution to a puzzle, which they call *the puzzle of "feeling well."* They observe that many phenomena that seem to involve an individual's heightened awareness of her emotional and bodily states (including high levels of emotional granularity and low levels of alexithymia) seem to be intimately connected to her well-being, and they ask why this is the case. In their words: "why it is that good mental health is to some extent dependent on having bodily/affective/emotional self-knowledge, and why a lack of such knowledge contributes to poor mental health and well-being?" (5).

The puzzle of "feeling well" is highly pertinent to our own discussion, but as stated, it seems to presuppose an important point that I wish to remain agnostic about for now (and ultimately reject), regarding the nature of the explanatory connection between self-knowledge and mental well-being. According to Stichter and Fridland's formulation of the puzzle, self-knowledge is supposed to be the explanatorily relevant causal antecedent (along with other conditions, perhaps) of the many positive health and well-being outcomes that they document. The way the

puzzle is formulated already suggests a sketch for its solution: a subject experiences a blooming, buzzing array of affective feelings, comes to know what they are—or what they *signify*— and then uses the knowledge in auspicious ways. The explanatory gap that needs filling concerns the peculiar auspiciousness of emotional self-knowledge: we suppose that this form of knowledge is great, but what, exactly, is so great about it? But it is important to flag at the outset that an alternative construal of the conditions leading to the puzzle could suggest that self-knowledge is not, in the final analysis, what is doing the explanatory work. Instead, a subject's expansive emotional self-knowledge might be *symptomatic* of another fact about her, which ultimately explains the mental health and well-being outcomes that Stichter and Fridland (and I) review. I will get back to this point very soon.

For the present moment, however, let us set this contention aside and concentrate on Stichter and Fridland's solution to the puzzle. According to them, "self-knowledge is indispensable for knowing what matters *to us*" (10, emphasis original). Our emotions, they argue, involve felt appraisals of our standing relative to our own goals, values, interests, cares and concerns (which they collectively dub 'goals+'); they draw our attention to situational features that are relevant to the attainment of our goals+; and they prepare us to act in ways that promote them. This means that, apart from granting us "access to the evaluative or normative properties of the world" (13), emotions can also offer us precious insight about *ourselves.*

> Emotions reveal what matters to us and thus constitute a method for acquiring such critical self-knowledge. Knowing how we feel in different situations, with different people, while engaging in different activities, and at different times is the only way we have for learning about the value that these different situations, people, and activities have for us. (13)

Our goals+, according to Stichter and Fridland, are far from always being transparent to us. Hence, attending to our emotions can be an invaluable way of acquiring crucial knowledge about ourselves that is not readily available to us otherwise. In turn, this kind of self-knowledge

regarding our goals+ could be used for "decision-making, goal setting, and action planning" (26). By understanding the nature, causes, and targets of our emotions, we can understand what our values, cares, and concerns are, rationally assess them, and either change them, or develop plans for their pursuit.

Call the suggestion that we are considering the Knowledge Account. If the Knowledge Account is broadly correct, then affective states and processes are *not* impish. Articulating, interpreting, or conceptualizing them does not directly change them. Instead, such acts can yield important *knowledge* that, in turn, may lead the interpreter to adopt, for instance, better strategies for coping with them. In and of themselves, however, acts of interpretation are not acts of emotion regulation.

In fairness to Stichter and Fridland, right before concluding their book, they briefly discuss the idea that interpreting one's emotions can lead to qualitative changes in her emotional experience. They observe that, for people who are relatively adept at differentiating their emotions,

> it's not just that you know more about what you're feeling, because you're also changing what it is that you're feeling. Those who were skilled [at emotion differentiation] had qualitatively different affective experiences [than unskilled emotional differentiators] because what they were feeling wasn't 'broad, undifferentiated, and diffuse.' (52)

However, here too, they can be read as suggesting that the mediating causal factor between a person's acts of interpreting her affective experiences and their alteration is the attainment of self-knowledge: "The process of acquiring self-knowledge about their affective states was leading [the skilled differentiators] to transformations in their affective states" (52). The proposal seems to me to consist in the idea that people with high EG—and, we may add, high AL and low alexithymia—are spared the confusion and distress that result from a person's *lack of knowledge* with respect to the identity, sources, and targets of her occurrent emotions. As they say, high EG "allows one to take an otherwise global and diffuse affective state (or mood), like 'I'm feeling

bad,' and recognize that one is feeling a specific emotion" (52). On the other hand, unskilled emotional differentiators "experience their distress as a diffuse negative mood which doesn't have a clear source and doesn't indicate a particular action to take to alleviate the distress—hence the confusion" (53). They do not realize that what they are experiencing is anger, disgust, fear, etc., and this lack of knowledge adds the insult of confusion to the injury of their initial distress.

I do not wish to engage too heavily in textual exegesis, and I acknowledge that the Knowledge Account is, in no small part, a rational reconstruction of Stichter and Fridland's considered view.[29] What matters for my purposes is that the Account seems like a coherent counterproposal to a view according to which emotions are impish. According to the impish story, articulating our emotions is (sometimes, at least) sufficient to change them. In contrast, according to the Knowledge Account, while the act of articulating our emotions cannot in itself change them, it can nevertheless provide us with indispensable knowledge that, first, we can utilize to alter and control them; and that, second, can alleviate our uncertainty with respect to them, and *hence* ameliorate them. On the latter proposal, emotional regulation is the result of emotional knowledge; on the former, such knowledge might be indicative of, but it is not (primarily) responsible for our capacity to regulate our emotions.

As I said, I think the Knowledge Account should be rejected. Most importantly, because there is a mode of obtaining knowledge about our desires, cares, and concerns that may render some of them thoroughly destructive. To exemplify this mode of awareness, you can imagine (perhaps even recall), a graduate student who, while optimistic about her career prospects, is also cognizant of the terrible—and ever-worsening—conditions of the academic job market. Still, as I

[29] After personal correspondence with Fridland, I have become convinced that her considered view is quite close to mine. Ultimately, our disagreement might boil down to whether and to what degree this view transpires in their Element.

ask you to imagine her, she typically does not consciously concern herself with her future (un)employment. Instead, following the encouragement of her advisors, she does her best to concentrate on her dissertation writing, thinking that she'll cross the dreaded application bridge when she gets to it. Over time, however, her attempts at repressing her anxieties take their toll, and after many a restless night she can no longer avoid interpreting her distress as pointing to the fact that the only career she can see herself pursuing—the only project that comes alive for her as a path towards a meaningful life—is academic philosophy.

Such a realization may motivate some people to try and change their evaluative perspective in a thoroughgoing way, but others (including, as I imagine her, our unfortunate grad student) may find it utterly debilitating. The thought that the structure of your current goals, cares, and concerns is such that, *by your own lights*, you will never lead a life of worth unless you secure a tenure-track position can just as easily *worsen* your anxieties and cripple your agency with respect to them. In these cases, you may reasonably view your deep-seated motivational and evaluative makeup as a misfortune that befell you, and that bullies you into enduring the pains that are required to satisfy it.

While perhaps not pervasive, this mode of becoming aware of our values, cares, and concerns through our affective feelings is not rare either. People assume it when they experience their feelings as signifying that they still love their abusive ex-partners; that they are much more susceptible to peer pressure than they'd like to think; or that, as Mitchell and Webb put it, they are, in fact, "the baddies." Sometimes, emotional self-knowledge of the kind Stichter and Fridland are concerned with can help us ameliorate our motivational-evaluative perspectives, but on many occasions, it can increase our alienation from them, and our pain in the face of this alienation. Self-

knowledge is an exceedingly sharp double-edged sword. As such, it cannot be what explains the Interpretation-Regulation Connection.[30]

4. Theoretical and Deliberative Modes of Awareness

Still, I think that the Knowledge Account is correct in emphasizing two affinities: (1) an affinity between the beneficial effects that we've been considering and our values, cares, and concerns; and (2) an affinity between *knowledge* of those values, cares, and concerns and subjective acts of articulating, interpreting, and conceptualizing our emotions. These affinities, however, are not what the Knowledge Account takes them to be. The Account conceives of our values, cares, and concerns as waiting to be revealed through engagement with our emotions, and it construes the discovery as robustly beneficial. I think we should deny these suppositions, but insist on the very real nature of both affinities.

My preferred impish account—call it the Commitment Account—does this by, first, suggesting that typically, articulating our emotions is a way of *gaining knowledge* about our values, cares, and concerns, because it is a way of *determining* what our values, cares, and concerns are. Second, the account suggests that in turn, determining our values, cares, and concerns is a way of exercising our agency with respect to them, and hence also with respect to our emotions. Determining our evaluative-motivational perspectives is (at least often) committing to these perspectives.

[30] Of course, this is not a knockdown argument. For instance, a modified Knowledge Account might allow that cases of alienating self-knowledge are quite common, but insist that still, *usually*, more self-knowledge is prudentially better. This response seems to depend precariously on the thought that our evaluative perspectives are such that, if we were to become *theoretically* aware of them, we would typically either (1) approve of them or (2) be able to work our way towards altering them. To me, both (1) and (2) sound overly optimistic, and furthermore, option (2) does not seem to explain the at-times rapid ways in which, for instance, affect-labelling operates. At the very least, these worries and the availability of an alternative account should serve as an invitation for potential defenders of the Knowledge Account to spell out the view further. Thanks to Manuel Vargas for raising this question.

To make headway on the first point, I want to import a prevalent distinction between questions the answers to which are based on evidence, and questions that are answered through decisions (Hampshire & Hart, 1958). While not without its critics, this distinction is commonly used to contrast, for example, the ways in which agents come to form beliefs about their intentions on the one hand, and to predict their behaviors on the other. According to many theorists (although by no means all. See Dennett, 1989, for one example of a highly-influential counter), the usual mode of becoming aware of—and forming beliefs about—our intentions is *making up our minds* about them, of *deciding* what to do. Call this the *deliberative* mode (I borrow the terminology from Moran, 2001). On the other hand, in many cases, agents can *anticipate* that they will act in certain ways, perhaps *despite* their pre-formed intentions. In such cases, agents form their beliefs about their future behaviors through evidence-gathering, which might include introspecting their motives, weaknesses, and soft spots. Call this mode of awareness *theoretical*. Presumably, we feel the tension between these two modes of awareness, or stances, whenever we expect to be akratic, for instance.

To be clear, I do not wish to commit to any specific account of intentions. I invoke the literature on this topic only to illustrate the differences between the deliberative and theoretical modes of awareness. If you think that this distinction is misguided in this context, you may think about it as operative in other contexts. For instance, it characterizes the difference in how one learns about Bilbo Baggins's adventures in the Misty Mountains (theoretically, by reading *The Hobbit*), and how one becomes aware of the ending of her own cyberpunk Tolkien fan-fiction novel, *The Return of the Ping* (deliberatively, by writing it).

Now, one way of articulating the problem at the core of the Knowledge Account is by pointing to the idea that it takes the theoretical mode as typical of how we become aware of our

emotions' significance. I wish to suggest that often, the relation in which we stand to our affective states and processes when we interpret, articulate, and conceptualize them is qualifiedly *deliberative*. Engagement in such activities involves provisionally deciding and committing to which cares and concerns our affective reactions stand for; and hence, such activities involve *making up our minds* with respect to our cares and concerns.

According to the Commitment Account, interpreting our affective reactions involves assuming the deliberative stance with respect to our evaluative perspectives. This idea elegantly explains the tight connection between such acts of interpretation and knowledge of our cares, concerns, goals, and values. Such knowledge is symptomatic of our capacity to articulate our emotions, since articulating our emotions *is*, at least in the usual case, making up our minds regarding what we care about or value.

I want to take a moment to spell out where we stand in this paper's dialectic progression. We started by examining a few phenomena that seem to point to a crucial link between our capacity to talk about our emotions, and our capacity to control them. This is the Interpretation-Regulation Connection. We tried, unsuccessfully, to explain the Connection without invoking the impish thought that talking about our emotions can change them. Along the way, and with Stichter and Fridland's help, an important idea presented itself: the Interpretation-Regulation Connection goes through what matters to us. Being able to talk about our emotions is indicative of a heightened awareness of our cares; and a heightened awareness of our cares is indicative of a better capacity for emotional regulation. Then, in this section, I proposed that often, articulating our emotions *just is* determining our values, cares, and concerns. This proposal, I noted, can eschew the problem that the Knowledge Account faces, according to which self-knowledge (acquired in the theoretical

mode) can often alienate us from our evaluative perspective and hence exacerbate rather than alleviate our predicament.

So, we now have a motivation for thinking that sometimes, interpreting our affective reactions is a way in which we can assume the deliberative stance with respect to our own evaluative outlook—to own, authorize, or commit to it. The question is, are our affective reactions amenable to this motivation? Specifically, can assuming the deliberative stance with respect to our *affective reactions* amount to assuming it with respect to our *evaluative perspective*? The next section aims to provide an affirmative response to this question.

5. The Sense of Import

In the previous chapter, I argued that humans' evaluative and motivational economies include stable, integrated patterns of finding certain things important to us, which resemble—yet do not rise to the level of—full-blown cares as these are typically understood by contemporary moral psychologists. I called such patterns our *Senses of Import* (henceforth, SoIs), and I suggested that they consist in dispositions to undergo similar sets of rapid, affective and attentional reactions in the face of similar practical contexts. Agents develop such characteristic patterns of attending and emoting to many features of our practical circumstances: we can have such patterns for contexts that include certain activities (we can have SoIs for, say, driving, cooking, arguing, or writing); certain people (our SoIs can be sensitive to our best friend, our high school sweetheart, and our arch-nemesis); and certain places (our SoIs can incorporate differential reactions to our childhood home, our department office, and the DMV), to name just a few categories.

Now, affective and attentional reactions are *evaluative*: they present things in ways that render us non-indifferent to them. Undergoing such reactions in response to their eliciting conditions entails taking these conditions to be *relevant*, or *to matter* to us—at least in a

rudimentary sense.[31] This means that, as patterns of affect and attention, our SoIs are *sensitive to reasons*: to (at least pro tanto) considerations we take to count in favor or against certain actions (Arpaly, 2003; Carman, 2018; Döring, 2003; Jones, 2006; Railton, 2017). Furthermore, affective—as well as attentional—reactions are *motivational*:[32] they facilitate decision-making and prepare us for action in rapid, effective ways that do not necessarily go through explicit practical deliberation (Damasio, 2005; Frijda, 1986).[33] As I argued, because SoIs are comprised of patterns of reactions that are both reasons-responsive and fast, they are also *ubiquitous*. Creatures like us have a standing need to be able to direct our actions in ways that both take into account the many considerations that matter to us in a given situation, and that do not rely on our slower deliberative faculties to recommend a single course of action. Our SoIs are what constitutes this ability.[34]

SoIs, then, are habitual modes of taking certain features of our practical contexts as mattering to us. And like our cares, concerns, and valuing attitudes, they can be thought of as complexes of many different dispositions to instantiate psychological attitudes. Yet they do not amount to *bona fide* cares as these are typically conceived of by contemporary moral philosophers for a number of reasons (surveyed in the previous chapter). I want to dedicate a few paragraphs to concentrate on one of these reasons, which will prove to be important in what follows.

Specifically, our habitual SoIs are not *referentially unified* in the way that our cares are. The various attitudes that comprise our full-blown cares are directed at a *single* intentional object. We care about a person, a project, a pet (figure 1). On the other hand, the multitude of affective

[31] This way of putting things is supposed to be neutral between content-based and attitudinal (Deonna & Teroni, 2015; Jacobson, 2019) theories of affect.
[32] For empirical support for the motivating role of attention, consult, for instance, Ransom et al., (2020); and Todd et al., (2020).
[33] Notably, theorists such as Wayne Wu (2023) are so impressed by the connection between attention and action that they view it is conceptual. On Wu's account, attention *just is* preparation for action.
[34] A few more variations on the theme of our affective reactions' necessity for action, include Brownstein (2014); Pacherie (2002); and Rietveld (2008).

and attentional reactions that comprise our SoIs with respect to a given practical situation can be directed at many different features of that situation (figure 2). Your SoI at the company party may involve slight agitation at the prospect of socializing, boredom at the meaninglessness of the small talk, discomfort at the blazer that hangs unnaturally on your body, minute nervousness about the possibility of making a *faux pax*, and an almost imperceptible drive to just screw all of this and go home. That is, your evaluative-motivational reactions are directed at, inter alia, other people, abstract aspects of the situation, the semantic content of occurrent conversations, your body, social affordances and threats, and alternatives to your presence in the room. Call this feature *Referential Disunity*.[35]

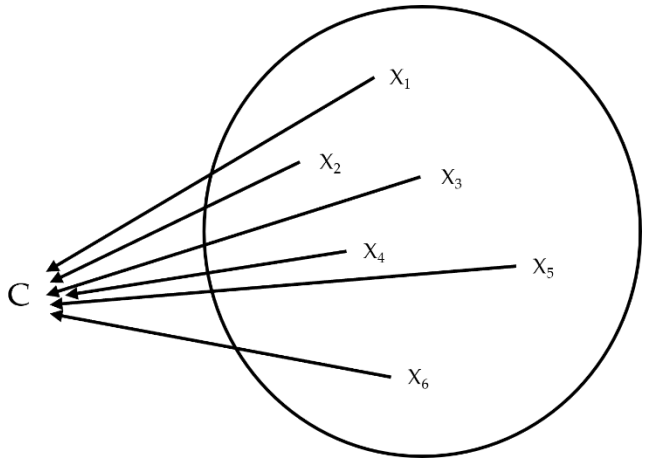

Figure 1. Cares are referentially unified. Each of the X_i represents an attitude that partly comprises the entire care. The C represents the care's intentional object.

[35] In Helm's (2001) terminology, like cares, SoI's may have many *targets*, but unlike cares, SoI's do not have a unitary *focus*.

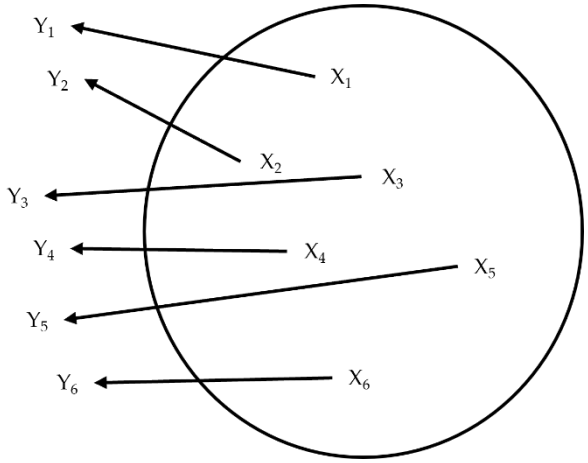

Figure 2. The sense of import is referentially disunified. Each of the X_i represents an attitude that partly comprises the sense of import with respect to a given situation. Each of the Y_i represents the respective attitude's intentional object

Now, the Commitment Account of the Interpretation-Regulation Connection maintains that articulating our affective reactions is a way of assuming the deliberative stance with respect to our cares and concerns. That is, it is a way of deciding and committing to what our cares and concerns *are*. The question is *how* can this be the case. What our affective reactions are like, such that interpreting them could possibly amount to committing to, or settling on a certain evaluative perspective? My suggestion is that we can answer this question if we realize that many (perhaps most) of our affective reactions are parts of the patterns of attitudes that constitute our SoIs. Hence, many cases in which we talk about, differentiate between, or label our emotions are, *ipso facto*, cases of talking about or interpreting our SoIs. In turn, keeping in mind that SoIs are referentially disunified can help us see what the SoI is like, such that interpreting it can (and often does) amount to assuming the deliberative stance with respect to it. And, since, as I have said, our SoI just is (at least a part of) our evaluative outlook, then assuming the deliberative stance with respect to the former just is assuming that stance with respect to the latter.

Here is the story. The SoI is referentially disunified in the sense that the reactions that comprise it do not aim at a single intentional object. This, I suggest, means that the SoI is *multiply*

interpretable. There is more than one legitimate, appropriate, or fitting way to understand what it signifies for its subject (figure 3). Take our party example. Your affective and attentional reactions in that context are not about any *one* specific thing. As I said, they are directed at many different features of your practical context—they are referentially disunified. But collectively, they characterize your action-guiding, evaluative-motivational perspective in the situation. This makes it possible to articulate different appropriate interpretations of your evaluative perspective. The various reactions that you are undergoing can be interpreted as signifying, for example, a concern for your precarious social standing within office politics, or anxiety about being perceived as intellectually inferior, or a disdain for (what you take to be) shallow gatherings. Still, none of these interpretations is mandatory, precisely because SoIs do not amount to full-blown cares.[36]

In other words, your SoI *fits*—in the proprietary sense that is sometimes invoked in contemporary meta-normative discussions—a number of cares or concerns. The facts of one's situation plus any one of the cares and concerns that serve as legitimate candidate interpretations of her SoI in that situation would render these reactions appropriate. The former would provide *normative support* for the latter.[37] The crucial point, again, is that there can be more than one fitting interpretation for the occurrent reactions that characterize your SoI in that context. This is why talking about, articulating, or interpreting the SoIs can amount to taking the deliberative stance with respect to them.

---

[36] I am willing to grant for the purposes of this essay that there may be cases where there is only one fitting interpretation for the SoI. But I suspect that these are quite rare (in fact, I suspect that they are non-existent), at the very least, because I don't think that human values, cares, and concerns are, in the usual case *maximally determined.* There are always more nuanced, refined, or specific ways of spelling out the contents of our cares. Compare Nguyen's discussion of value determination in his (2024).

[37] This formulation is meant to be neutral with respect to the question of whether fittingness is reasons-providing. It is a common assumption in recent discussions about fittingness that the fact that an attitude is fitting entails that there is a reason to hold that attitude (see, e.g., Howard, 2018, en. 4). However, while common, this assumption is not without detractors. Still, even those detractors would agree that both *being a reason for* and *making-fit* are relations of normative support (see Maguire, 2018: 779).

This means that often, your SoI is, in an important sense, *partially discretionary*. On the one hand, it is constituted by patterns of rapid, action-guiding psychological reactions to your situation that are crucially non-volitional. On the other hand, while these reactions are attributable to you, you certainly do not have to own them, or avow them, or commit to their significance for you. But you *can* do so by assuming the deliberative stance with respect to your SoI, by settling (in the deliberative mode) on *one* interpretation for it.

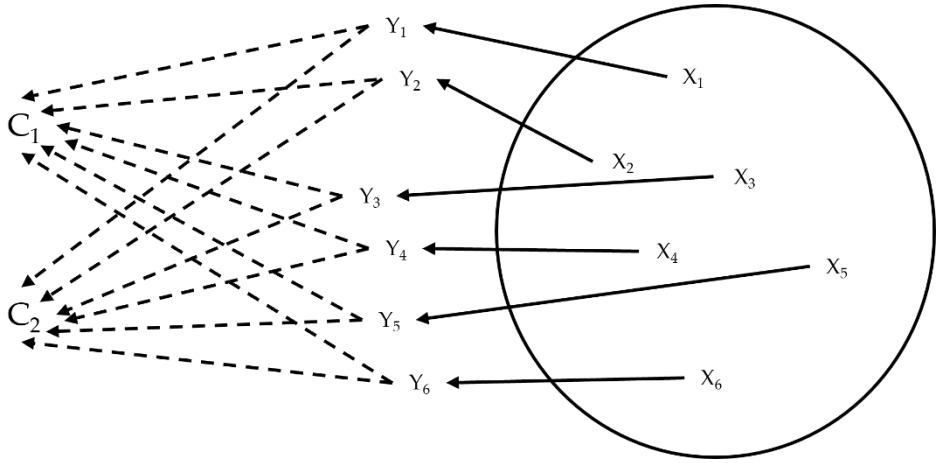

Figure 3. The sense of import is multiply interpretable. Each of the C_i represents a care that constitutes a potential, fitting interpretation for the attitudes that comprise the sense of import

So, articulating or interpreting the attitudes that collectively comprise your SoI can be more than just reporting on them. The SoI is such that interpreting it can be a matter of *deciding* which commitments, cares, and concerns it reflects, because which of these commitments, cares and concerns it reflects is *an open question*. And this, in turn, explains why interpreting your affective reactions can be a matter of deciding what your cares are.

6. From Emotions to Commitments

The Commitment Account that I favor, then, suggests that *labeling* or *interpreting* our affective reactions involves assuming the deliberative stance with respect to our multiply-interpretable SoI. It involves determining its significance for us, by settling on a certain interpretation of it. When subjects label or describe their affective reactions by using more or less

differentiated emotion terms, they thereby commit to *caring* (in the bona fide sense) about something in particular.

There are two questions that remain to be answered, however, for the Commitment Account to be satisfactory. First, often, when subjects describe their emotions, they do not explicitly refer to their cares, concerns, and values. So what, in talking about one's emotions, makes reference to her internal evaluative commitments? And second, why does committing to care about something allow for the kind of regulatory effects that we observe in people with a heightened capacity to articulate their emotions? This section answers the former question, and the next section will provide a tentative response to the latter.

Notice, first, that when people interpret their affective feelings, they often do not merely state *what* they are feeling, but also the object of their state. To use an illustration from a recent study, a person high on EG may report feeling "guilty *for one's clumsiness*," whereas a person low on EG may suggest that she is "generally upset" (Hoemann et al., 2023: 291, my emphasis). Furthermore, training people to increase their EG and AL skills involves prompting them to include the object of their affective feelings in their reports. For example, people who trained in affect labeling during exposure therapy for arachnophobia were encouraged to respond to a spider with statements like "I feel anxious the disgusting tarantula will jump on me" (Kircanski et al., 2012: 1087); and AL training protocols for primary school students involve modules dedicated to "understand[ing] emotions by analyzing *the types of events and situations that cause and result* from various emotions" and by "learning about the *triggers* and consequences of the emotional experiences of adults" (Brackett et al., 2012: 220, my emphases).

But more principally, one essential feature of emotions is precisely their intentionality. Emotions are *about* certain particular and formal objects, and they are at least partly individuated

by those objects. To use some of the usual examples, fear is directed at *the dangerous*, anger at the *offending*, admiration at the *admirable*, and so on. So, when an agent interprets her SoI as a tokening of a certain emotion, she thereby also (implicitly or explicitly) ascribes a formal object to her SoI, out of a range of putative fitting candidates—this is a manifestation of the SoI's amenability for multiple interpretations. If, for example, you label your SoI at the company party as an expression of *social fear*, you thereby (implicitly or explicitly) label your situation as socially dangerous. For instance, as posing a risk to your standing in the department's unofficial hierarchy. According to the Commitment Account, this entails at least a tacit commitment to treating the situation as socially dangerous, on the assumption that *the socially dangerous* is the appropriate formal object of your fear. But treating the situation as socially dangerous, as posing a risk to your social standing means, in this context, *caring* about your social standing.[38] In general, the Commitment Account suggests that, when an agent is interpreting her attentional and affective reactions as a specific emotion, this implies a commitment to caring about the emotion's object, regardless of whether the object features in the explicit self-ascription by the agent.

7. From Commitments to Regulation

According to the Commitment Account, interpreting one's occurrent array of attentional reactions and affective feelings is, in the usual case, provisionally committing to care about some object that those reactions and feelings can be appropriately interpreted as directed at. That is, in the usual case, self-interpretation consists in assuming the deliberative stance with respect to our SoI. The lingering question is: why should committing to care about something entail the

[38] There are, arguably, cases of mental states that can represent, say, *the dangerous* in a way that does not imply caring about their intentional object in the relevant sense. For instance, when we calmly think about the ascent to the peak of K2. This, however, is not the case here. The SoI, as I stressed, is already evaluative and motivational such that undergoing the reactions that comprise it amounts to taking things as mattering to us, at least rudimentarily, in some attenuated way.

regulatory effects that I have recounted in section 2, such as a better capacity to control our reactions in the face of strong feelings, and to be less distressed by negative emotions? Why is committing to care about one of our SoI's potentially-fitting objects constitute regulating our SoI?

To answer this question, recall the problem we raised for the Knowledge Account: increased self-knowledge cannot explain the Interpretation-Regulation Connection, because some forms of (theoretical) self-knowledge can easily alienate us from our evaluative-motivational perspective, in a way that can increase our distress in the face of our affective reactions towards things we take to be of worth. Gaining knowledge about our values, cares, and concerns in the theoretical mode implies a kind of passivity with respect to the nature of these values, cares, and concerns.

A similar kind of passivity towards our SoI prior to its interpretation is, in a significant sense, the default case. While our SoI is, to some degree, a byproduct of our actions, it develops largely unbidden, shaped by a host of contingent factors to many of which we are far from fully privy. In most cases, we simply find ourselves occupying a specific evaluative-motivational standpoint which includes dispositions to undergo both highly salient emotional responses as well as barely-noticeable affective reactions. We are saddled, so to speak, with import. Thus, large swaths of our evaluative makeup are akin to what Calhoun calls *passive commitments.* That is, "psychological attractions" that we have "independently of deliberative decision making and subsequent employment of commitment conventions" (Calhoun, 2009: 616-7).[39] But, as I noted, these attractions need not be—and they often are not—internal to us.

[39] Somewhat confusingly, Calhoun calls these attitudes 'commitments,' despite the fact that she is quick to suggest that they "do not look much like commitments at all. They are simply deep psychological attractions. Being moved on the basis of such involuntary psychological attractions to pursue a project, maintain a relationship, care for the well-being of some entity, and the like is not yet to be committed" (617)

This passivity is transcended when we assume the deliberative stance with respect to our SoI. By interpreting it in the deliberative mode, we move from having our practical contexts impressing themselves on us to actively committing to care about something-or-other in particular—again, at least provisionally. In this way we *authorize,* or *accept responsibility for* our evaluative-motivational outlook as it is manifested by our affective reactions to our situations.

To spell out this point further, recall our unfortunate graduate student, who comes (theoretically) to the realization that the only life-plan that she may find valuable involves a career in academia. The question whether she is responsible for her cares *is* in some sense pertinent to her condition. Specifically, we might think that she is prudentially blameworthy for the shape of her present motivations. We (and she) may think that she should have taken steps towards diversifying her interests or articulating other compelling visions of the good life for herself. She is responsible, then, insofar as the cultivation of her myopic evaluative-motivational makeup was *up to her*, and hence she is criticizable for its present state. Therefore, she is also an appropriate target for reactive attitudes and agent-regret concerning it.

Yet this notion of responsibility is, as Pickard explains, essentially backward-looking, "whereby a person is judged and potentially condemned for their past behaviour" (2017: 175). But one way of analyzing the severity of our grad student's predicament involves acknowledging her deep sense of helplessness in the face of her own SoI, her role in bringing it about notwithstanding. By assuming the theoretical stance with respect to it, she treats it as a fact about herself that needs to be dealt with, but that she is in no qualitatively unique position to change. Rather than being inapplicable to her, the question: "is this the way I want to be?" is moot. Her attitude towards herself corresponds to what Lear calls *holding responsible*: "in holding oneself responsible one is

essentially taking a third-personal stance with respect to oneself and deciding whether one's character, actions, thoughts or feelings are worthy of praise or blame" (Lear, 1998: 66-7).

What our graduate student is missing is what Pickard identifies as forward-looking responsibility that involves a capacity "to identify where there exists capacity for change thanks to the presence of choice and control" (2017: 175). Her psychological position does not allow her to (re)gain a sense of agency with respect to her own cares and aspirations. She does not, in Lear's parlance, *accept responsibility* for her evaluative makeup. When a person accepts responsibility for some of her features, "she is able to acknowledge them as hers. Accepting responsibility is essentially a first-person relation. In accepting responsibility I acknowledge who or what I am" (67).

Now, several authors have discussed practices of responsibility-taking, but they often do not distinguish the passive, backward-looking aspects of such acts from their active, forward-looking aspects. These latter aspects—which I read both Lear and Pickard as trying to spell out—comprise what I will call *owning up* to your actions or attitudes. Mason, for instance, says that

> Someone who genuinely takes responsibility is willing to engage in the blame conversation. They understand and are willing to acknowledge blame, and to move through the steps of acceptance, apology, reparations, and so on. (Mason, 2019: 187)

And Wolf characterizes the "nameless virtue" of taking responsibility as involving "living with an expectation and a willingness to be held accountable for what one does, understanding the scope of 'what one does,' particularly when costs are involved, in an expansive rather than a narrow way" (Wolf, 2000: 13).

I think a charitable reading of both Mason and Wolf suggests that sometimes, *thoroughly* taking responsibility for one's actions or attitudes does not only include offering justifications, or excuses, or apologies, or feelings of remorse, or a willingness to be blamed for them. Rather, taking responsibility involves *owning up* to them (in my stipulative sense) as well. Owning up to your

actions and attitudes involves, roughly, (1) avowing them as one's own; as well as (2) a commitment either to change them, or to act in accordance with these avowals; accompanied by (3) some (veridical) sense that you *can,* indeed, follow through with those commitments.

Hence, by owning up to our attitudes, we do not only acknowledge our potential liability and blameworthiness for them, but also view them as something that *is* (rather than merely was) up to us. Something that we are in a unique and privileged position to either alter or to stick to. My proposal is that this is precisely what assuming the deliberative stance with respect to our attitudes by interpreting them amounts to. As such, assuming this stance blocks the kind of alienation from our evaluative-motivational makeup that we may experience by learning about it in the theoretical mode.

The transition from a backward-looking, third-personal, theoretical attitude to a forward-looking, first-personal deliberative attitude is possible because commitments are both active and, importantly, projective. In taking the deliberative stance with respect to our affective feelings, and interpreting them as specific emotions, directed at determinate objects, we provisionally commit ourselves to behaving in specific ways. These include promoting the welfare or integrity of our cared-for object, believing that it is reason-giving for us, and referring to it when providing justification for some of our actions in the context of responsibility practices.

Lastly, I said that assuming the deliberative stance with respect to our attitudes is *in the usual case* owning up to them. But there are cases where our self-interpretations are highly diffuse or generic. The Commitment Account can help us see why coming up with such interpretations for one's SoI is often insufficient for the beneficial regulatory effects that people with expansive EG and AL capacities enjoy. Interpreting one's emotional comportment as, for instance, a feeling of *general badness* is precisely *not* committing to care about anything in particular. It does not

project a pattern of specific, actionable behaviors and attitudes. Even if such a diffuse or non-specific interpretation is given in the deliberative mode, it amounts to keeping one's SoI in an indeterminate, inchoate state. The interpretation is *fitting* of the SoI, but trivially so. It is, simply put, an interpretive failure.

8. Conclusion

In the usual case, when we interpret our affective reactions, which constitute our SoI, we commit to caring about something in particular. This means, first, that by interpreting our SoI we render it *more determinate*; we *fix the content* of our multiply-interpretable affective reactions, and we fashion them into patterns that better resemble cares proper. This is why our affective reactions are impish. Voluntary acts of interpreting them *can indeed* cause them to change in ways that render them more akin to the content ascribed to them via interpretation. Second, because such acts of self-interpretation are also self-commissive, they carry with them important benefits for the self-interpreter's mental-health and wellbeing. Most importantly, by self-interpreting one's SoI in the deliberative mode, one internalizes her own affective reactions, and accepts responsibility for them. When we take our affective reactions as expressions of the values, cares, and concerns to which we have deliberatively committed the question: "do I want to be this way?" comes alive for us as practically relevant. As crucially up to us.

By way of concluding, I want to alleviate, at least tentatively, a potential discomfort some readers might have regarding my account. I haven't given explicit consideration to the invaluable role that *others* play in our commissive acts of self-interpretation. But, as a growing number of authors forcefully argue (Dover, 2022, 2023; Westlund, 2023; Yao, 2023), we often cannot (and in some cases, we ought not) self-interpret *all by ourselves*. We are intimately, essentially intermingled with others, and this intermingling includes self-constitutive hermeneutical activities

of the kind I have dedicated this paper to exploring. I want to accept this highly pertinent point wholeheartedly. I think others play several pivotal roles in enabling us to self-interpret at all. In fact, I think the acts of self-interpretation are *primarily dialogical*. First, acts of self-interpretation are *more successfully commissive* when they are addressed to another. But more importantly, I think that our capacity to self-interpret in a monological or solitary mode is etiologically dependent on our capacity to do so in dialogue. While I cannot elaborate on this point here, I intend to do so in future work.

Chapter 4. ENGAGEMENT, RESONANCE, AND THE AESTHETICS OF ETHOS

Out of the night that covers me,
Black as the pit from pole to pole,
I thank whatever gods may be
For my unconquerable soul.

In the fell clutch of circumstance,
I have not winced nor cried aloud.
Under the bludgeonings of chance
My head is bloody, but unbowed

Beyond this place of wrath and tears
Looms but the Horror of the shade,
And yet the menace of the years
Finds and shall find me unafraid.

I t matters not how strait the gate,
How charged with punishments the scroll,
I am the master of my fate,
I am the captain of my soul.

(Henley, 1920)

Just say if I'm disturbing you,
he said as he stepped over the threshold,
and I'll leave immediately.

You do more than disturb me,
I answered.
you turn my whole existence upside down.
Welcome.

(Kilpi, 2014)

1. Introduction

A growing theme in recent discussions of transformative experiences involves a critique of the fundamental assumptions and motivations that have been shaping key debates in this literature. According to Vida Yao (2023), for instance, questions about the potential rationality of a decision to undergo a transformative experience draw their rhetorical force and urgency from a lurking solipsistic ideal of "self-realization;" and Daniela Dover (2022, 2023) has suggested that the possibility that others may influence us into undergoing a transformative experience is worrisome

only insofar as we are committed to what she calls an *ethos of authenticity.* In turn, both authors forcefully argue against a conception of the good life that is narrowly guided by such sensibilities. We have good reasons, they suggest, to reject as a regulative ideal a life that is unqualifiedly characterizable as rationally autonomous, self-governed, authentic, or independent from other people's influence. Accordingly, they recommend a reorientation of debates regarding transformative experiences in light of a complementary ideal of human relatedness or receptivity, which I will call *the ethos of connection.*

While I agree with Dover and Yao that adopting an ideal or ethos of rational authenticity can be particularly conducive to certain worries and anxieties (worries and anxieties that have been dominating the transformative experience literature), it can do much more than that. I argue that such an ethos incorporates an orientation towards a specific *aesthetic* value—*engagement value*—that has been left unexamined by authors working on transformative decision-making. On the other hand, I claim, the competing or complementary outlook that seems to underlie Dover's and Yao's critiques is shaped to a large degree by attunements toward a different kind of aesthetic value—which I dub *resonance value*. The aim of this paper is to uncover these values as powerful motivations for adherence to, respectively, the ethos of authenticity and of connection, and to ask whether both values can be achieved in contexts of self-transformation and maintenance.

Section 2 of this chapter sets the stage. It reviews Paul's worries regarding the putative (ir)rationality of decisions to undergo transformative experiences; and Dover's and Yao's views, according to which such worries only come alive for a person with a tacit or explicit adoption of the ethos of authenticity. Section 3 strives to extrapolate a relatively precise notion of *ethos* that can be gleaned from Dover's use of the term, by utilizing Camp's (Camp, 2013, 2015, 2019, 2025) construct of "perspective." Ethoses, I will suggest, are *perspectives on the good life*. Equipped

with this characterization, in section 4, I aim to articulate the alternative ethos—the ethos of connection—that Dover and Yao (and others) appear to endorse.

One important component of ethoses as I understand them is an orientation towards different *values*. In section 5, I argue that, apart from promoting certain characteristic *anxieties*, the ethos of authenticity also orientates us to the value *of engagement* in certain self-forming activities. If this is correct, then the way Dover and Yao represent that ethos is incomplete. They do not give sufficient attention to the unique aesthetic goods that the internalization of the ethos of authenticity harbors. On the other hand, as Section 6 explains, the competing or complementary ethos of connection sensitizes its adherents to a different kind of aesthetic value—*resonance value*. The value that inheres in allowing our own subjectivity to be shaped, molded, and stirred by the world and others. The tension between the two perspectives—between the ethos of connection and authenticity—boils down to more than the tension between succumbing to and overcoming certain anxieties.

The chapter concludes (in sections 7 and 8) by asking whether the tension between engagement and resonance can be transcended. Specifically, are there certain self-forming activities that can provide us with both engagement and resonance value? While I cannot provide a complete answer to this question, I attempt to highlight one specific avenue through which a positive response might be delivered.

2. Authenticity

Roughly, an experience is transformative in the sense that I am interested in if it "teaches you something new, something that you could not have known before having the experience, while also changing you as a person" (Paul, 2014, 2014: 17). Undergoing such experiences means becoming "an unknown, dramatically changed self" (51). In the wake of Ullmann-Margalit (2006),

and even more vigorously after the publication of Paul's massively influential book, much ink has been spilt on discussions regarding the putative rationality of a decision to undergo (or abstain from undergoing) such experiences. How is it rational, from a decision-theoretical perspective, to elect to undergo a transformative experience given (1) that we cannot know in advance what undergoing the experience is like; and (2) that our preferences prior to the transformation are bound to change in significant ways after the transformation?

The threat to standard models of rational decision-making posed by (1) is that, according to such models, to rationally decide between possible outcomes, agents must determine the value they assign to each of these outcomes. But (1) means that, from their own first-personal, non-fungible perspective, agents are in principle ignorant about the value a potential transformative experience would harbor for them. In other words, they cannot determine how subjectively valuable the outcomes of such transformations would prove to be. And this inability means that agents cannot rationally compare the values of undergoing a transformation and of abstaining from it.

To exacerbate matters, (2) introduces another layer of ignorance that agents who face *transformative decisions*—decisions whether to undergo a transformative experience—must grapple with. The preferences that guide an agent's deliberation prior to a transformative experience might be radically different from her preferences after it. Even if the agent *could* (as Paul says, *per impossibile*) assign a specific value to undergoing a certain transformative experience at the time of deliberation, this value may change in profound ways from her perspective after the experience. "Transformative choices, then, ask you to make a decision where you must manage different selves at different times, with different sets of preferences" (48). And

standard decision-making models are silent with respect to which evaluative perspective, which of her temporally situated "selves," a decision-maker ought to privilege.[40]

In short, according to Paul—and many in her wake—agents faced with transformative decisions cannot rely on their own evaluative outlooks when deliberating about such decisions. This is the case both because these outlooks are principally partial, and because they are bound to change in thoroughgoing ways. As Paul summarizes, in transformative decision-making:

> The natural and appealing notion of living your life by regularly engaging in a process of reflective deliberation, considering a range of acts with various possible outcomes, and choosing the acts that best realize your hopes and preferences for your future, cannot be rational in the way we prereflectively thought it could be. (2014: 112)[41]

This chapter does not aim to contribute to the bustling cottage industry of hashing out the rational credentials of transformative decisions. Instead, I want to follow a thread that's been laid down by Daniela Dover and Vida Yao, and that concerns the underlying motivations for philosophers' interest in these rational credentials. According to Dover and Yao, the fact that so many philosophers seem preoccupied with transformative experiences' threat to our capacity to live both rationally and authentically is rooted in deep-seated fears that ought to be scrutinized—and at least partially rejected. I think Dover and Yao make a critically important point, but I also think that the roots of our interest in such questions strike deeper, or wider, than fear. Before attempting to excavate these roots, I want to dedicate some time to rehearsing Dover's and Yao's contentions.

Dover, for instance, highlights that questions regarding our possibility to decide rationally and authentically between transformative outcomes present themselves as pressing or important

[40] However, see, e.g., Reuter and Messerly (2018), and Pettigrew (2015) for critiques of Paul's conclusion that transformative decisions cannot accord with the normative standards of rationality.
[41] This conclusion mirrors Ullman-Margalit's suggestion that "very 'small' decisions on the one hand and … very 'big' decisions on the other" pose challenges to standard decision theory (2006: 157).

only given a certain underlying *ethical* conception of what humans are, or ought to aspire to be. Specifically, Dover identifies what she calls an *ethos of authenticity* that serves to orientate thinkers in this literature and to sensitize them to questions regarding the rationality of transformative decisions. She identifies this ethos rearing its head in Paul's work in passages like this:

> You are expected to take charge of your own destiny. You chart your future, deliberating and reflecting on who you really are and what you really want from life…. You live an authentic life by faithfully modeling your preferences, and you live a rational life by matching your choices to these preferences. Rational authenticity, then, is hewing as close as you can to the kind of life that best realizes your dreams, hopes, and aspirations. (Paul, 2014: 105)[42]

Dover suggests that, while the deep motivations for Paul's inquiry are at times vague or imprecise, "we can all recognize the general outlines of the moral vision Paul is invoking." According to this vision, "what it is to be good at living is to be good at taking charge of your own destiny and pursuing your own dreams" (2023: 3).

Yao, too, identifies similar sensibilities in what she describes as Paul's "commitment to the value of *rational mastery* over one's life" (2023: 2, emphasis original). Like Dover, Yao articulates a theme that runs through Paul's book, and that involves a preoccupation (an obsession?) with ensuring that, in living your life, you "proceed as rationally as possible" (Paul, 2014: 1), or that you "choose the smartest option" (2), in order to satisfy as many of your deepest or core preferences, and hence live authentically. But instead of celebrating the felt need (that at least some of us experience) for such rational mastery over the course of our lives, Yao suggests that it stems from "anxieties, which are then exacerbated by the particular rationalistic ideal that

[42] Paul herself admits that such statements are representative of our "predominant cultural paradigm" (Paul, 2014: 25, 2015: 12).

Paul assumes" (2). And like Dover, Yao is concerned with reducing the anxieties that such rationalism brings with it rather than catering to them.

Both authors, then, think that the moral vision that configures Paul's work (and much of the literature that engages her work) places at its center a "self who should be 'armed' against the world in order to maximize the satisfaction of her preferences independently of the potentially coercive influence of others" (Yao, 2023: 4). A self who is (regrettably) malleable, susceptible to corruption by others and the world-at-large; a self who, simply by living, is prone to change in ways that undermine her faithfulness to who she *really is*. Such a self is invariably faced with the daunting task of "staying true" to herself (cf. Dover, 2023: 5), and her armamentarium—her line of defense against the potentially corrupting influence of her environment (social or otherwise)—must include first and foremost her rational will.

Tellingly, both Dover and Yao seem to think that the ethos of authenticity draws much of its appeal from characteristic worries, anxieties, or even neuroses that philosophers in highly individualistic societies are likely to exhibit. If you are prone to think about your agency, or dignity, or even *humanness* as dependent upon a certain kind of integrity, and if you think that achieving this integrity is only possible by sufficiently partitioning yourself off from exogenous influence, you are bound to develop a certain fear of "contamination by others" (Dover, 2022: 209). Similarly, if you take your deepest or most authentic *self* to include only those parts of you that are under the sway of your rational powers, then experiences that threaten to wrangle such parts out of your control are bound to be thoroughly unnerving for you. As Yao illustrates with Callard's self-report about her experience of falling in love:

> Agathon says Eros is soft and delicate, because he only makes his home in the tender regions of the most soft-hearted people. That doesn't describe me. Nor does it describe Ester [the protagonist of two novels by Lena Andersson] …What resonates most with me in Andersson's novels is the theme of confusion to which she returns again and again: "The

> worst part of all was not understanding this thing she was in the midst of, this thing that had her in its clutches. There is no pain like the pain of not understanding. (Callard, 2022, in Yao, 2023:11)

Now, as I'm sure you've noticed, both Dover and Yao invoke several heavy-duty theoretical terms to characterize the philosophical-anthropological image that they are skeptical of, and that they think is recognizable in Paul's work: autonomy, authenticity, rationality, self-governance, independence… all of which have been used in a variety of ways, by a variety of authors, and for a variety of (sometimes conflicting) purposes. I do not wish to try and reconstruct biconditionally-articulated definitions of what their specific usage of such terms amounts to, most importantly because I think it would be a mistake. As Dover herself puts it, what she aptly describes as an *ethos* of authenticity is "no more than a collection of hazy representations whose precise normative implications, if any, remain to be specified" (2023: 3). I happily accept that this 'ethos of authenticity' as well as what Yao refers to as the 'ideal of rational mastery' tap into a vague, yet recognizable *philosophical orientation*.

Still, I think there is an informative and more precise interpretation of what such orientation or *ethos*, in this context, may amount to, which would prove useful for the discussion to follow. Specifically, understanding what ethoses *are* will help us see that they consist of more than fixed characteristic descriptions of the world and people's place in it. So, before articulating the alternative ethos I take Dover and Yao to be recommending, I wish to take a brief detour to spell out this interpretation.

3. Ethos

My suggestion utilizes Camp's notion of *perspectives* (Camp, 2013, 2015, 2019, 2025). According to her, a perspective is "an open-ended disposition to interpret whatever one encounters within the perspective's target domain, by *noticing*, *connecting*, and *responding* to information in an intuitive way that is partly but not entirely under one's voluntary control" (2025: 23).

Perspectives are propensities to synthesize and update complex, holistic, multi-dimensional, and intuitive representations, which Camp calls "characterizations." Camp's go-to example for characterizations are social kind stereotypes like "woman" or "Jew" (2019: 19, 2025: 23), but whereas stereotypes are often culturally uniform, characterizations "can be restricted to a sub-discipline, a particular conversation, or an isolated individual" (2019: 19).

An important contrast class for characterizations are *concepts*. Like concepts, characterizations are cognitive-representational structures that underlie thought, talk, and action, but they differ from the former in several key ways (Camp, 2015, 2019). First, while concepts abstract away many features of their targets, characterizations are "informationally, experientially, and affectively rich" (2019: 20). Among other things, they tend to associate their targets with emotionally charged and motivationally potent contents. Second, while concepts have a relatively simple structure "of deductive and material inference" (20), characterizations attribute a range of diverse, non-extension-determining features to their target, which ground some of the ways in which that target *matters* to an agent. They can determine, for instance, which features of the target are salient—more prone to capture and guide the agent's attention—and which are central for the way the agent thinks about the target (2025: 23-29). Thirdly, where concepts are relatively stable, in the sense that they allow us to represent the same content on multiple occasions, characterizations are *malleable*. They are prone to change and be changed by our occurrent "ethical and emotional responses, as well as judgments about causal structure and statistical probability" (2019: 22).

Importantly, characterizations need not be reflectively endorsed by their agents. In fact, as Camp suggests, a characterization can remain intuitively compelling to a person, and even be deployed by her, despite the fact that she may reject its aptness, as in certain cases of implicit racial

bias, or unwelcome insinuations (2013). As Camp summarizes, "[c]haracterizing, as the complex intuitive construal of a subject, is … in general partly but not entirely under voluntary control" (2019: 23).

If characterizations are rich, action-guiding, intuitive ways of construing a target, then *perspectives* are *dispositions to characterize*. They are propensities to form and activate certain characterizations across different contexts. As such, perspectives are not themselves attitudes or representations, but "modes of interpreting whatever situation an agent encounters within the perspective's domain of application." As Camp succinctly puts it, they are "*tools* for thought rather than thoughts themselves (2025: 29, emphasis original). And just like characterizations, they are often evaluatively rich, intuitive, and not necessarily endorsed or avowed by their agents.

To get back to the discussion at hand—and to gain a better handle on what perspectives are—I suggest we think of *ethoses* as *perspectives on the good life*. A personal or communal ethos is a disposition to construe certain aspects of the good life (good for the person or community) in ways that are open-ended, complex, and intuitively compelling. It is a propensity to deploy certain ways of thinking about, inter alia, the subject, trajectory, and conditions of a good life; and of the activities, goals, virtues, and values that feature in it.[43] So, for example, the ethos of the Zionist youth movements of the early 1900s included dispositions to characterize the subject of the good life as the *New Jew*, a self-reliant, physically competent, and geographically-rooted figure, who stands in opposition to the exilic, passive Jews of the eastern European diaspora. This ethos also included propensities to construe the good life as centered around agricultural work, manual labor,

[43] Ethoses, then, might be a part of—or complement—what Russell, following Murdoch (2001), calls "soul pictures." That is, 'images of man' that are "(1) *holistic* , (2) *value-rich* and *psychologically realistic* (for Murdoch these hang together), (3) *reflexive* , (4) *therapeutic* , and (5) makes use of *imaginative* language" (Russell, 2026: 302).

and physical exercise; and tendencies to think about that life as geared towards facilitating a cultural and national revival of the Jewish people in its ancestral homeland.

Similarly, I think it is helpful to think about the *ethos of authenticity* that I take Yao and Dover as aiming to flesh out as a perspective of just this kind. That is, as a set of dispositions to characterize the good life (for the characterizing agent), which are prevalent among many members of our contemporary philosophical community. This perspective includes, roughly, dispositions to characterize the subject of the good life as a self-sufficient, authentic individual who displays rational mastery over her life's trajectory. An agent who can reflect on what different futures might be like for her, and who "use[s] this reflection … to authentically determine [her] preferences about [her] future, and thus to decide how to rationally act in the present." (Paul, 2014: 107). Accordingly, this ethos includes dispositions to construe certain activities like rational reflection, deliberation, and planning as particularly valuable for the good life; and certain occasions—like value-changing, uncontrolled, non-foreseeable transformative experiences—as potentially threatening to such a life, in part because they undermine an agent's capacity to meet the constitutive standards of the aforementioned activities.

In line with these characterizations—including their affective and motivational charge—the ethos of authenticity also involves dispositions to be particularly moved by the typical "anxieties of influence" that both Dover and Yao call out. As I will soon argue, while I concur with Dover and Yao that this ethos can include and be reinforced by certain anxiety-inducing construals of threats to a good life, I think that it also involves some deeply positive aspects that we ought not neglect. In section 5, I spell out in detail one such positive aspect: its capacity to attune its adherents towards a specific kind of aesthetic value. Indeed, a key point of this essay is that each of the ethoses we're examining motivates us to engage with different kinds of aesthetic

value. Before I can tell you what kinds of value these might be, I need to briefly hash out the competing or complementary ethos that Dover and Yao appear to me to try and motivate.

4. Connection

Both Dover and Yao urge a reconsideration of the ethos of authenticity. Tellingly, they both do so by spelling out in detail activities they take as particularly valuable for the realization of a good life, and which go against the characterizations of the values, goals, ideals, and actions that the ethos of authenticity motivates. Recognizing the value of such activities, they argue, might encourage readers to *welcome* and *celebrate* others' influences on us rather than bemoan them. Thus, it seems to me that both authors advance a competing, or complementary ethos, an alternative *perspective* (in Camp's technical sense) on what the good life amounts to.

Dover, for instance, points out that, to have intimate, elective relationships, we must be both intelligible and amiable to our significant others: "we have to become and remain sufficiently attractive (in a broad sense) to others for them to elect to engage in elective relationships with us in the first place" (12). Subverting Korsgaard's (1996) famous notion of practical identity, Dover argues that, if I am to be your friend, I must develop "a description under which I value *you* and find *our relationship* worth preserving and pursuing" (2023: 15) To help me develop and maintain such a relational conception of you, "you will… for the most part unconsciously—endeavor to be, not who you think I *want* you to be … but rather who you think I *take* you to be" (16, emphases original). Following Bagley (2015), and Cocking and Kennett (1998), Dover argues that "valuable activities, emotions, practices and relationships—such as love, friendship and conversation" (2022: 219), are predicated on "letting one's rational guard down in order to admit the other's influence" (221). To enter into such committed relationships we must, at times, make ourselves vulnerable to others in ways that often involve flaunting the cannons of instrumental rationality.

Yao too thinks that some of the most precious—and thoroughly transformative—experiences that humans can undergo essentially involve a kind of thoroughgoing, irrational susceptibility to external influence. She takes as her paradigms cases of being passionately, erotically in love:

> falling in love … is to see oneself as now radically incomplete and imperfect, compared to the relative completeness and perfection attainable if only one were united with the beloved. And it is to incorporate the beloved *into* one's new conception of who one is: to have one's self "adulterated", and "mixed up" with another person. (8)

Not only does the lover's conception of herself drastically change to accommodate the beloved. The at-times violent, shocking, *penetrating* nature of erotic love precisely undermines one's status as a rational agent:

> love conflicts with certain defining marks of standard views of both practical and epistemic rationality: it introduces incoherence and conflict into both one's preferences and actions, it provides an aspiration that is itself not grounded in a set of autonomously-determined or reflectively endorsed values, it has a tendency to project qualities onto the object of the beloved that may not in fact be there, and it itself is generated in ways that are not rationally justifiable. (9)

The point, if I understand it correctly, is not that external influence necessarily goes against one's status as a rational agent, but that some of the most precious experiences that a subject can undergo *are* experiences of being externally influenced in rationality-undermining ways. So, on what I identify as Dover's and Yao's alternative ethos—call it the *ethos of connection*—the kind of practical questions that ought to be asked given our substantive (perhaps constitutive) malleability, impressionability, or porousness should not be directed at understanding how to *restrict* other people's influence on us. Instead, they should aim at making sure that such influence is valuable, or worthy, or meaningful:

> Rather than asking, 'Am I too open to influence?', what if we went back to asking, just like our parents did about us back in the day, 'Is this person a good influence?' Rather than asking, 'Is this what I really want?' or 'Will I still be me if I do it?', what if we simply asked, 'Should I do this?' Rather than asking, 'Am I truly myself around this person?',

> what if we instead asked questions like: 'Do I actually enjoy being around this person? Does this person make me feel free and alive, or dull and constricted'? (Dover, 2023: 19)

In fact, we may think that some of those practical questions should be geared towards helping us become *more* impressionable, or receptive to others (including by foregoing, at times, the cognitive procedures and protocols that instrumental rationality recommends). As Yao suggests, resisting the ideal of rational mastery over our actions, our lives, and who we are can make available to us "[m]eaningful pleasures and intensities, deep intimacies with and attachments to the world that we may experience when we are able to be erotically open to them" (18-19).

The ethos of connection, then, includes dispositions to think about the subject of the good life as a person who does not obsess over self-control, and who is happily willing to "abdicate her interpretive sovereignty" (Dover, 2022: 210). Furthermore, it involves propensities to relish the fact that we are likely to change in multitude, unpredictable ways thanks to others; it urges us to think about human impressionability as something that ought to be cherished; and it encapsulates a disposition to think about intimate interactions as particularly valuable constituents of the good life, not because they are ways of getting at what we truly are deep down, but because they are, among other things, ways of relaxing our need for personal integrity.[44]

Now, a neat thing about perspectives is that a subject can have more than one. People can and often do change, or alternate between perspectives on a certain target domain, in part thanks to external cues. Perhaps as a testament to my own impressionability, I admit that, when reading Yao and Dover, I find it hard to resist the thought that they give voice to something precious—coarse-grained and nebulous, perhaps, but precious all the same—that has often been neglected by our philosophical culture. I also confess that I feel the subtle, yet bitter, bite of shame when I

[44] I think that Carrol Gilligan's work can be read as an important formulation of (at least some aspects of) this ethos. Yet, I also think that her work points to the idea that connection does not necessarily conflict with some highly plausible standards for rational agency (Gilligan, 1993).

recognize in myself the kinds of "anxieties of influence" that they deftly describe. Still, I am tempted to think that the ethos of authenticity has something going for it that is not ultimately rooted in anxiety about one's fragile integrity, or "solitude" (Dover 2023: 6), or "aloneness" (Yao, 2025). Of course, other people may take advantage of our permeability and exert very real—and malicious—influence on us (as described by, e.g., Tsai, 2014). And undoubtedly, we *should* be worried about and safeguard ourselves from such bad influences. But there are also worthwhile, positive, "meaningful pleasures and intensities" that the ethos of authenticity compels us to pay close attention to.

At the same time, I think Yao's and Dover's attempts at centering the all-too-often marginalized ethos of connection—as well as their recommendation of formulating new questions regarding transformative experiences—are absolutely on point. We *can* be quite myopic in our concerns for rationality, or control, or integrity, and there *are* more valuable discussions to be had if we relax these concerns' grip on us. In the remainder of this essay, I want to characterize further the unique *pleasures and intensities* that both the ethos of authenticity and the ethos of connection might encourage us to notice and to seek out. Hopefully, these characterizations could help us formulate a more precise answer regarding why *both* these ethoses can seem (to some of us, at least) positively compelling. To make headway, it will be useful to move away, at least for a while, from the more poignantly *ethical* dimensions of either ethos, and to concentrate attention on their *aesthetic* aspects.

5. Games

To do this, I wish to consider two classes of agent-involving happenings that humans are prone, occasionally, to find profoundly pleasurable and intensely meaningful. These two classes, I should point out, are not jointly exhaustive of the kinds of happening that can and should feature

in a good life. However, I hope that like me, you will find some members of each class both personally gratifying and generally valuable.

The first class I want to attend to involves activities like *playing chess*, *writing a haiku*, *solving a philosophical puzzle*, *running a marathon, planning a birthday party*, *coming up with a new recipe,* and *climbing El Capitan*. Especially when we think about such happenings as making valuable contributions to a good life, I suppose we tend to view them as appealing partially because they pose *agential challenges*. They consist of problems or obstacles that agents should solve or overcome through an *active form of struggle*. Such activities can be thought of as members of a set that generalizes what Nguyen calls *striving play*. In striving play, agents adopt the official, stated goal of the activity they are pursuing, but this is not the *point* of the play. Instead, adopting the goal of the activity allows the player access to the value inherent in the struggle to overcome the challenges that the game, or game-like activity furnishes. This value partially originates from exerting control over, or influencing, one's body, mind, and environment to meet the demands of one's situation. Experiences of striving play are "experiences of acting, deciding, solving, and doing—not only of appreciating the movement or solution, *but of originating it*" (Nguyen, 2019: 430, my emphasis).

Striving players must try to *win*—to achieve the goal of the activity—but winning isn't the *point* of playing. Rather, the point is to realize the value of *doing*, or *performing*, or *engaging in* the activity. Thus, striving play involves a kind of motivational inversion: "[i]n normal practical life, one takes the means for the sake of the ends. In striving play, one selects the ends for the sake of the means it puts one through" (Nguyen, 2020: 1143). Striving players modulate their agency in response to the demands established by the game's architecture. This modulation allows players

to attain, or experience, or embody *engagement value*. To mention one of Nguyen's favorite examples,

> The aesthetics of climbing is not only an aesthetics of the climber's own motion, but an aesthetics of how that motion relates to the rock. It is not only that my movement is elegant, it is that my movement is elegant as a solution to a particular puzzle (2019: 431)

Of course, it is lovely and wonderful to *attain* one's goals but as Nguyen, I think, convincingly argues, the value agents may derive from game-playing is not wholly reducible to the value of *achievement*. Instead, there is final and aesthetic value in exerting one's powers and abilities in different ways in the context of goal pursuit, regardless of whether one succeeds in achieving that goal. Processes of *modulating* or *manipulating* our own agency while engaging with difficulties give rise to aesthetic qualities that are appreciable or worthwhile in themselves. That is why, for instance, we can appreciate games not just for the sake of their difficulty, but for the sake of the *specific ways* in which they demand we operate to overcome them (see Nguyen, 2019: 432-3).

Importantly, *bona fide* instances of playing (what Ridge, 2022 calls "play full-stop") don't seem to be the only activities that are particularly suitable for the realization of this kind of *engagement value.*[45] In what follows, it will be useful to have a term to refer to such activities, whether these are readily recognizable as cases of play or not. I will use the crisp "games," with the caveat that this label serves as a technical term of art. That is, it is not meant to delineate all or only activities that could be labeled as a game in ordinary language use.[46] Instead, I will use the term to designate the class of activities that generalizes Nguyen's notion of striving play: activities

[45] Both Nguyen (2020, 2023) and Riggle (2024a, 2024b), for instance, argue that practices of aesthetic appreciation also seem to exhibit a motivational structure that is akin to games, and that renders such practices especially conducive to the realization of engagement value.

[46] As we are well-aware (thanks to Wittgenstein's influence), "games" is a paradigm case for a linguistic category that seems particularly resistant to definitions given in terms of necessary and sufficient conditions.

wherein we take on certain challenges (at least partially) because taking on these challenges is prone to facilitate or realize the value of engagement. Games, in the technical sense that I am trying to characterize, are sites of engagement value.

Now, I claim that, apart from the peculiar anxieties that it may cultivate, the ethos of authenticity can also orientate us towards the kind of value that we may readily find in games. Adopting the ethos encourages us (probably unconsciously) to place a high premium on and to become particularly attuned to *engagement value* in contexts of self-formation and self-maintenance. Characterizing (in the technical, Campian sense) such processes as aiming at "overall rational unity" (Rovane, 1998, 8), or at "organiz[ing] ... experience in an ongoing, self-reflexive narrative" (Schechtman, 1996), or at "faithfully modeling your preferences, and … matching your choices to these preferences" (Paul, 2014: 105) means thinking about them as delineating certain agential problems or challenges. These challenges require that agents seriously engage their introspective, narrative, and hermeneutical abilities to identify and contextualize the different attitudes that populate their psychologies; and their deliberative faculties and rational will (and, perhaps, also their emotional makeup) to commit to certain cares, values, and projects. In other words, it requires that we modulate our agency in characteristic ways in our struggle towards integration or individuation.

But this means that the ethos of authenticity encourages us to view or to frame self-formation as a kind of *game*, or *game-like activity*, with its own constitutive rules, aims, and constraints. Consider Paul's gloss of how rational decision-making ought to proceed:

> To choose rationally, you determine the approximate value of each relevant outcome, you determine the weighted value of each outcome given the probability of the world being such that the relevant state would occur, and then use this information to estimate the expected value of each act, choosing the act with the highest expected value. (2014: 75).

I invite you to view this passage as a list of *game instructions.* To me, it reads very much like a suggestive rundown of what playing the game of rational self-making amounts to. In such a game, your goal consists in something like making yourself *by* yourself: "you want to determine *your* preferences in terms of *your* expected subjective values" (2014: 109. emphases original), while taking into account the constraints imposed by the normative standards of rationality.

If that is true, then while rational, authentic self-mastery might be the *goal* of the self-making activities that different versions of the ethos of authenticity highlight, self-mastery oughtn't be the *point* of these activities. Rather, the point is engaging, or *immersing* oneself in such self-making activities. It's the deliberation, interpretation, planning, imagining, and committing that rational, authentic self-making necessarily involves.[47] Indulge me, if you will, and consider Yeats's short and beautiful *The Lake Isle of Innisfree:*

> I will arise and go now, and go to Innisfree,
> And a small cabin build there, of clay and wattles made;
> Nine bean-rows will I have there, a hive for the honey-bee,
> And live alone in the bee-loud glade.
>
> And I shall have some peace there, for peace comes dropping slow,
> Dropping from the veils of the morning to where the cricket sings;
> There midnight's all a glimmer, and noon a purple glow,
> And evening full of the linnet's wings.
>
> I will arise and go now, for always night and day
> I hear lake water lapping with low sounds by the shore;
> While I stand on the roadway, or on the pavements grey,
> I hear it in the deep heart's core. (1889)

The poem, I think, wonderfully captures the distinctive value and joy of planning one's future in detail, down to the number of bean-rows one will cultivate. It also, I suggest, highlights how profoundly our interpretive, imaginal, and creative—indeed, our aesthetic—capacities are

[47] If I'm right, then rational deliberation might not be antithetical to certain aspects of what Gingerich calls "spontaneous freedom" (2022). Perhaps the self-making activities of rationally planning, interpreting, deliberating, and committing can in fact help us "experience ourselves as sources of freshness and novelty" (57).

involved in acts self-making deliberation. On my reading, the speaker's recognition, for instance, that the "low sounds" in his core are the quiet-yet-persistent waves of Innisfree is made possible only with the engagement of his intellectual *and* his creative powers. And his masterful—and motivating—construal of the sights and sounds of the isle are only possible thanks to his cultivated (and remarkable) cognitive skills.

Taken this way, the problem with others' potential influence on us (when it *is* a problem) is not so much (or not just) that it threatens to corrupt or to damage us, but that it threatens to spoil the fun—the unique, intense value that we may gain from authentic self-making and self-maintenance while trying to adhere to standards of rationality, and to our prior knowledge of ourselves and our circumstances. While this perspective is often left unarticulated in relevant debates (perhaps drowned by the very real anxieties that Dover and Yao hash out), I think it does show up in certain places in the literature. Consider, for instance, what Kane describes as our "basic desire to be selves or psyches with the power to be independent sources of activity in the universe" (1996: 96). As he puts it,

> When the child learning to walk says, "unhand me, I want to do this on my own," I believe she is expressing a primordial desire that inevitably arises in creatures who attain a certain level of intelligence and self consciousness: the desire to be "somebody" whose contribution to the world is her own. That is to say, creatures of higher consciousness such as we are have an unquenchable thirst for individuality and personhood (which I think are inextricably linked). (1996: 100-101)

What stands at the root of the child's desire for "individuality and personhood" that Kane identifies, *could* be interpreted as her wish to be alone, or partitioned from the rest of the world. But I think it is better understood as a motivation to be a source of activity. More specifically, it is a motivation to volitionally shape, or mold her agency on her own. It is a desire to *try out* the difficulty, to meet the challenge without the (real or metaphorical) training wheels. To be, if you will, the "captain of her soul."

A similar sentiment, I think, can be gleaned from Akhlaghi's remarks about what he calls our "moral right to revelatory autonomy." That is, our right to "autonomously decide to discover how one's life will go and who one will become by making a transformative choice" (2023: 8). According to Akhlagi, the importance of our (presumed) right to revelatory autonomy is grounded in the moral value of "autonomous self-making:" of "autonomously making choices to learn what our core preferences and values will become." Such autonomous self-making, he argues, grants us a level of control over ourselves that is "crucial for us and others to see ourselves as *ours*elves—selves we have become at least partly through transformative choices we have made" (8).

Now, I am not convinced that this particular notion of "seeing ourselves as *our*selves" is as morally significant as Akhlaghi thinks it is.[48] But I *do* think that Akhlaghi picks up on something important when he is trying to characterize the value for a flourishing life of rationally reflecting, by ourselves, about the kind of people we hope to become. As I've argued, this value inheres, at least in part, in undertaking and following through with these forms of reflection. It inheres, in other words, in the engagement value that we realize when we play the game of rational, authentic self-making.

To be clear, I do not think that Paul, or Kane, or Akhlaghi—choose your favorite representative of the ethos of authenticity—necessarily consciously thinks about rational decision-making processes as game-like. My point is that we *could* construe them as such, and that this construal could help us gauge the unique value that such processes involve, over and above their functioning as buffers against the anxieties of influence that Yao and Dover articulate. In short,

[48] I am unsure, for instance, that my (or your) perception of yourself as the product of the *transformative choices* that *you've made* is important to ground your right to make your own future decisions, or to be free from coercion; nor do I think it's important to ground your dignity or respect. But these issues, admittedly, are neither here nor there.

the ethos of authenticity may include an inchoate disposition to care about the *engagement value* that certain self-making processes encapsulate.

6. Rides

The ethos of authenticity includes dispositions to construe certain forms of self-making and self-maintenance as valuable in a specific sense: as *engagement valuable*. I will now turn to argue that the ethos of connection involves dispositions to construe (often different) forms of such processes as valuable in a different sense.

Consider a class of agent-involving happenings, which, like games, I hope we can agree should feature in a good, well-rounded life. Paradigm cases of such happenings might include things like *receiving a massage, riding a roller coaster, getting crowd-worked by a comedian, being lulled into sleep, melting into someone's hug, submitting to a lover's erotic authority,* and *being led on a guided meditation.* To contrast them with games, let us call such happenings *rides* (remembering, again, that for our purposes, these labels serve as technical terms of art). While the value of games partially derives from the fact that they involve agents striving to achieve certain goals by actively shaping their agency, (some of) the value of rides seems to me to inhere precisely in agents' *suspension* of their own agency. Rides can be both highly exhilarating or arousing, as well as relaxing or comforting, but they all involve agents letting themselves be shaped, or moved, or handled by other people or by their environments.

Perhaps to echo Yao and Dover, it may be difficult to accurately express the kind of value that *relinquishing* our agency could encompass from within the purview of contemporary academic philosophy. Our stereotypically individualistic, rationalistic and secular intellectual positioning might have left us with an impoverished vocabulary to refer to the things that make such happenings appreciable or worthy (although the significance of surrender is a positively

pervasive theme in many religious and contemplative traditions). But perhaps we can gauge it, albeit vaguely, by again concentrating on the aesthetic aspects of such experiences.

And sure enough, many of those experiences can be deeply aesthetic. They can instantiate properties like delicacy, elegance, intensity, surprisingness, intricacy, and suspense. To bring into relief rides' contrast with games, let us concentrate on an aesthetic property that characterizes both the experience of game-playing and of ride-going: harmony (cf. Nguyen, 2019: 430). In games, harmony is achieved when "one's abilities fit the demands of the situation." That is, when "one's abilities are just right to cope with the situation at hand." This requires a sharp enough delineation between self and world, or, as Nguyen says, "self and challenge" (431). Between an agent that enjoys sufficient integrity, and a world that offers enough resistance. If you like metaphors, we might say that the harmony between self and world we experience in striving play (when we experience it) is similar to the harmony that holds between two matching puzzle pieces: between two well-defined entities that fit together by being pressed against each other.

In rides, on the other hand, harmony is achieved when a subject becomes *absorbed* or *dissolves* into her environment. When she lets things take their course, or run through her. The kind of harmony that rides aim at involves a subject who allows herself *not* to struggle in response to the pressures the world exerts on her. That is, a subject who is capable of assuming (to some degree at least) a spectatorial role with respect to herself, in a way that enables her to appreciate the ways in which she is moved and stirred by the world. If the aesthetics of games is defined by a volitional modulation of one's agency in response to one's environment, the peculiar aesthetics of rides involves an *a-volitional modulation of one's subjectivity by one's environment*.

Here is another attempt at describing what I'm getting at. Recall the particular elegance of rock climbing, according to Nguyen. As he says, it is not only that a rock climber's movement is

elegant, but that it is elegant *as a solution to a particular puzzle* (2019: 431). Compare this with the elegance of your body's motion when you allow yourself to be carried gently by the ocean's waves. Here too, I claim, your movement is not merely elegant. Instead, your movement is elegant *as a manifestation of the world's influence*. The specific movements of your body are directly derived from the movements of the waves. The former are not just *coordinated with* or *follow* the latter, rather, they are resonances or reverberations of the waves. In the water, your body is elegant as a conduit or transmitter of your milieu's activity.

Like games, many rides (although not all of them) have completion conditions. A roller coaster ride comes to an end when the train safely reaches the station; a massage successfully concludes when muscle tension is reduced; a gentle lulling felicitously terminates when baby is sound asleep. And just like games, the value of rides for the ride-goer is not exhausted by—or wholly derivable from—these conditions' achievement. A massage is not valuable just because it loosens your muscles, being lulled is not valuable merely because it is conducive to napping, and going on a roller coaster is not desirable just because it affords the satisfaction of having survived it. Analogously to games, there is something valuable in *the unfolding* of such processes (or the experience thereof). To put it in general terms, whereas games make available to the player the unique, intrinsic value of striving to achieve a goal, rides allow us the opportunity of having our own subjectivity shaped by something other than ourselves. When we go on rides, we are afforded the chance to let go, get taken care of, whisked away, moved, touched… We are afforded the opportunity, in short, to open ourselves up to the world's and to others' impact on us in highly diverse, and aesthetically profound ways.[49]

[49] I sense a potential response, according to which what I am trying to characterize here is simply *felt pleasantness*, or some broader kind of hedonic value. To stress, I do not take a stand about the metaphysics of the values that I am attempting to point to here, other than the that they are *aesthetic*. Many philosophers view aesthetic value as essentially hedonic (e.g., Beardsley, 1982; Levinson, 1996; Matthen, 2017; Schopenhauer & Payne, 1819; Stang, 2012; Walton,

Call the value that can be derived from having our subjectivity modulated by our environments *resonance value.* I claim that, just as the ethos of authenticity is prone to make us attuned to and seek out engagement value in contexts of self-formation and transformation, the competing ethos of connection can motivate us to appreciate such contexts as potential sources of resonance value.

Recall the agent-involving happenings that feature in Dover's and Yao's construals of a good life: falling in love, letting one's friends determine one's cares and concerns, being interpreted by one's intimates, or, to use Eeva Kilpi's words, having one's existence "turned upside down." The defining features of such happenings are precisely the suspension or abdication of one's agency or authority over one's identity. Moreover, these features, according to Dover and Yao, are precisely what makes these happenings attractive; and what renders a life that is wholly devoid of them lacking in some significant ways. But if what we've said about rides is true, then the ethos of connection involves a disposition to characterize valuable self-making happenings as *rides*. As such, the ethos of connection *also* involves a propensity to think about the particular value of such happenings as resonance value. Yao's comments provide a vivid illustration of the kind of attitude towards ride-like identity-shaping happenings that, I think, is characteristic of the ethos of connection:

> One may simply want to lose control, to be moved about in the world by forces that seem outside of oneself, to come to understand oneself in relation to another particular person (or other object of erotic or passionate attachment), to be unsure of where one ends and begins, to be unable to find rational justification for why one wants to do what one wants to do (2023: 15)

1993), or at least essentially related to pleasure (Gorodeisky, 2021), but there are some notable recent dissenters (Lopes, 2018; Matherne, 2020; Nguyen, 2020; Riggle, 2024a). As far as I can tell, my view of the aesthetic value of rides is compatible with these dissenters. In future work, I aim to make precise what this value amounts to, in ways that explicitly avoid its reduction to hedonic value.

Moreover, I have said that, generally, rides afford the ride-goer the opportunity of having her own subjectivity shaped by something bigger, more encompassing, or simply *other* than herself. This opportunity (and, presumably, its attendant *resonance* value) is intensified when the rides under consideration constitute processes of *self-formation.* Falling in love, developing a profound friendship, or joining a tightly-knit community are not just processes wherein you let others determine what will *happen* to you. In such processes, you allow others to have a (sometimes definitive) say about *who* this *you* will end up being. Such ride-like happenings *shape you* as a subject. In a slogan, intimate relationships are the ultimate rides.

7. Woods and Gardens

If there is one lesson to be learned from Dover and Yao, it is that loosening the "anxieties of influence's" grip on us can encourage us to care about different things than being contaminated by others in contexts of self-formation and transformation. I hope to have shown that there are at least two ways of reorientating our attention away from those anxieties when dealing with such contexts. First, by motivating a perspective that is complementary to the ethos of authenticity; and second, by trying to elaborate the ethos of authenticity itself. Doing *both* can help us characterize (in Camp's sense) two kinds of broadly aesthetic value that self-making happenings can make available to us. Some self-making processes can be more easily or appropriately construed as games—and hence more conducive to engagement value—and others might be better interpreted as rides—and hence more conducive to resonance value.

Where does this line of thinking leave us? If you are like me, you'd want to know whether these two classes of self-making processes are mutually exclusive. You'd like to know, that is, whether engagement value and resonance value (particularly in self-making contexts) are at odds with each other. Can agents undergo self-making processes that offer them the possibility of

realizing *both* kinds of value? To put it slightly more practically: can we have our cake and eat it too? Should we (or can we) hope or try to realize the values of both games and rides; or ought we, when evaluating the way we "do" identity work, acknowledge that a balance must be struck between them? I want to conclude this paper by offering a few highly tentative remarks on these questions.

First, if my characterization of resonance and engagement value is on the right track, there seems to be a straightforward sense in which these values stand in tension with each other. Engagement value, I have said, is the kind of value that we may gain from (for instance) striving play: the value of actively modulating our agency in response to challenges. On the other hand, resonance value is made available to us by what appears like the converse: by *abdicating* agency in certain ways. If that's the case, then prima facie, these two categories seem incompatible in at least one narrow sense: participating in activities that are conducive to engagement value means that we cannot abdicate our agency in a wholesale manner. Hence, the pursuit of such activities puts a cap on the intensity, or amount, or thoroughness of resonance value we can simultaneously realize, or enjoy, or attain. Assuming this impression is correct, what does it bode for our prospects of realizing *both* engagement and resonance, authenticity and connection, in self-formative contexts?

Perhaps most pessimistically, we might just have to choose—or otherwise be moved to embrace—one value over the other. Value conflicts are a constant, if unfortunate, fact of our sublunary lives, and self-making processes are no exception. This might be tantamount to saying that generally speaking, (at least in good cases) identity work comes in two flavors: it can take the form of either games or rides, with their distinctive attendant values. Ultimately, you'll have to

choose—or let something else choose—which of these values will dominate the processes through which you come into your own.

Second, relatedly, perhaps the sheer ubiquity and variety of self-formative processes means that an agent could land (either deliberately or not) on certain satisficing policies. Maybe you could strike a balance between engagement and resonance by participating in certain self-making *games* at some points, and self-making *rides* at others. In the conclusion to her (2023) Yao considers one such satisficing, "moderately rationalist" policy that may be adopted by some of her readers:

> By all means fall in love with a person or a tract of land or a piece of music, [the] *moderately* rationalist and reasonable philosopher might think, but still try your best to do so as part of an overall rational plan of life. Maybe don't do so when you are too young, but equally don't wait until it's too late; allow it to inform and color your major life decisions, but don't let it interfere with all of them; be more moderate and careful in how you feel and how you act, given the risks of immoderation and recklessness. (15)

Yet, as Yao remarks, this highly (if not perfectly) reasonable life plan does, in the final analysis, include giving up on richly compelling, and deeply valuable alternatives. In Cora Diamond words (which Yao mentions approvingly, 2023: 16):

> Lawrence's image for what Rawls calls a life-plan is a garden-plan for one's soul. And there could, within one's garden-plan, be a space for a certain amount of wilderness, a place where things do not grow in accordance with a plan, but spring up on their own. But a planned garden, even with some space to allocated wilderness, is an altogether different conception from a 'dark forest', an 'Hercynian Wood'. Seeing one's soul as a vast forest is not merely *enlarging* the space in the garden allotted by the gardener-planner to letting things go wild (Diamond, 1997: 226)

Notice, though, that just as some people might be inclined to adopt a moderately rationalist life plan, others might choose—or better, be thrown into—a "moderately irrationalist" life of wandering in inner and outer wilderness. You may plunge yourself into the turbulences of boondocking in the high Mojave, or succumb to the ecstasies of WWOOFing in Australia. Bartend in Manali for a season, then guide tours in Goa come winter. Switch lovers, careers, homes; allow your heart to break and heal according to its old brag. But maybe also try to find joy in saving for

retirement, or in planning your next van upgrade. To adopt the Lawrentian metaphor, every now and then, you might find a clearing in your Hercynian wood. When you do, maybe try planting some bean-rows.[50]

Now, such satisficing policies might be pretty good, but if they're all we can hope for, they might still be disappointing. I suspect that what makes both engagement value and resonance value especially attractive is not just the prospects of *engaging* with them, but of *immersing* ourselves in them. Of being completely devoted to, for instance, shaping ourselves on our own; or of being utterly willing to relinquish our authority over who we might become. The occasions in which we manage to do so are those in which those values truly shine. That's when we really *get* them—both in the sense of acquiring them, and of understanding what they are about.

8. Is This Some Kind of Joke?

But there might be another option. Maybe there are some self-forming processes that are *both* game-like *and* ride-like in important respects. Maybe certain ways of becoming call for a structured oscillation between activity and abdication. Perhaps some formative processes require that we both fully immerse ourselves in the game of *giving an account* of ourselves, or of rationally determining our preferences in terms of our expected subjective values; *and* afterwards (or beforehand) that we completely relinquish our authority about who we might turn out to be. If that is the case, then perhaps, in some fortunate instances, we don't have to decide. We may be engaging with both engagement value and resonance value simply in virtue of undergoing such formative processes.

One reason for thinking that this might be the case (and hence one cause for optimism) is that some mundane activities *do* seem very much like hybrids of games and rides. Take jokes, for

[50] I thank Nick Riggle for encouraging me to think about and articulate this point.

instance. According to Gimbel (2018), joking is a discursive activity that exhibits a "master and slave pragmatics" (66). Joking contexts, like lectures and court hearings, involve conversational power asymmetries between speaker and hearer. In an academic lecture hall, for example, the instructor occupies a discursively privileged position: they are expected to arbitrate between different positions, grant class-members permission to speak, challenge views, and often do most (if not all) of the talking. Such power asymmetries occur in joking contexts as well, but "what makes the joking context so fascinating is that unlike the philosophy classroom or the court of law, the power asymmetry is dynamic. Conversational power shifts three times during the course of the speech act" (68).

Telling a joke requires one of the participants in a conversation to willingly move to the conversational position of passive listener. To convince her to do so, the joke teller must generally coax the potential listener by "promising a positive result" with phrases like: 'wanna hear a good one?' Thanks to her ability to frustrate the joke sequence, the potential listener enjoys a dominant conversational position until she grants the teller permission to proceed.

But once permission has been granted, the joke teller's addressee moves to a subjugated position. "The joke teller … becomes the center of attention with the ability to speak at will and command responses of an expected sort from the listener" (69). The teller gets to instruct the listener in constructing the "mental world" of the joke, to accommodate new information, to cultivate certain expectations, and to upend them, all while keeping silent, or uttering certain conventional and highly circumscribed responses on cue (think knock-knock jokes).

Lastly, once the punchline has been delivered, the listener is again granted asymmetrical power over the teller. The former "becomes the judge of the aesthetic act. Was it a good joke? Was it funny? Was it well told? Was it appropriate? Did it expose any morally problematic beliefs on

the part of the teller?" (71). The teller's status as successful jokesmith is at the mercy of her audience.

> If the joke delivers, then the joke teller has delivered. S/he has shown the listener that allowing him/herself to be conversationally vulnerable was a good idea and that the teller can be trusted with the conversational power and thereby with other power. (72)

I contend that the dynamic power asymmetry of joking contexts makes joke-crafting and telling both game-like *and* ride-like. First, the Jokesmith must carefully and intentionally utilize her cognitive resources to construct a conspicuous, playful demonstration of her cleverness, and she must present it in a masterful way when telling the joke. In other words, the jokesmith must voluntarily engage in a particularly delicate agential challenge. Attempting to overcome this challenge can grant her the exact kind of engagement value that characterizes other game-like activities.

But once the joke is told, the teller must abdicate her agency and let her audience determine the joke's fate. She must take a step back and let things take their course, hoping that the joke receives its intended uptake. This kind of potential thrill puts this aspect of joke-telling on a par with other, perhaps more salient rides. Although I cannot argue for this claim here, I conjecture that many jokesmiths' primary motivation for engaging in comedy is precisely the resonance value that may be attained by allowing one's audience to make up their minds about one's jokes.

If this is true, then at least some ordinary discursive activities are both stringently game-like *and* uncompromisingly ride-like. As such, they seem to be able to afford their agents the chance of engaging with both resonance and engagement value. The optimistic note that I wish to end on is that perhaps some patently self-forming activities have this structure too. Perhaps *some* formative or transformative contexts must exhibit alternating game-like and ride-like characteristics to be felicitous. If that is the case, then maybe (and pace what I read as Yao's skepticism) we *can* enjoy the pleasures and intensities that are emphasized by both the ethos of

connection and the ethos of authenticity. The hope, then, is that self-making is some kind of a big joke.

Chapter 5. THE ETHICS (AND AESTHETICS) OF UPTAKE REFUSALS

1. Introduction

Many authors point to the pernicious effects that denying a speaker's self-ascribed identity components might have for that speaker. In many cases, refusing to give the intended uptake to such self-ascriptions can lead to palpable and substantive harms to speakers' autonomy, dignity, material conditions, and even bodily integrity. Given the massive harmful potential of such uptake refusals, it is perfectly understandable that their positive aspects have been largely ignored—and, in some cases, ruled out categorically.

While not denying their very real capacity to harm, in this paper, I wish to explore some of the ways in which refusing to accept speakers' self-identifications may prove to be prudentially beneficial for those speakers. I focus attention on a specific kind of identity-related utterance that regularly appears in intimate dialogue, which I will call "existential self-ascriptions" (ESAs). That is, utterances in which a speaker self-ascribes aspects of her own evaluative-motivational perspective, like her concerns, values, projects, and other related, identity-relevant features. Expressions of the form "I care about this job," "I value my privacy," "I despise laziness," and "I love you," are paradigms of ESAs, but I take it that such ascriptions need not involve explicit reference to psychological attitude terms. ESAs can include expressions (uttered in certain contexts) like "I am my mother's daughter," "I'm such a Pisces," and "I'm done playing nice."

In this paper, I argue that in addition to the harmful potential that uptake refusals may harbor, sometimes, refusing to give uptake to a certain uttered ESA can help the speaker shape their evaluative outlook in prudentially—and perhaps even morally—auspicious ways. Given that uptake refusals can be both highly beneficial in some cases, and positively destructive in others, teasing apart the conditions under which they are indeed permissible—and, perhaps, obligatory—

is a pressing question in the ethics of dialogue. I offer a few clarifications regarding what these conditions might include. The chapter concludes by proposing that, surprisingly, the mere possibility of uptake refusal makes uttering ESAs in contexts of sincere, intimate dialogue a particularly valuable way for someone to engage in identity formation and maintenance

2. Uptake: Some Preliminaries

For the purposes of this chapter, I will adopt (and, when necessary, adapt) Kukla and Lance's (Kukla & Lance, 2009; Kukla & Lance, 2023; Lance & Kukla, 2013) framework for analyzing speech acts.[51] In a somewhat Brandomian key (1994), they maintain that speech acts can be usefully understood in terms of the normative statuses—the sets of social entitlements and commitments—that enable them and that they effect.[52] Different kinds of speech act can be construed as different functions within discursive practices, each with its own distinct "inputs" and "outputs." Generally speaking, the "input" of a speech act is the set of normative statuses that entitle its performance, and the "output" of a speech act is the set of changes to normative statuses that its "proper production strives to make" (2009:28).[53] To use one highly worked-out example, the input of a wedding pronouncement—the normative conditions that entitle a person to issue it—include her socially accepted status as officiator, being in the right kind of ceremonial setting, having acquired the consent of the betrothed, etc. Whereas the output of the pronouncement—the

[51] I adopt Kukla and Lance's framework because it is particularly amenable to the kind of social and ethical concerns that I will deal with here (specifically, unlike other approaches, they do not view assertions as explanatorily prior to other speech acts). However, as far as I can tell, everything that I will say can be translated into more traditional "scorekeeping" approaches (e.g.,Lewis, 1979; Stalnaker, 2002). Thanks to Nick Riggle for asking clarification on this point.

[52] I say that Kukla and Lance's key is *somewhat* Brandomian because, while Brandom thinks that the commitments and entitlements that are relevant to determining the pragmatic kind of an utterance are primarily *discursive*, Kukla and Lance are much more permissive. As I will explain momentarily, on their account, the relevant normative statuses include far more than just entitlements and commitments to *say* certain things.

[53] Kukla and Lance's attributions of "strivings" to speech acts might raise worries of spooky anthropomorphisms, but as they say, the "striving" language is a perfectly natural way to characterize anything with a with a functional structure: "anything—a machine, a policy, an action, a vital organ—that has a function may be said to strive, by design, to fulfill this function" (2009: 16).

changes it makes to community members' entitlements and commitments—might include shifts in the couple's tax status, family relations, and obligations to attend certain dinner parties (Kukla, 2014: 442). According to this account, identifying a speech act's inputs and outputs can provide a useful way of determining its pragmatic kind.

The output of a speech act, I just said, is the difference that it makes to the social fabric of the relevant discursive community, the set of concrete changes to normative statuses that it generates. In turn, changing the community's "normative landscape" (in Owens's apt phrase, 2012) via the production of a speech act can only occur through certain acts and events of *recognition* by relevant community members. Such acts can be more-or-less explicit, as is the case when people assent to the declarative "Morticia and Gomez are married" following their wedding; and they can be implicit, as is the case when sister-in-law Ophelia finally adds Gomez to the Frump family group chat. In both cases, community members *enact* their recognition of Morticia and Gomez's marital status in response to their wedding pronouncement. In general, I will use the term 'act(s) of recognition' to refer to the concrete acts that are performed by hearers in response to a (token) speech act. The impacts on the normative landscape that are enacted through acts of recognition in response to a given speech act constitute that speech act's *uptake*. As I said, such acts of recognition *can* be covert (hearers can change their mental states and attitudes in response to a speech act), but in this paper I will be mostly concerned with the concrete, material, overt acts that hearers perform in response to an uttered speech act.

To have performative force, a speech act must secure uptake. That is, "other people must recognize [the speech act] as shifting normative statuses, so that it can have a systematic impact on people's behavioral dispositions" (Kukla, 2014: 443). If the priest's utterance of the wedding pronouncement's words fails to effect changes in the way Morticia and Gomez are treated by their

community, then the utterance did not have the output it was striving to produce. In Austinian parlance, the utterance *misfires*.[54]

To be clear, it is not obvious that all normative landscape changes must involve *speech acts*. The death of the king, for instance, is patently not a speech act (it's not an act at all), but it shifts into gear a host of normative status updates. More contentiously, you might think changes to the normative landscape of a given community could be enacted without any *acts of recognition*. Some authors argue, for instance, that the mere act of 'mentally' or covertly releasing me from an obligation to refrain from, say, φ-ing, constitutes consent to my φ-ing and hence changes the normative landscape between the two us—with or without uptake (Alexander, 1996; Ferzan, 2015; Renzo, 2022). Be that as it may, what matters for my purposes here is not whether normative landscape changes can occur without uptake, but how such changes are generated *in response to speech acts* (and which of these changes are morally permissible). In other words, I am interested in what amounts to uptake when uptake is present (and what is the moral or prudential significance of giving which kinds of uptake). To reiterate, in the framework that I am borrowing from Kukla and Lance, uptake is the totality of acts of recognition that a speech act elicits.

3. Two Kinds of Uptake Refusal

Notice that it is a result of this framework (a happy result, if you ask me) that acts of recognition encompass more than full-blown *approvals* or *acceptances* of the changes that a speech act strives to enact. Acts of recognition—acts that change the normative landscape of a

[54] Yet notice that, at least according to one common reading of Austin, the uptake of a speech act consists in the recognition of its speaker's intentions. For instance, Hornsby and Langton (1998) ascribe to Austin the view that "[c]onditions for performing an illocution typically include a hearer's recognition of a speaker's intention to perform an illocution" (80). Similarly, Maitra writes: "I gloss [Austin] as follows: uptake requires both understanding of the content expressed by the utterance, and recognition of the speaker's illocutionary intention(s)" (2009: 313) (but consult Bianchi (2021) for a different reading). The important thing to note is that Kukla and Lance explicitly reject that idea that uptake amounts to a recognition of a speaker's *intentions*, in favor of a view that defines uptake as the impact a speech act *in fact* has on social space (2009, 2013, 2023; Kukla, 2014).

speaker's discursive community in response to her speech act—can be partial or even contradictory to a given speech act's purpose, or function, or intended result. To clarify this point, I want to survey a few examples. Start with

Counterproposal
Pugsley: Let's join an expedition to climb K2!
Wednesday: How about we take a stroll to the graveyard instead?

In this exchange, Pugsley attempts to get Wednesday to adopt a certain—highly demanding—shared project, which will involve, we may assume, drastic changes to their normative positioning. Wednesday does not accept Pugsley's entreatment, but she does not outright reject it either. Instead, she comes up with a counterproposal to Pugsley's extravagant bid. Wednesday suggests that the normative landscape change—that she and Pugsley undertake a certain commitment together—but not in the way that Pugsley's speech act was striving to enact. In the technical terminology that I will be applying, Wednesday's response counts as an act of recognition. Next, consider

Disengagement
Pugsley: Let's join an expedition to climb K2!
Wednesday: Umm… I think I'll pass (makes a mental note to stay away from Pugsley)

Here, Pugsley's invitation succeeds in instituting changes to the normative statuses that obtain between Pugsley and Wednesday. But these changes are quite antithetical to the changes that he was hoping to incur. From now on, Wednesday will act in ways that will presumably amount to being *less* committed to Pugsley and to their shared projects. For our purposes, responses such as Wednesday's will also count as acts of recognition, insofar as they enact changes to normative statuses. Lastly, let's look at

Disregard
Pugsley: Let's join an expedition to climb K2!
Wednesday: Anyway, last night's horror movie was magnificent (continues talking)

In this case, Pugsley's speech act is normatively inefficacious. It fails to institute any kind of change to Pugsley's and Wednesday's normative landscape. Wednesday brushes off Pugsley's invitation in its entirety, in an instance (a relatively benign instance) of the phenomenon philosophers of language call *illocutionary silencing* (Hornsby & Langton, 1998; Langton, 1993). In illocutionary silencing, an attempted speech act fails to have any performative force at all. In this case, Wednesday's response is *not* an act of recognition. To reiterate, in contrast to the illocutionary silencing portrayed in *Disregard*, in *Disengagement* and (to a lesser extent) *Counterproposal*, the normative landscape *does* change, albeit in ways unintended by the speaker. Such cases exemplify what Kukla calls *illocutionary distortion* (Kukla, 2014). We may say that what distinguishes cases of distortion from cases of full-blown silencing is that the former but not the latter involve acts of recognition.[55]

In other words, *uptake refusals come in two flavors*: silencing and distortion. In silencing, a speech act completely fails to secure uptake; whereas in distortion, the uptake that a speech act manages to secure is different from the uptake it was striving to obtain. The main thesis of this paper is that at least some cases of illocutionary *distortion*—even distortion of highly intimate, and vulnerable speech acts—might be prudentially beneficial for the speaker.

4. Covert Exercitives

In this chapter, I'm interested in fleshing out the conditions under which it is permissible to deny uptake to a specific kind of speech act, which I've called *existential self-ascriptions*, or *ESAs*: first-personal ascriptions of identity-related features. One important and pervasive feature

[55] What should we say about cases in which, for example, a hearer simply refuses (maybe even politely refuses) a speaker's invitation? I take it that in such cases the hearer does not silence the speaker. A plausible necessary condition for an invitation's success is that it institute a commitment on the part of the invitee to *respond* to the invitation. In portner's (2004) terms, invitations update the hearer's 'to do list' by adding to it an item like "either accept or reject the invitation." So invitation refusals seem to me to be unproblematically classifiable as acts of recognition, in the sense I'm interested in. I thank Dana Nelkin for a few clarificatory questions regarding various aspects of this section.

of such ascriptions is that, while their surface grammar often corresponds to that of paradigmatic declarative sentences, their function is quite different. This, mind you, is a reflection of a general pragmatic point: the surface grammar of an utterance does not reliably indicate the pragmatic kind to which it belongs. For instance, under the right circumstances, "we will visit the Addams family" can function as an assertion, a promise, or even a threat. Indeed, according to the framework I am adopting, speech acts are partly *individuated* by their actual outputs—the concrete ways in which they impact social space.

In this section, I point to the fact that, at least in usual cases, ESAs function to enact (or attempt to enact) certain changes in how their speakers are treated by their audiences. In other words, ESAs are an important sub-class of what McGowan calls *conversational*, or *covert exercitives* (2004; McGowan, 2009).

Generally, exercitive speech acts strive to enact 'permissibility facts': facts about what is permissible or impermissible in a certain context. In a series of important papers, McGowan points out that there is a huge class of exercitive speech acts that differ from textbook cases like commands, orders, instructions, and resignations, which 'wear their pragmatic kind on their sleeve,' so to speak. Consider one of McGowan's (highly mundane) examples:

> Suppose that, when discussing Mike's dog, I say "We had a hyperactive Irish setter named Finbar who stole undergarments from neighborhood clotheslines and so we had to get rid of the dog". By introducing Finbar into the conversation, I made Finbar the most salient dog and that is why I managed to refer to Finbar with the expression 'the dog'. (2004: 100)

McGowan's utterance regarding Finbar changes which entities may be referred to with the phrase 'the dog' in her specific conversational context. Put differently, it changes (conversational) permissibility facts. This makes the utterance different from standard exercitives in two important respects. First, standard exercitives typically enact permissibility facts via the exercise of the speaker's pre-established (and often institutionally recognized) authority. Second, and more

importantly, standard—but not covert—exercitives "[express] the content of the permissibility fact being enacted" (McGowan, 2004: 96). In the usual case, issuing a command like "drop and give me twenty!" involves an overt articulation of the state of affairs that the command is attempting to achieve—as well as an indication of its pragmatic kind (via the use of the imperative "drop," for instance). This, however, is not the case with covert exercitives. In the previous example, nothing in the contents of McGowan's utterance corresponds to something like: "until further notice, you may only use to expression 'the dog' to refer to Finbar."[56]

As I stated, I will follow a significant theme in feminist philosophy of language according to which certain ESAs are highly important covert exercitives, insofar as they are attempts at enacting permissibility facts with respect to speakers' identities. That is, among other things, they constitute attempts to alter what is and is not permissible in the norm-governed activity of relating to that speaker.

For instance, Bettcher argues that certain avowals of mental states, attitudes, and other aspects of one's identity "are not reports of fact or at least not *mere* reports of fact" (emphasis original). Rather, such avowals serve to communicate one's acknowledgement of, and responsibility for the avowed features:

> In publicly avowing an attitude, the first person has in some sense staked a social claim and certified a view about their mental life on which we can "bank." In avowing an attitude, one authorizes a view of one's mental life that is then fit for circulation. (2009: 102)

Dembroff and Saint-Croix make similar claims, using their notion of "agential identity." That is, the aspects of a person's identity that she publicly externalizes or expresses—the aspects

[56] In the preceding section, I said that *any* speech act constitutes an attempt to change the normative landscape of the speaker's discursive context. Hence, it follows that any speech act is (at least covertly) an exercitive. This is not an unhappy consequence. Here, I am closely following McGowan, who argues that "any conversational contribution that invokes a rule of accommodation is an exercitive speech act in virtue of changing what is permissible in that conversation" (2004: 99). This result is not trivial because, as we'll see, ESAs are an especially *sensitive*, or *vulnerable* kind of exercitive.

that she puts forward as candidates for social recognition (2019). According to them, a person's agential identity does not necessarily overlap with her *social position*, which is the position into which people are placed when they are recognized as belonging to a certain social kind. Nor must it overlap with her *self-identity*, which is the aspects of her identity that she herself takes to define her actual kinship relations and the norms she is beholden to. On Dembroff and Saint Croix's view, publicly identifying as belonging to a certain social kind is, among other things, laying a claim to a certain agential identity. It is both *reporting* on your *self*-identity, but also demanding, suggesting, proposing, or requesting that others recognize you as having the identity you externalize and that they proceed to enact this recognition (see also Ásta, 2018: 19).

Building on Dembroff and Saint-Croix, Kukla and Lance suggest that "claiming an agential identity is not an act of making a traditional truth claim … Rather, it is a petition for a certain normative social position, which calls for recognition" (2023: 1147). In calling you my friend, for example, I am not trying to convey information about the state of our relationship. Rather, I am trying to *establish* or *reinforce* such a relationship: "I am not telling you what empirical properties you or I have, but rather letting you know how I will treat you and how I expect you to treat me." In section 6, I will pursue this dialectical route. I will argue that often, we lay claims to agential identities without having fully settled our *self*-identities. Acknowledging the possibility (and, potentially, pervasiveness) of such cases has interesting ramifications for the ethics of dialogue.

5. The Dangers of Uptake Refusals

Before this, I want to spell out some moral implications of the view according to which ESAs are covert exercitives. ESAs are, inter alia, attempts made by speakers to negotiate, restructure, or reinforce the ways in which they are made intelligible and materially treated by

others. They lay claims to certain positions in social space. So, *refusing* to give uptake to such self-ascriptions amounts to denials of the claims that they make.

Such refusals can be (and they often are) harmful to the utterers of those ESAs.[57] Moreover, the harms perpetrated by such refusals can (and they often do) wrong those ESAs' utterers. People have a prima facie right to self-determination, and this right encapsulates, at least to a large extent, their social positioning. So denying a speaker's ESA entails denying their rightful attempt to self-determine. As Camp and Flores summarize, "[b]eing labelled by others in a way that one rejects and that fails to accurately track the traits that are central to one's self-conception clearly constitutes a violation of autonomy and a moral harm" (Camp & Flores, 2024; Cf. Dembroff & Saint-Croix, 2019: 587; Kukla & Lance, 2023: 1151; Whiteley, 2023).

Versions of this argument have been deployed in different ways by a number of authors, particularly with respect to first-person *gender* ascriptions. For example, Dembroff and Wodak (2018) suggest that gender identity labels "provide others with a guide or a blueprint for interpreting and evaluating one's behavior or speech" (378). Misgendering someone (which, I assume, is a salient form of refusal to give uptake to their first-personal existential gender ascriptions) involves using a social blueprint they did not authorize—and potentially outright rejected—to make sense of them and their actions:

> Consider gender-specific norms like *women shave their legs*. Misgendering McGregor [a senior transgender military officer] would undermine the intelligibility of her actions and choices vis-à-vis this norm regardless of whether McGregor shaves her legs: she can only be recognized to conform to or violate the norm if the norm is recognized as applying to her. So if McGregor is misgendered as a *he*, the intelligibility of her actions and choices is undermined. She is stripped of certain forms of speech and self-expression. (378)

[57] Of course, uptake denials might be harmful *even if* it turns out that ESAs are not exercitives. For instance, refusing to give uptake to an *assertion* can count as a form of testimonial injustice (Fricker, 2007).

Indeed, this could explain why, as Bettcher highlights, trans people are often stereotyped either as being intentionally deceptive or tragically delusional when they self-present in a way that is perceived to be misaligned with their genital status (2007). The blueprints that they themselves authorize for use in interpreting their own actions, attitudes, and identities are *not* utilized by their social environments. Apart from the fact that such refusals to use trans people's elected blueprints can and do cause significant material harms, they also undermine their autonomy with respect to how these individuals are to be made intelligible by others.

As such, certain uptake refusals can also be interpreted as a form of hermeneutical injustice (Fricker, 2007). That is, an epistemic agent's unjust deficit in "communicative intelligibility," which harms agents insofar as it renders "some significant area of one's social experience obscured from collective understanding" (154). As I've explained, refusals to give uptake to trans persons' first-personal gender ascriptions seem precisely to block them from being understood by their communities in the ways that they choose, with devastating results. As Fricker and Jenkins put it:

> [T]rans people may suffer the full range of harms associated with hermeneutical injustice: unjust intelligibility deficit (the intrinsic, primary harm), its negative practical consequences (secondary harms), and moreover those extended and specifically identity-related secondary harms concerning both social perceptions (what one "counts" as) and one's actual self-identity. (Fricker & Jenkins, 2017: 275; cf. Henckel, 2025; Turyn, 2023)

Given everything that has been said so far, it is completely understandable why many theorists reject almost categorically the permissibility of uptake refusals for at least certain kinds of ESAs. Most saliently, perhaps, for gender-related ESAs: "[M]ere denial [of a person's first-personal gender ascription] is a kind of second-personal denunciation of their status that brings a dimension of discursive and social violence that is hard to justify" (Kukla & Lance, 2023: 1152).

I am rehearsing (some of) the potentially harmful consequences that certain ESA uptake refusals may harbor, because it is crucial to acknowledge that the nature of this topic requires

handling it with care. My positive thesis is that *sometimes*, some ESA uptake refusals are permissible, and may even be prudentially beneficial for the ESA's speaker. The paradigm cases to which my thesis applies are quite different from those that feminist philosophers typically concentrate on when they are engaged in ameliorative and emancipatory projects. Still, my view *will* suggest that occasionally (perhaps on rare occasions) even certain refusals to give uptake to sincere self-ascriptions that involve a speaker's, say, gender-identity could be permissible (and maybe even prudentially beneficial for the speaker). But given the severe consequences of many uptake refusals it is of paramount importance to establish what the permissibility conditions for such refusals are. I will say a few things about this question in section 7. For now, I want to make progress towards this paper's positive thesis.

6. Evaluative Underdetermination

In previous chapters, I have argued at length for the idea that our evaluative outlook is not maximally determined. I will not rehearse the arguments here. Instead, I'll restrict myself to a few brief remarks to elucidate what I mean by that. Hopefully, the picture that will emerge from these remarks will sound at least tentatively compelling enough for you to bear with me (and with the other the authors I will mention, who share this picture). The key point is that the characteristic underdetermination of our evaluative perspective, together with the idea that ESAs are covert exercitives, will prove to have important ethical consequences.[58]

First, our avowed values and cares are often *open-ended*. Valuing and caring attitudes are *intentional*: they are directed at or refer to certain objects. I value *you*, or care about *my dog*, or are

[58] In case you do not, ultimately, agree with me, I invite you to read this paper's conclusion as a conditional. *If* our evaluative perspective is undetermined in at least some of the ways I mentioned, *then* sometimes, uptake-refusals are permissible (or even obligatory). If this sounds to you like a reductio of a view you think is wrong, then I've just provided you with another argument against it. For what it's worth (as will become clear in the next section), I think that many of the feminist philosophers I engage with here *should* be open to the ideas I am about to present.

concerned with *my cholesterol levels*. Yet often, such valued or cared-for objects might be highly diffuse and non-specific. We might care about or value abstract things like "health," or "happiness," or "friendship." And honing or spelling out what amounts to these in our own particular, non-fungible case might be a long—and potentially arduous—task (Millgram, 1997). As Nguyen explain, with the example of valuing "a good education":

> "A concept like … 'a good education' admits of different interpretations. Different people may work out their own interpretation of what counts as a good education—and so evaluate their understanding of the term. You want to know more useful things; I want to indulge my sense of curiosity—both are viable understandings of what one might want out of an education." (Nguyen, 2024: 482)

Another salient form of underdetermination in our evaluative-motivational perspectives often characterizes singular experiences of specific emotions. It is widely accepted that emotions are evaluative. In part, this is because they involve appraisals of certain aspects of our situation with respect to their perceived relevance for promoting or hindering our well-being, goals, and motives.[59] But often, the same token emotion can be appropriately interpreted as a response to different practical considerations. As with values and cares, the intentional and formal objects of our specific emotions may be importantly underdetermined. For instance, I may feel intense envy upon learning about your recent prestigious publication. But my bout of envy can be appropriately interpreted as disclosing my care for my own prestige, or my desperate desire to produce high quality work, or (in a more sinister key) my clandestine wish to see you fail.

Furthermore, our evaluative-motivational outlooks are comprised of more than full-blown, referentially unified cares and values. People are disposed to undergo similar patterns of affective

[59] This thought is not restricted to so-called 'appraisal' theories of emotion (e.g., Ellsworth, 1991, 1991; Frijda, 1986; Lazarus, 1991; Moors et al., 2013), but it is a commonplace assumption in most if not all accounts of emotion. To give an incomplete list, Ekman (1992), suggests that emotions involve automatic appraisal processes; Tappolet argues that emotions are perceptions of evaluative properties (2016); Goldie maintains that emotions are "evaluative feelings" (2000); Helm develops a view of emotions as "felt evaluations" (2001); and according to Prinz, emotions are "embodied appraisals" (2003).

and attentional reactions in similar contexts (see chapters 2 and 3 in particular). These patterns can be quite stable across time, but often, they are not directed at the same intentional object at all. For example, whenever I sit at my desk to write, I may be reliably disposed to feel, inter alia, anxiety about the force of my argumentations; restlessness about expected emails; a sense of insufficiency at my academic output; and a burning urge to check if my colleague is in her office. This *referentially disunified* bundle of psychological attitudes may be fittingly interpreted as a manifestation of several different values and cares. It might be construed as an expression of my perfectionism towards my intellectual work; a debilitating concern for others' opinions; my (perhaps over-)valuing of membership in an intellectual community; or my ultimate weariness of academic life. In general, our reliable, habitual patterns of emotion and attention are often fittingly interpretable as pointing to or grounded by different cares or values.

Often, then, our cares, values, and emotions, as well as our barely-noticeable affective and attentional responses can all be legitimately specified or articulated in different ways. That is, they are all *multiply interpretable*. The facts about our affective-motivational makeup often grant us considerable—although, by no means complete—leeway in understanding their significance. In turn, this means that often, acts of interpreting our own values, cares, and emotions are *partially discretionary*. The specific, fine-grained details regarding the contents of our values and cares are underdetermined by our evaluative, motivational, and affective states, processes, and dispositions.

Of course, not everything goes. Our psychological makeup rules out the propriety or fittingness of at least some potential interpretations. There is something weird—if not outright delusional—in saying that my envy towards your accomplishment is an expression of my care for your well-being, or of my uncompromising love of wisdom. We might say that in such cases of indeterminacy in our cares, concerns, and values—calls this phenomenon *evaluative*

underdetermination—the contents of such attitudes are *determinable*, but not radically so. So, when agents set out to articulate, interpret, or spell-out the objects of their values, cares, and concerns, they must ensure that their candidate interpretation is a genuine *determinate* of such determinable values, cares, and concerns.

7. Dialogical Self-Interpretation

Our characteristic evaluative underdetermination means that we often self-ascribe a valuing or caring attitude without having settled the question of what *exactly* it is that we value or care about *prior* to the interpretation. Another way of seeing this is noticing that our valuing and caring attitudes are what Jones calls "trajectory dependent properties" (Jones, 2008). That is, properties that obtain in virtue of something being a trajectory. In turn, a trajectory is an ordered, temporally extended sequence of events. To use one of Jones's examples, "being on a quest" is trajectory dependent. Whether one instantiates it at a particular moment depends on events that happen else*when*. In particular, such property is vulnerable to future contingencies. Whether you are *now*—as you are reading this sentence—on a quest likely depends on what happens afterwards. Moreover, whether you are now on a quest might plausibly depend on the description under which you view your present situation. Hence, "being on a quest" is not only a trajectory-dependent, but an *interpretation-sensitive* property:

> Once you conceptualize your activity as quest, you have available a range of ''quest scripts,'' scripts that identify the need for resoluteness, as well as scaffold your attempts at resolve by telling you how to build it, how, for example, to create your own ''fellowship.'' Now that you know what you are doing, you know how to go about doing it. (Jones, 2008: 273).

Now—and this is precisely Jones's point—we instantiate many other properties that are interpretation-dependent. Jones concentrates specifically on the property "being in love," but I think that her argumentation can be naturally extended to our valuing and caring attitudes in general. As many theorists note, caring about or valuing something is constituted by a temporally

extended, projectible pattern of emotional, conative, and cognitive dispositions (Helm, 2001; Jaworska, 2007b, 2022; Jaworska & Wonderly, 2017; Shoemaker, 2003). It involves, among other things, dispositions to experience positive emotions when that object is flourishing and negative emotions when it is harmed; to entertain desires to promote and protect it; to view it as a source of reasons for one's actions; and to pay particular attention to it and to considerations that are relevant to how well or poorly it fares. In other words, caring or valuing is (at least) a temporally extended pattern of psychological vulnerability, or attunement towards the cared-for or valued object.

This means that caring or valuing are trajectory dependent. Moreover, I think that these attitudes (or rather, attitude complexes) are also, to an extent, *interpretation sensitive*. As I've said, an agent's evaluative perspective is often underdetermined. Our affective-motivational makeup at-a-time can be interpreted in multiple ways. But once our self-interpretation has been issued (in the form of an ESA) it can trigger a cascade of interpersonal responses that can lead to changes in our standing evaluative perspectives.

Now, as I noted, ESAs are covert exercitives. As such, they express something like a request, demand, or invitation to treat us in ways that accord with their content. They strive to change the normative statuses in our social milieu—the sets of commitments and entitlements that characterize our discursive community; to secure, or negotiate, or shift our position (and the positions of others) in social space. In turn, such normative status changes inevitably lead to alterations in what we notice, desire, feel, or pay attention to. In short, they alter our evaluative-motivational perspectives (cf. Camp & Flores, 2024).

Here is another way of spelling this out. I mentioned Dembroff and Saint-Croix's notion of agential identity—the aspects of ourselves that, wittingly or otherwise, we express or make available for public scrutiny. I've also said that Dembroff and Saint-Croix conceptualize agential

identity as importantly different from what they call our *self-identity*, which is comprised of something like our felt sense of who we are. One tempting thought is that the causal relationship between our self-identities and our agential identities is unidirectional. That we need to *first* settle on a relatively highly-elaborated self-identity before we can establish an agential identity via overt (linguistic or non-linguistic) performances.[60] I think—and I suspect Dembroff and Saint-Croix would agree—that this is not the case. That acts of expressing our agential identities can routinely help us fix or determine the contents of our self-identities.

According to Dembroff and Saint-Croix, part of what comprises an agent's self-identity is a *felt relationship to certain norms*, which Jenkins (from whom this idea is borrowed, 2018, 2023) calls "identity as norm-relevancy." Roughly, identity in this sense is constituted by a person's experience of which norms and expectations apply to her. This does not mean that the person necessarily *abides* by the norms, but rather that "the norms function as a reference point for one's own experience of one's behaviour … Norms can play this role for one even if one adopts a critical stance towards them" (Jenkins, 2023: 159). In Jenkins's example,

> Imagine a woman who does not remove her underarm hair, who wears clothing that reveals that hair, and who feels subtly as if she is doing something transgressive when she does so. This woman is experiencing a feminine norm ("women should not have underarm hair") as relevant to her, even though she does not comply with it. (160)

While both Jenkins and Dembroff and Saint-Croix are particularly interested in *gender* norms, clearly our self-identities involve our felt-relationships with many norms that are not primarily gender-related. Moreover, our characteristic evaluative underdetermination means that, unlike norms such as "women should not have underarm hair," the contents of some of the norms

[60] As part of their *definition* for agential identity, Dembroff and Saint Croix assert that "to have an agential identity, one must self-identify as a member of the group in question" (576). I do not take myself to be denying this (after all, I claim to be using *their* concept). What I do deny is that our self-identity must be maximally—or even highly—determined before we have a corresponding agential identity. Indeed, that our *agential* identity can be more determinate, nuanced, or specific than our self-identity.

that feature as reference points for our self-identities are highly diffuse, inchoate, or underspecified. We can feel the normative force and relevance to our condition of imperatives like "you should improve your health," or "you need to fix this!" or, in the words of Rilke, "you must change you life" (1982). And we can externalize our felt relationships with such imperatives despite the fact that we haven't settled the question of what *exactly* they mean in our case.

One—especially prominent—way of externalizing our relationship with the norms that we take as relevant to us is by issuing, in dialogue, a self-interpretation in the form of an ESA. But, as I've said, once an ESA has been uttered, its uptake (whether this uptake was intended or not) consists in normative status updates. These updates, in turn, lead to changes in our own evaluative-motivational makeup. In particular, they can and do make our values and cares more determinate. Issuing a self-interpretation can be a way of jamming ourselves into a specific coordinate in social space, and in doing so, to further specify our self-identities.

Consider a clichéd example: in their early forties, Alex is starting to feel unease with the shape their day-to-day life has assumed over the years. Alex has a loving partner and a steady job, a nice house and a small-yet-close circle of friends. But a growing air of dissatisfaction is starting to subtly but persistently permeate Alex's being. This feeling is experienced by Alex as normatively significant. Alex, that is, is starting to feel certain normative claims as deeply relevant to their condition. However, these claims may be highly inchoate or diffuse. They might be expressed in no more accurate terms than "do something about this!" Alex's self-identity, their evaluative-motivational makeup, is tinged with a hue of frustration, but what is Alex's malaise specifically *about*? What I've tried to suggest is that such stereotypical uncertainties are not merely (or always) epistemic. The phenomenon of evaluative underdetermination means that some aspects of our self-identities just *are* non-specific prior to their articulation.

At this point, I should stress that, as Alex's case demonstrates, the indeterminacy of (some of) our evaluative attitudes does not necessarily mean that such attitudes are *motivationally weak*. Indeed, Alex's unease might intensify to a point of being close to unbearable. Alex could feel that they must take some drastic actions to abide by (or, for that matter, resist) the norms that they experience as applying to them. Indeed, part of what makes Alex's predicament so frustrating is just such misalignment between the indeterminacy of their evaluative makeup and its ever-increasing motivational force.

One way in which Alex can alleviate their distress is by making the content of their evaluative perspective *more* specific. And this can be—and often is—done by issuing a certain ESA in dialogue: "I think I want kids," "I should quit my job," "I need to go on an adventure." The way that the utterance would be taken up by Alex's discursive community would, per the framework I am adapting from Kukla and Lance, consist in certain updates of normative statuses. Depending on the content Alex's ESAs, Alex (and members of their social environments) could be granted new entitlements and assigned new commitments; and these in turn would further *determine* Alex's evaluative-motivational outlook.

8. Refusals Done Right

By now, the hard part of my argument is over. To summarize, uttering an ESA can be a way of *fixing* or *determining* our evaluative-motivational outlook. In cases of evaluative underdetermination, ESAs can function as attempts to induce normative landscape changes—changes to the set of commitments and entitlements possessed by members of the speaker's discursive community—which in turn will lead to further changes to the speaker's values, motivations, and cares. This means that often, utterances of ESAs are important intersubjective

choice points. They are forks in the speaker's road to further individuation or self-(trans)formation, as well as opportunities to negotiate the social positions of both speaker and hearers.

But just as with other identity-establishing choice points, speakers can make *bad* decisions with respect to their own self-interpretations. A person's specific self-interpretation might be *hermeneutically appropriate*, that is, it might *fit* her underdetermined evaluative perspective in the sense that it is a proper *determinate* of that perspective. But it may be inappropriate, or simply *not right* in different respects. The uptake that some utterances require—the social landscape changes that they are striving to institute—may be harmful to the speaker or to members of their community. In such cases, hearers might be permitted, or even obligated, to refuse to grant full uptake to the speaker's ESA.

Consider Alex's case. In a moment of vulnerable openness, Alex might self-interpret their midlife malaise as indicating that they want to leave their partner, or to quit their job, or to embrace a nomadic lifestyle. Each of these interpretations might be hermeneutically appropriate, but accepting each of them carries different, and potentially drastic implications for Alex *and* for others. Indeed, granting uptake to ESAs that express such interpretations might be straightforwardly harmful to Alex and to their intimates. Insofar as we think that sometimes, stopping someone from harming themselves or others is permissible, we should, I maintain, think that sometimes, *not* granting full uptake to a speaker's ESA is permissible. This is because, as I said, granting uptake to a speaker's ESA might jam them (and us) into social positions it'd be best for them (and for us) not to occupy.

Does this mean that such benevolent uptake refusals constitute a form of paternalism? While I cannot grant this question a comprehensive treatment, I want to mention a few

considerations in favor of a (qualified) negative response. This discussion will also enable us to gauge how hearers may, so to speak, refuse uptake *the right way*.

For the sake of the discussion, let's adopt a version of standard accounts of paternalism, according to which an agent P's action towards a patient Q is paternalistic if, roughly, "by doing (or omitting to do) R, (1) P limits Q's liberty or interferes with Q's decision-making; (2) against Q's will, without his consent, or contrary (or with indifference) to his values, preferences, or desires; and (3) for Q's own good" (Ben-Moshe, 2022; Dworkin, 2020; Scoccia, 2018).[61] Of course, the potentially morally problematic features of this account are conditions (1) and (2). As I will suggest, we have good reasons to believe that in many circumstances, benevolently refusing to grant uptake to certain ESAs does not satisfy either of these conditions (and many need not satisfy condition (3)). Specifically, such circumstances involve cases wherein a certain ESA is an attempt to determine a-yet-to-be-determined evaluative outlook.

Regarding condition (1), at least insofar as Alex's evaluative outlook is still underdetermined, their ESAs regarding it may not have the illocutionary force of a full-blown order, demand, or plea. Rather, the ESA might resemble a primarily self-regarding *motion* or *proposition*. Alex's utterance of their self-interpretation might be likened to suggesting that the normative landscape change in certain ways. Such suggested changes could be (clumsily) explicitly articulated as "I propose to hereby care about expanding my family," or "prioritize myself over my present or future partners," or "value existence on the open road and under the great blue sky." This means that many ESAs are relatively *weak* exercitives in a specific way, which is worth spelling out.

[61] But see, for example, Groll (2012) and Shiffrin, (2000) for other accounts of paternalism.

Notice that different exercitives present different reason(s) for their hearers to follow the speaker's instructions regarding how to shift the discursive community's normative landscape (My discussion here is indebted to Riggle (2022), See also Harris, (2014); Schiffer, (1972)). In the case of formal orders, for instance, the *presented reason* might be the speaker's institutionally recognized status. In other cases, like demands, that reason might be the authoritative force of the speaker's desire; and in invitations, the reason might be that the speaker (or the hearer) will value the hearer's following the instructions. Now, if my analysis of the illocutionary force of many ESAs is correct, that is, if ESAs often function as self-regarding motions, or propositions, then it seems that the reasons they present for their own satisfaction are relatively weak. When Alex suggests that they hereby care about, say, X, it seems that the reason their ESA presents is something like Alex's *sense*, or *suspicion* that they would appreciate the changes that would ensue from Alex's own caring about X.

But if the illocutionary force of Alex's ESA is more akin to that of an invitation or a proposition than a demand, then there is nothing particularly paternalistic about refusing to accept it. There *is*, to be sure, a sense in which it "interferes" with Alex's decision-making, but this sense is benign: when you refuse my invitation or proposition to do something together (in the cases under consideration, to rearrange our shared social space), you send me back to the drawing board. Your refusal dictates that I come up with another plan of action, or a different proposal, if I wish to secure your cooperation. An account of paternalism that entails that such refusals are paternalistic seems to be flirting with its own reductio. So at least in cases where ESAs function as weak exercitives, the first condition for paternalism does not seem to be satisfied.

Furthermore, in cases of evaluative underdetermination, there are many ways of refusing uptake in a manner that avoids the satisfaction of condition (2) as well. If a speaker's ESA

functions as a weak exercitive in the sense I just spelled out, then refusing to grant it uptake does not have to count as going against the speaker's will, values, preferences, or desires. This is because the speaker's will (or values, or preferences, or desires) is precisely *not fully determined* yet. As I said, many ESAs function as *propositions* made by the speaker to care about something in particular, rather than *demands* to respect a pre-formed, specific care or concern. If that is the case, it is possible to respect a person's evaluative outlook *while* denying uptake to their ESA. I wish to dedicate the remainder of this section to spell out precisely what I think (at least some of) these conditions are. It is important, however, to keep in mind that while I'm relatively confident that such conditions are *necessary* for the pro-tanto permissibility of a certain uptake refusal, I am less certain with respect to whether they are also *sufficient*.

The first condition, as I said, is that the speaker's outlook must exhibit genuine *evaluative underdetermination*. Their values, cares, and concerns must be under-specified in the sense that they admit to multiple fitting potential interpretations. Secondly (again, as already suggested), the ESA should be relatively *weak* in that its presented reasons are the speaker's defeasible belief, or sense, or suspicion, that they (or their hearers) will value the changes that would result from their caring about the object(s) that feature in the ESA's content. Thirdly, another condition that I mentioned is that the speaker's proposed changes to the normative landscape of their discursive community will plausibly harm either the speaker or members of their community.

Fourth, it seems to me that the hearer's refusal to grant uptake to a given ESA must be a case of illocutionary *distortion* rather than *silencing*. Recall that in illocutionary distortion, a given speech act *does* manage to change the normative landscape of a situation in certain ways, which, however, do not conform with the speech act's intended function. In distortion—but not in

silencing—the speaker manages to secure uptake in the form of certain *acts of recognition* by their hearers with respect to the speech act.

I wish to remain relatively permissive regarding which acts of recognition render an ESA's uptake refusal permissible. However, at the very least, it seems to me that the hearer must enact their acknowledgement that the speaker is currently attempting to determine—to render more specific—their evaluative outlook. That is, that the speaker is attempting to figure out what moves and stirs them—what, we may say, they are about. This might entail, minimally, that the hearer does not deny that the speaker is undertaking this particular quest, and perhaps less minimally, that they offer help, perhaps in the form of an alternative to the speaker's proposed self-interpretation.

How should such help be offered? What are the constraints on the hearer's interpretation of a speaker's values or concerns? Here too I am inclined towards permissiveness, but I wish to mention what I take to be one rough-yet-crucial guideline: at the very least, if I deny uptake to your ESA by putting forward my own alternative interpretation of your evaluative perspective, my interpretation must be a genuine *determinate* of that *determinable* perspective. My interpretation of *you* must abide by the same standards of hermeneutic propriety as your original self-interpretation. The interpretation must fit the facts regarding your evaluative-motivational makeup. Proposing—in the sense of putting forward a motion—that you care about something that you cannot care about in the right way is in the best circumstances a way of disrespecting your deep values and concerns, and in less ideal circumstances an act of violence against what may lay lie at the core of who you are—inchoate or vague as it may be.

At this stage, I wish to stress one crucial point: clearly, there may be some cases where it is *pro-tanto impermissible* to deny uptake to a speaker's ESA, but still *all-things-considered permissible*. To adapt Dembroff and Saint-Croix's example (2019: 591-2), denying uptake to a

person's self-ascribed white nationalist identity is clearly permissible (if not outright obligatory), regardless of whether harms the individual whose uptake is being denied. As they assert, "any harm done to [the person] is vastly outweighed by the good of preventing uptake of [his] agential identity as a *noble white nationalist*" (592). But still, Dembroff and Saint Croix maintain (in an unqualified way) that "to deny uptake is to cause harm to the individual denied" (591). *Pace* Dembroff and Saint Croix, I have argued that, in many cases, it is neither harmful nor even *pro tanto* impermissible to deny uptake to the individual denied, and I have tried to enumerate (some of) the conditions that secure this permissibility.

The distinction between determinable and determinate evaluative perspectives—and the conditions for hermeneutic propriety that it entails—can help us analyze further what goes awry with uptake refusals of *gender*-related self-ascriptions. Usually, when a speaker has finally decided to give voice to their gender-nonconforming identity—to include it in their *agential identity*—their gender-related evaluative perspective has already been determined to a significant degree. As Henckel writes, while not indefeasible, trans individuals' (to take one example) gender-related ESAs typically result from "substantial cognitive and emotional labor, time spent reflecting, and mindfulness of one's own gendered experiences, as well as special knowledge of resistant epistemic resources and experience in resistant communities" (2025: 9). In the vast majority of cases, telling someone that they are *not* trans is (among other harmful things), a way of flaunting their settled, already-preestablished gender identity. It is trying to shift the speaker's normative landscape in a way that simply *rejects* their evaluative, motivational, and affective makeup. In other words, it is hermeneutically unjust (in part) because hermeneutically *inappropriate*. And this impropriety can be captured by saying that it is an attempt to assign the speaker values, cares,

concerns, and motivations that are *not* genuine determinates of their existing (if, perhaps, determinable) evaluative-motivational outlook.

Contrast such malign illocutionary distortions with the following exchange:

> A: I love you!
> B [bitterly shaking their head, tears welling up in their eyes]: You don't love me, you just love the idea of me.

I suggest that such cases are quite often (in all likelihood *typically*) permissible reappraisals of the speaker's evaluative perspective in part *because* this perspective is still unsettled. In such cases, as I argued, A's exclamation functions as a weak covert exercitive. It is akin to a *motion* that A puts forward to hereby love B. In such circumstances it is not even *pro-tanto* impermissible for B to refuse uptake to A's ESA; to *distort* A's proposition for normative landscape change by putting forward their own alternative interpretation of A's outlook. What renders B's alternative interpretation apt is that it *is* a legitimate, genuine determinate of A's determinable perspective regarding B. Assuming that A is sincere, we might plausibly suppose that they are experiencing an array of intense, B-related feelings, but their object—whether it's B themselves, or the 'idea of B,' or something else—is not yet determined.

Or, conversely, consider the following, short-yet-vulnerable dialogue between, we may suppose, two close friends:

> A: I think I should get a divorce.
> B: Totally. Might as well just get it over with.[62]

Especially if A is expressing this thought to B for the first time, B's unquestioning uptake-granting to A's speech act might (I think *should*) strike us as inappropriate, or off kilter. It may seem to us that B is *not* being a good friend to A by sheepishly affirming A's utterance. Again, I think that the view I have been developing can help us see why. In such cases, the person uttering

[62] Thanks to Nick Riggle for the example.

the ESA did not yet settle the question of whether they want to get a divorce. Again, their ESA functions as a weak covert exercitive. It is striving to determine a still *indeterminate* agential identity in a certain way, and this way might be deeply harmful to A and others. In such cases, it is natural to think that B has strong reasons to deny uptake to A's exercitive, to refuse to accept A's motion for this specific rearrangement of A's evaluative outlook and social space without at least some kind of thorough questioning. In fact, we might think that this is, in part, what friends are for.

9. Engagement and Resonance

Before closing, I want to make a final—and, admittedly, speculative—proposal: the sheer *possibility* that our ESAs will be denied uptake makes their pronouncement in dialogue a particularly valuable way of engaging in self-formation and self-maintenance. As I'll argue, dialogical utterances of ESAs can allow their speakers to realize two kinds of broadly *aesthetic* value, which I have called in the previous chapter *engagement* and *resonance* value. And at least *resonance* value can only be realized on the condition that *who we will become* following the ESA's utterance is not entirely up to us. The possibility that an ESA will be denied uptake is what guarantees this option, and hence what allows for the realization of resonance value in dialogue.

9.1. Engagement

Let's start with engagement value. As I said, our evaluative perspectives are characteristically underdetermined, but not radically so. This means that the activity of honest self-interpretation—of, for example, providing reasons for our actions or feelings—poses a unique agential challenge. First, when we self-interpret, we must be faithful to the facts of our actual, concrete evaluative-motivational outlook. We must take into consideration our histories, psychological makeups, and environmental conditions in the stories that we weave about ourselves (we must respect what Schechtman, 2014 call "the reality constraint"). Thus, as a particularly

elaborate form of engagement in the 'game of giving and asking for reasons,' self-interpretation must abide by the rules of discursive rationality. On the other hand, our common (and perhaps constant) condition of evaluative underdetermination ensures that there is seldom one uniquely correct interpretation for the reasons, values, and cares that our actions and feelings express. As I have said, we often have significant hermeneutic leeway—or authorial freedom, if you will—with respect to our self-interpretations.

And this means that both *correct* and *incorrect* self-interpretations can exhibit a host of aesthetic qualities. They can be elegant or messy, simple or intricate, polished or coarse, masterful or clumsy. You can take on the agential challenge of *determining who you are as an agent* in aesthetically valuable or disvaluable ways. The activity of self-interpretation provides an opportunity to adhere to the facts of our condition and to the canons of rationality, but to do so in more or less creative, thoughtful, or pleasing ways.

Given these observations, I think that occasions for self-interpretation are occasions for engagement with the distinctive value of what Nguyen call *striving play* (2019, 2020). In striving play, agents adopt the official, stated goal of the activity they are pursuing—paradigmatically, certain *games*—but this is not the *point* of the play. Instead, adopting the goal of the activity allows the striving player to access the value inherent in the struggle to overcome the challenges that the game, or game-like activity furnishes. It allows them, that is, to modulate their agency in response to the demands established by the game's architecture. Striving players must try to *win*—i.e., to achieve the goal of the activity—but winning isn't what the playing is about. Rather, the point is to realize the value of *engagement* in the activity. Thus, striving play involves a kind of *motivational inversion*: "[i]n normal practical life, one takes the means for the sake of the ends. In striving play, one selects the ends for the sake of the means it puts one through" (2020: 1143).

Importantly, *bona fide* games don't seem to be the only activities that allow for the realization of this kind of *engagement value.* As both Nguyen (2020, 2023) and Riggle (Riggle, 2022, 2024b), for instance, argue, practices of aesthetic appreciation also seem to exhibit this kind of motivational structure. As Riggle puts it, "[t]he primary value of aesthetic valuing consists in the activities of *generating* good aesthetic beliefs and not in *possessing* good aesthetic beliefs" (2024b: 394, my emphasis). Or, as Nguyen asserts, "[i]n aesthetic appreciation, we aim at correctness, but correctness is not the purpose. It is only the right goal to adopt in order to become engaged in a desirable form of activity" (Nguyen, 2020: 1143).

Riggle offers helpful remarks about what makes aesthetic engagement and the practices that promote it so valuable. In short, such practices cultivate our capacities for what he calls *discretionary valuing* and *volitional openness*. First, while some aspects of our values, cares, and concerns are compulsory or heavily conditioned by social factors, others are not. All of us are subject to normative pressures to, say, value our family, or to disvalue torture. However, our evaluative outlooks are not completely foisted on us. We often can *choose* what and how to value. Discretionary valuing is the set of processes of selecting, honing, determining, and acting on our values. As Riggle writes,

> [w]hen we engage in discretionary valuing, we choose to devote our time, attention, care, and concern to the thing of our choice. And we take a certain pride in, even identify with, the results. We become food people, travelers, music lovers, fashionistas, rappers, guitar players, break dancers, car lovers, wine aficionados—in other words, we become *individuals*. When we exercise our discretionary valuing capacities, we establish patterns of discretionary valuing and cultivate our individuality. (2024a: 128)

Secondly, the capacity for *volitional openness* is the capacity to "seek out, be immediately responsive to, and engage with the value in one's immediate environment" (2024b: 404). It is the capacity to "open ourselves up" to different sources of values, and, as Gingerich puts it, to "experience ourselves as sources of freshness and novelty" (2023: 57). Volitional openness thus

allows us to "act from a generous, curious, exploratory, or spontaneous place" (Riggle, 2024a: 129).

While I agree with Riggle and Nguyen that practices of artistic or aesthetic valuing are *especially* effective in facilitating engagement (in part thanks to the unique ways in which they cultivate and activate our capacities for discretionary valuing and volitional openness), I think that dialogical self-interpretation is another, common-yet-overlooked context in which engagement value can shine.

Self-interpretation constitutively involves a struggle to make sense of our inner turmoil and our external circumstances. But as I stressed, we usually enjoy some measure of hermeneutic leeway in this endeavor thanks to the indeterminacy of our evaluative-motivational perspectives. We can, to changing degrees, *choose* which *determinate* interpretations we offer for our *determinable* perspectives. This means that, in the usual case, self-interpretation is a way of exercising our capacity for discretionary valuing. For similar reasons, self-interpretation can *also* provide us with opportunities to exercise volitional openness with respect to sources of evaluative novelty. When we are encouraged to give an account of our values or cares—especially in dialogical contexts—we are usually under some pressure to say something *substantive*. Self attributing cares, values, or reasons that are overly unspecific or noncommittal is usually not acceptable as a serious response to questions like "why did you do *that*?" But the fact that our evaluative perspectives are often underdetermined means that, to come up with something substantive to say, we need to be attuned to potential sources of value. To be on the lookout for some determinate value or care that it could make sense to attribute to ourselves in our concrete situation. This is exactly what being volitionally open consists in. In short, the capacities whose exercise allows us to issue satisfactory, hermeneutically-apt ESAs are the same capacities whose

operation is highly visible in practices of aesthetic valuing. Interpreting our *selves* is like interpreting works of art.

9.2. Resonance

So, dialogical self-interpretation allows agents to realize the kind of engagement value that is typical of striving play and of aesthetic valuing practices. But finally, I want to suggest that it also enables us to realize another kind of value. The realization of this value, however, is predicated on the possibility of being denied uptake.

The kind of value I have in mind is what I have called, in chapter 4, *resonance* value. It is the value that people seek when they go on roller coaster rides, or melt into a lover's embrace, or receive a deep tissue massage, or are led on a guided meditation. In such subject-involving happenings, people allow themselves to abdicate certain features of their agency in order to be moved, or shaped, or handled by their environments. If engagement value is marked by an active form of struggle to meet the demands of a certain task, the signature of *resonance* value is a form of willful passivity. It is achieved, that is, when we allow ourselves to be *subjected* to, or immersed in the pressures that the world exerts on us.

Now, one part of what makes resonance value *valuable* is the fact that it consists, in part, in the cultivation and exercise of our capacities for volitional openness. But more than that, I think that resonance value is also—perhaps primarily—valuable because it is partly grounded in our capacities to stand in what we may call *ontological intimacy* with others. When we are ontologically intimate with something or someone, we render our identity vulnerable to it. We enable ourselves to be impacted by, or beholden to, or dependent upon its effects, or actions, or fate in pervasive ways. In other words, intimacy with another person, object, or place enables us to *belong* to it. And belonging, I think, is something that we desire and look for in our friendships

(Cocking & Kennett, 1998; Dover, 2022, 2023), our romantic relationships (Yao, 2023), our attachments (Wonderly, 2016), and even our professional affiliations.

I am not sure that I have the time—or, if I'm being honest, the talent—to make this highly impressionistic claim more precise, but perhaps Yao's remarks might help. As she says, in some identity-shaping happenings,

> [o]ne may simply want to lose control, to be moved about in the world by forces that seem outside of oneself, to come to understand oneself in relation to another particular person (or other object of erotic or passionate attachment), to be unsure of where one ends and begins, to be unable to find rational justification for why one wants to do what one wants to do (2023: 15)

If this description strikes you as portraying some things that you too might find valuable and important (at least occasionally), then I take it that you are sufficiently open to the thought that resonance value has a role to play in a well-rounded, flourishing life. Furthermore, I take it that you are sufficiently open to the idea that resonance value is partly grounded in our capacity for standing in ontological intimacy. And this capacity, in turn, is predicated on our deep—one might say *existential*—vulnerability to others.

If this makes sense to you, I think we are in a position to appreciate why the utterances of ESAs in vulnerable dialogue are opportunities to realize resonance value; but *also* why such opportunities depend on the possibility of uptake refusals. For ontological intimacy to be possible, we must be able to abdicate certain features of our agency. Specifically, we must be able to willfully relinquish our control over what will happen to us. Trivially, this means that what will happen to us ought not be entirely in our hands. Now, as I described at length, in the usual case, when we issue an ESA, we are putting forward a motion regarding who we might become. But this motion *could* be denied by our interlocutors. Such denials will impact the commitments and entitlements that shall characterize us; and hence they will impact our evaluative perspectives. They will alter or, more precisely, *determine* what we care about in ways that do not lie fully at

our discretion. Put differently, that our ESAs *could* be refused uptake means that vulnerable dialogues are sites for our potential self-(trans)formation *by* others. This, in turn, means that such dialogues are particularly conducive to resonance value. In vulnerable dialogue, once we have issued an ESA, we do not merely (partially) relinquish control over what will *happen* to us. we relinquish control over who we might become.

Self-interpretation in dialogue, then, has an interesting, dynamic pragmatic profile. Typically, when you self-interpret, you must engage your rational, creative, and hermeneutic capacities to come up with a satisfactory ESA to make intelligible to yourself and others your motives, reasons, and cares. The capacities that are exercised in producing this ESA are just those capacities that are crucial for the realization of *engagement* value. But once the ESA has been uttered, you are, so to speak, at the mercy of your hearers. Your ESA delineates a potential version of your future self, but whether and to what extent this self will bloom into actuality depends to a significant degree on the uptake you manage or fail to secure. And this, at least if your hearers are not your lackeys, is not up to you. The other side of such willful relinquishing of control, of such rendering oneself vulnerable to others, is precisely what allows for certain happenings to be conducive to resonance value—and the openness and intimacy that it consists in. At the risk of sounding pretentious, uttering our self-interpretations is akin to taking a big plunge into a deep, dark lake, with all the terrifying beauty that such a plunge may engender.

10. Conclusion

When we utter an ESA in the context of sincere, intimate dialogue, we render ourselves (potentially) radically vulnerable to other people's influence on who we might become. This is because ESAs are often covert exercitives that aim to fix, or make more *determinate*, an underdetermined evaluative-motivational perspective. This feature of our dialogical self-

interpretations renders refusals to grant uptake to such interpretations, when these refusals are done *wrong*, so massively pernicious. On the other hand, sometimes, uptake refusals can save us from hammering ourselves into an evaluative perspective that will harm us and others.

But moreover, the fact that issuing ESAs in dialogue requires *both* the exercise of our capacities for rational, creative, autonomous self-interpretation; *and* the willful abdication of our exclusive authority over ourselves makes dialogical self-interpretation a particularly valuable form of self-making. One that enables the realization of both engagement and resonance value. When we self-interpret, we must use all the hermeneutic resources at our disposal to create a draft for who we might become. But then we must send this draft out into our social milieu, to be reviewed, revised, and, if we are lucky, resubmitted.

EPILOGUE

I have tried to provide a detailed, empirically-informed, and convincing account of how self-interpretation may be a particularly valuable way in which humans can form and transform who we are. According to the framework I have developed in this dissertation, humans' evaluative perspectives are indeterminate in non-trivial ways—ways which render these perspectives multiply interpretable. Moreover, I have argued that interpreting such perspectives is, in the usual case, a way of making provisional commitments to care about some object or other in particular. In turn, these commitments can be accepted or denied by members of our discursive communities in manners that bear significance for the values and cares we end up adopting. Therefore, ascribing reasons, values, and cares to ourselves means opening ourselves up to external influence in profound ways. And this, lastly, means that engaging in existential self-ascriptions in sincere, intimate dialogue, is a particularly valuable way of changing and holding on to who we are. It is fitting, I think, to end a dissertation on the power of articulation by pointing to several things that have been left unsaid.

First, although it seems plausible to me that dialogical contexts can enable interlocuters to realize both engagement and resonance value in quite *unique* ways, I did not explicitly argue for this claim here. However, I sincerely hope to explore this thought in future work. Indeed, I have a sense that the potential aesthetic value of dialogue means that certain intimate conversations are paradigms of *beauty*.[63]

Second, while I think speech can provide us with distinctively valuable avenues for self-formation and transformation (I am, after all, a committed Buberian), I believe that the theoretical

[63] I am indebted to Nick Riggle for encouraging me to think about this possibility.

apparatus developed in this dissertation can help us explain non-discursive modes of self-making as well. To mention just a few examples, going on adventures, listening to music, engaging in mental imagery, and touching each other can all, I think, fix our inchoate senses of import in pervasive, important ways. I intend to explore such possibilities in the future.

Finally, as I have noted at the outset, this dissertation is an attempt to clarify, for myself at least, some comments made by Martin Buber throughout his oeuvre. But it is patently *not* an exercise in Buber exegesis. Still, Buber's work is an inexhaustible font of inspiration and wisdom, and I sense that the time will soon come when I feel called to engage more thoroughly with it on its own terms. Among other things, then, this dissertation is a promissory note toward one specific dialogue that I hope to initiate soon.

REFERENCES

Akhlaghi, F. (2023). Transformative experience and the right to revelatory autonomy. *Analysis*, *83*(1), 3–12. https://doi.org/10.1093/analys/anac084

Alexander, L. (1996). The Moral Magic of Consent (II). *Legal Theory*, *2*(3), 165–174. https://doi.org/10.1017/S1352325200000471

Ameral, V., & Palm Reed, K. M. (2021). Envisioning a future: Values clarification in early recovery from opioid use disorder. *Journal of Substance Abuse Treatment*, *121*, 108207. https://doi.org/10.1016/j.jsat.2020.108207

Arpaly, N. (2003). *Unprincipled virtue: An inquiry into moral agency*. Oxford University Press.

Ásta. (2018). *Categories We Live By: The Construction of Sex, Gender, Race, and Other Social Categories*. Oxford University Press.

Bagley, B. (2015). Loving Someone in Particular. *Ethics*, *125*(2), 477–507. https://doi.org/10.1086/678481

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion*, *15*(6), 713–724. https://doi.org/10.1080/02699930143000239

Beardsley, M. C. (1982). *The Aesthetic Point of View*. Cornell University Press.

Beaudoin, M.-N., Moersch, M., & Evare, B. S. (2016). The Effectiveness of Narrative Therapy With Children's Social and Emotional Skill Development: An Empirical Study of 813 Problem-Solving Stories. *Journal of Systemic Therapies*, *35*(3), 42–59. https://doi.org/10.1521/jsyt.2016.35.3.42

Bélanger, J. J., Lafrenière, M.-A. K., Vallerand, R. J., & Kruglanski, A. W. (2013). When passion makes the heart grow colder: The role of passion in alternative goal suppression. *Journal of Personality and Social Psychology*, *104*(1), 126–147. https://doi.org/10.1037/a0029679

Ben-Moshe, N. (2022). Internal and External Paternalism. *Canadian Journal of Philosophy*, *52*(6), 673–687. https://doi.org/10.1017/can.2023.5

Bettcher, T. M. (2007). Evil Deceivers and Make-Believers: On Transphobic Violence and the Politics of Illusion. *Hypatia*, *22*(3), 43–65. https://doi.org/10.1111/j.1527-2001.2007.tb01090.x

Bianchi, C. (2021). Discursive Injustice: The Role of Uptake. *Topoi*, *40*(1), 181–190. https://doi.org/10.1007/s11245-020-09699-x

Brackett, M. A., Rivers, S. E., Reyes, M. R., & Salovey, P. (2012). Enhancing academic performance and social and emotional competence with the RULER feeling words curriculum. *Learning and Individual Differences, Noncognitive Skills in Education: Emerging Research and Applications in a Variety of International Contexts*, *22*(2), 218–224. https://doi.org/10.1016/j.lindif.2010.10.002

Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard University Press.

Branstetter-Rost, A., Cushing, C., & Douleh, T. (2009). Personal Values and Pain Tolerance: Does a Values Intervention Add to Acceptance? *The Journal of Pain*, *10*(8), 887–892. https://doi.org/10.1016/j.jpain.2009.01.001

Bratman, M. (1987). *Intention, plans, and practical reason*. CSLI Publications.

Bratman, M. (2007). *Structures of agency*. Oxford University Press.

Brownstein, M. (2014). Rationalizing flow: Agency in skilled unreflective action. *Philosophical Studies*, *168*(2), 545–568. https://doi.org/10.1007/s11098-013-0143-5

Buber, M. (2023). *I and Thou* (W. Kaufmann, Tran.; 100th anniversary reissue). Free Press.

Burklund, L. J., Creswell, J. D., Irwin, M., & Lieberman, M. (2014). The common and distinct neural bases of affect labeling and reappraisal in healthy adults. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00221

Burnston, D. C. (2017). Interface problems in the explanation of action. *Philosophical Explorations*, *20*(2), 242–258. https://doi.org/10.1080/13869795.2017.1312504

Calhoun, C. (2009). What Good Is Commitment? *Ethics*, *119*(4), 613–641. https://doi.org/10.1086/605564

Callard, A. (2018). *Aspiration: The agency of becoming*. Oxford University Press.

Callard, A. (2020). Transformative activities. In A. Callard, *Becoming Someone New* (pp. 147–161). Oxford University Press. https://doi.org/10.1093/oso/9780198823735.003.0009

Callard, A. (2022). The eros monstoer. *Harper's Magazine*.

Camp, E. (2013). Slurring Perspectives. *Analytic Philosophy*, *54*(3), 330–349. https://doi.org/10.1111/phib.12022

Camp, E. (2015). Logical concepts and associative characterizations. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New direction in the study of concepts* (pp. 591–621). MIT Press.

Camp, E. (2019). Perspectives and frames in pursuit of ultimate understanding. In S. R. Grimm (Ed.), *Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology* (pp. 17–45). Oxford University Press.

Camp, E. (2025). Perspectival complacency, perversion and amelioration. In W. D. Riggs (Ed.), *The Virtue of Open-Mindedness and Perspective* (pp. 19–57). Oxford University Press.

Camp, E., & Flores, C. (2024). Playing with labels: Identity terms as tools for building agency. *The Philosophical Quarterly*, *74*(4), 1103–1136. https://doi.org/10.1093/pq/pqae092

Carman, M. (2018). Emotionally guiding our actions. *Canadian Journal of Philosophy*, *48*(1), 43–64. https://doi.org/10.1080/00455091.2017.1334440

Carruthers, P. (2018). Valence and Value. *Philosophy and Phenomenological Research*, *97*(3), 658–680. https://doi.org/10.1111/phpr.12395

Cassam, Q. (2018). The Epistemology of Terrorism and Radicalisation. *Royal Institute of Philosophy Supplements*, *84*, 187–209. https://doi.org/10.1017/S1358246118000607

Chao, L. L., & Martin, A. (2000). Representation of Manipulable Man-Made Objects in the Dorsal Stream. *NeuroImage*, *12*(4), 478–484. https://doi.org/10.1006/nimg.2000.0635

Christensen, W., Sutton, J., & Doris, M. (2016). Cognition in skilled action: Meshed control and the varieties of skill experience. *Mind & Language*, *31*(1), 37–66.

Cocking, D., & Kennett, J. (1998). Friendship and the Self. *Ethics*, *108*(3), 502–527. https://doi.org/10.1086/233824

Colombetti, G. (2017). Enactive Affectivity, Extended. *Topoi*, *36*(3), 445–455. https://doi.org/10.1007/s11245-015-9335-2

Conti, J., Heywood, L., Hay, P., Shrestha, R. M., & Perich, T. (2022). Paper 2: A systematic review of narrative therapy treatment outcomes for eating disorders—bridging the divide between practice-based evidence and evidence-based practice. *Journal of Eating Disorders*, *10*, 138. https://doi.org/10.1186/s40337-022-00636-4

Craig, E., Laengle, A., Schneider, K. J., Tantam, D., & Plock, S. du. (2019). *The Wiley World Handbook of Existential Therapy*. John Wiley & Sons.

Crane, T. (1998). Intentionality as the Mark of the Mental. *Royal Institute of Philosophy Supplements*, *43*, 229–251. https://doi.org/10.1017/S1358246100004380

Dahl, J., Luciano, C., & Wilson, K. (2005). *Acceptance and Commitment Therapy for Chronic Pain*. New Harbinger Publications.

Dahl, J., Plumb, J. C., Lundgren, T., & Stewart, I. (2009). *The Art & Science of Valuing in Psychotherapy: Helping Clients Discover, Explore, and Commit to Valued Action Using Acceptance and Commitment Therapy*. New Harbinger Publications.

Damasio, A. (2005). *Descartes' Error: Emotion, Reason, and the Human Brain*. Penguin.

De Araujo, I. E., Rolls, E. T., Velazco, M. I., Margot, C., & Cayeux, I. (2005). Cognitive Modulation of Olfactory Processing. *Neuron*, *46*(4), 671–679. https://doi.org/10.1016/j.neuron.2005.04.021

Dembroff, R., & Saint-Croix, C. (2019). 'Yep, I'm Gay': Understanding Agential Identity. *Ergo, an Open Access Journal of Philosophy*, *6*. https://doi.org/10.3998/ergo.12405314.0006.020

Dembroff, R., & Wodak, D. (2018). He/She/They/Ze. *Ergo, an Open Access Journal of Philosophy*, *5*. https://doi.org/10.3998/ergo.12405314.0005.014

Demiralp, E., Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Barrett, L. F., Ellsworth, P. C., Demiralp, M., Hernandez-Garcia, L., Deldin, P. J., Gotlib, I. H., & Jonides, J. (2012). Feeling Blue or Turquoise? Emotional Differentiation in Major Depressive Disorder. *Psychological Science*, *23*(11), 1410–1416. https://doi.org/10.1177/0956797612444903

Denborough, D. (2014). *Retelling the Stories of Our Lives: Everyday Narrative Therapy to Draw Inspiration and Transform Experience*. W. W. Norton & Company.

Dennett, D. C. (1969). *Content and consciousness*. Routledge and Kegan Paul.

Dennett, D. C. (1989). *The Intentional Stance*. MIT Press.

Deonna, J. A., & Teroni, F. (2015). Emotions as Attitudes. *Dialectica*, *69*(3), 293–311. https://doi.org/10.1111/1746-8361.12116

DeSocio, J. E. (2005). Accessing Self-development through Narrative Approaches in Child and Adolescent Psychotherapy. *Journal of Child and Adolescent Psychiatric Nursing*, *18*(2), 53–61.

Diamond, C. (1997). Moral differences and distances: Some questions. In L. Alanen, S. Heinämaa, & T. Wallgreen (Eds.), *Commonality and particularity in ethics* (pp. 197–234). St. Martin's Press.

Döring, S. A. (2003). Explaining Action by Emotion. *The Philosophical Quarterly*, *53*(211), 214–230. https://doi.org/10.1111/1467-9213.00307

Doris, J. M. (2002). *Lack of Character: Personality and Moral Behavior*. Cambridge University Press.

Doris, J. M. (2015). *Talking to Our Selves: Reflection, Ignorance, and Agency*. OUP Oxford.

Dover, D. (2022). The conversational self. *Mind*, *131*(521), 193–230. https://doi.org/10.1093/mind/fzab069

Dover, D. (2023). Identity and influence. *Synthese*, *202*(5), 167. https://doi.org/10.1007/s11229-023-04279-z

Drayson, Z. (2012). The Uses and Abuses of the Personal/Subpersonal Distinction. *Philosophical Perspectives*, *26*(1), 1–18. https://doi.org/10.1111/phpe.12014

Drayson, Z. (2014). The Personal/Subpersonal Distinction: The Personal/Subpersonal Distinction. *Philosophy Compass*, *9*(5), 338–346. https://doi.org/10.1111/phc3.12124

Dreyfus, H. L. (2013). The myth of the pervasiveness of the mental. In J. K. Schear (Ed.), *Mind, reason, and being-in-the-world: The McDowell-Dreyfus debate* (pp. 15–40). Routledge.

Dworkin, G. (2020). Paternalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2020/entries/paternalism/

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ellsworth, P. (1991). Some Implications of Cognitive Appraisal Theories of Emotion. *Book Chapters*. https://repository.law.umich.edu/book_chapters/369

Ellsworth, P., & Scherer, K. (2003). Appraisal Processes in Emotion. *Book Chapters*. https://repository.law.umich.edu/book_chapters/228

Erbas, Y., Ceulemans, E., Blanke, E. S., Sels, L., Fischer, A., & Kuppens, P. (2019). Emotion differentiation dissected: Between-category, within-category, and integral emotion differentiation, and their relation to well-being. *Cognition and Emotion*, *33*(2), 258–271. https://doi.org/10.1080/02699931.2018.1465894

Erbas, Y., Ceulemans, E., Boonen, J., Noens, I., & Kuppens, P. (2013). Emotion differentiation in autism spectrum disorder. *Research in Autism Spectrum Disorders*, *7*(10), 1221–1227. https://doi.org/10.1016/j.rasd.2013.07.007

Erbas, Y., Ceulemans, E., Lee Pe, M., Koval, P., & Kuppens, P. (2014). Negative emotion differentiation: Its personality and well-being correlates and a comparison of different

assessment methods. *Cognition and Emotion*, *28*(7), 1196–1213. https://doi.org/10.1080/02699931.2013.875890

Etchison, M., & Kleist, D. M. (2000). Review of Narrative Therapy: Research and Utility. *The Family Journal*, *8*(1), 61–66. https://doi.org/10.1177/1066480700081009

Ferzan, K. K. (2015). Consent, Culpability, and the Law of Rape Symposium: Rape Law Revisited. *Ohio State Journal of Criminal Law*, *13*(2), 397–440.

Forman, E. M., Butryn, M. L., Hoffman, K. L., & Herbert, J. D. (2009). An Open Trial of an Acceptance-Based Behavioral Intervention for Weight Loss. *Cognitive and Behavioral Practice*, *16*(2), 223–235. https://doi.org/10.1016/j.cbpra.2008.09.005

Frankfurt, H. (1998). The importance of what we care about. In *The importance of what we care about: Philosophical essays* (pp. 80–94). Cambridge University Press.

Frankfurt, H. (1999). On caring. In *Necessity, volition, and love* (pp. 155–180). Cambridge University Press.

Frankl, V. E. (2014). *The Will to Meaning: Foundations and Applications of Logotherapy*. Penguin.

Frewen, P. A., Dozois, D. J. A., Neufeld, R. W. J., & Lanius, R. A. (2008). Meta-analysis of alexithymia in posttraumatic stress disorder. *Journal of Traumatic Stress*, *21*(2), 243–246. https://doi.org/10.1002/jts.20320

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press. https://academic.oup.com/pq/article-lookup/doi/10.1111/j.1467-9213.2009.594_4.x

Fricker, M., & Jenkins, K. (2017). Epistemic injustice, ignorance, and trans experiences. In A. Garry, S. J. Khader, & A. Stone (Eds.), *The Routledge Companion to Feminist Philosophy* (pp. 268–278). Routledge.

Fridland, E. (2017). Skill and motor control: Intelligence all the way down. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, *174*(6), 1539–1560.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Frijda, N. H. (1993). The place of appraisal in emotion. *Cognition & Emotion*, *7*(3–4), 357–387. https://doi.org/10.1080/02699939308409193

Frijda, N. H. (2009). Emotion Experience and its Varieties. *Emotion Review*, *1*(3), 264–271. https://doi.org/10.1177/1754073909103595

Fulkerson, M. (n.d.). *Perceptual Nudges* [Unpublished manuscript].

Fulkerson, M. (2023). How thirst compels: An aggregation model of sensory motivation. *Mind & Language*, *38*(1), 141–155. https://doi.org/10.1111/mila.12369

Gendlin, E. T. (2012). *Focusing-Oriented Psychotherapy: A Manual of the Experiential Method*. Guilford Press.

Gilligan, C. (1993). *In a different voice: Psychological theory and women's development*. Harvard University Press.

Gimbel, S. (2018). *Isn't that clever: A philosophical account of humor and comedy*. Routledge. https://www.routledge.com/Isnt-that-Clever-A-Philosophical-Account-of-Humor-and-Comedy/Gimbel/p/book/9780367594923

Gingerich, J. (2022). Spontaneous Freedom. *Ethics*, *133*(1), 38–71. https://doi.org/10.1086/720778

Goldie, P. (2000). *The emotions: A philosophical exploration*. Clarendon Press.

Goldie, P. (2003). One's Remembered Past: Narrative Thinking, Emotion, and the External Perspective. *Philosophical Papers*, *32*(3), 301–319. https://doi.org/10.1080/05568640309485129

Gorodeisky, K. (2021). The authority of pleasure. *Noûs*, *55*(1), 199–220. https://doi.org/10.1111/nous.12310

Gregg, J. A., Namekata, M. S., Louie, W. A., & Chancellor-Freeland, C. (2014). Impact of values clarification on cortisol reactivity to an acute stressor. *Journal of Contextual Behavioral Science*, *3*(4), 299–304. https://doi.org/10.1016/j.jcbs.2014.08.002

Grèzes, J., & Decety, J. (2002). Does visual perception of object afford action? Evidence from a neuroimaging study. *Neuropsychologia*, *40*(2), 212–222. https://doi.org/10.1016/S0028-3932(01)00089-6

Grèzes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate action: An fMRI study of implicit processing. *European Journal of Neuroscience*, *17*(12), 2735–2740. https://doi.org/10.1046/j.1460-9568.2003.02695.x

Groll, D. (2012). Paternalism, Respect, and the Will. *Ethics*, *122*(4), 692–720. https://doi.org/10.1086/666500

Hampshire, S., & Hart, H. L. A. (1958). Decision, Intention and Certainty. *Mind*, *67*(265), 1–12.

Harris, D. (2014). *Speech Act Theoretic Semantics* [Doctoral dissertation]. https://academicworks.cuny.edu/gc_etds/222

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2011). *Acceptance and Commitment Therapy, Second Edition: The Process and Practice of Mindful Change*. Guilford Press.

Helm, B. W. (2001). *Emotional reason*. Cambridge University Press.

Helm, B. W. (2009). *Love, Friendship, and the self: Intimacy, identification, and the social nature of persons*. Oxford University Press.

Henckel, S. J. (2025). You Think You Know Someone: Trans Identities and Epistemic Injustice. *Hypatia*, 1–22. https://doi.org/10.1017/hyp.2025.4

Henley, W. E. (1920). Invictus. In *Poems*. Macmillan and Co.

Hilton Andersen, C. (2017, May 23). 19 People Reveal the Exact Moment When They Knew They Were In Love. *Redbook*. https://www.redbookmag.com/love-sex/sex/a50425/moment-people-knew-they-were-in-love/

Hoemann, K., Lee, Y., Kuppens, P., Gendron, M., & Boyd, R. L. (2023). Emotional Granularity is Associated with Daily Experiential Diversity. *Affective Science*, *4*(2), 291–306. https://doi.org/10.1007/s42761-023-00185-2

Hogeveen, J., & Grafman, J. (2021). Chapter 3—Alexithymia. In K. M. Heilman & S. E. Nadeau (Eds.), *Handbook of Clinical Neurology* (Vol. 183, pp. 47–62). Elsevier. https://doi.org/10.1016/B978-0-12-822290-4.00004-9

Honkalampi, K., Hintikka, J., Koivumaa-Honkanen, H., Antikainen, R., Haatainen, K., & Viinam□ki, H. (2007). Long-Term Alexithymic Features Indicate Poor Recovery from Depression and Psychopathology: A Six-Year Follow-Up. *Psychotherapy and Psychosomatics*, *76*(5), 312–314.

Hornsby, J., & Langton, R. (1998). Free Speech and Illocution. *Legal Theory*, *4*(1), 21–37. https://doi.org/10.1017/S1352325200000902

Hsing, C. K., Hofelich Mohr, A., Brent Stansfield, R., & Preston, S. D. (2013). Alexithymia Slows Performance but Preserves Spontaneous Semantic Decoding of Negative

Expressions in the Emostroop Task. *International Journal of Psychological Research*, *6*(SPE), 56–67.

Hu, G., Han, B., Gains, H., & Jia, Y. (2024). Effectiveness of narrative therapy for depressive symptoms in adults with somatic disorders: A systematic review and meta-analysis. *International Journal of Clinical and Health Psychology*, *24*(4), 100520. https://doi.org/10.1016/j.ijchp.2024.100520

Jacobson, H. (2019). Not Only a Messenger: Towards an Attitudinal-Representational Theory of Pain. *Philosophy and Phenomenological Research*, *99*(2), 382–408. https://doi.org/10.1111/phpr.12493

Jaworska, A. (2007a). Caring and Full Moral Standing. *Ethics*, *117*(3), 460–497. https://doi.org/10.1086/512780

Jaworska, A. (2007b). Caring and internality. *Philosophy and Phenomenological Research*, *74*(3), 529–568.

Jaworska, A. (2022). Agnecy and identification. In L. Ferrero (Ed.), *The Routledge handbook of philosophy of agency* (pp. 288–297). Routledge.

Jaworska, A., & Wonderly, M. (2017). Love and caring. In C. Grau & A. Smuts (Eds.), *The Oxford handbook of philosophy of love*. Oxford University Press.

Jenkins, K. (2018). Toward an Account of Gender Identity. *Ergo, an Open Access Journal of Philosophy*, *5*. https://doi.org/10.3998/ergo.12405314.0005.027

Jenkins, K. (2023). *Ontology and Oppression: Race, Gender, and Social Reality*. Oxford University Press.

Jones, K. (2006). Quick and Smart? Modularity and the Pro-Emotion Consensus. *Canadian Journal of Philosophy Supplementary Volume*, *32*, 3–27. https://doi.org/10.1353/cjp.2007.0031

Jones, K. (2008). How to change the past. In C. Mackenzie & K. Atkins (Eds.), *Practical identity and narrative agency* (pp. 269–288). Routledge.

Kane, R. (1996). *The significance of free will*. Oxford Univ. Press.

Kashdan, T. B., Ferssizidis, P., Muraven, M., & Collins, R. L. (2010). Emotion Differentiation as Resilience Against Excessive Alcohol Use. *Psychological Science*, *21*(9), 1341–1347.

Katsafanas, P. (2021). Review of Agnes Callard, Aspiration. *Philosophy and Phenomenological Research*, *102*(2), 464–469. https://doi.org/10.1111/phpr.12760

Kilpi, E. (2014). He stepped over the threshold (D. Adamson, Tran.). In *A landscape blossoms within me*. Arc Publications.

Kircanski, K., Lieberman, M. D., & Craske, M. G. (2012). Feelings Into Words: Contributions of Language to Exposure Therapy. *Psychological Science*, *23*(10), 1086–1091. https://doi.org/10.1177/0956797612443830

Kolodny, N. (2003). Love as Valuing a Relationship. *The Philosophical Review*, *112*(2), 135–189.

Korsgaard, C. M. (1989). Personal Identity and the Unity of Agency: A Kantian Response to Parfit. *Philosophy & Public Affairs*, *18*(2), 101–132.

Korsgaard, C. M. (1996). *The Sources of Normativity*. Cambridge University Press.

Korsgaard, C. M. (2008). *The constitution of agency: Essays on practical reason and moral psychology*. Oxford University Press.

Kriegel, U. (2019). The Intentional Structure of Moods. *Philosopher's Imprint*, *19*(49). http://hdl.handle.net/2027/spo.3521354.0019.049

Kring, A. M., Barrett, L. F., & Gard, D. E. (2003). On the Broad Applicability of the Affective Circumplex: Representations of Affective Knowledge Among Schizophrenia Patients. *Psychological Science*, *14*(3), 207–214. https://doi.org/10.1111/1467-9280.02433

Kruglanski, A. W., Bélanger, J. J., & Gunaratna, R. (2019). *The Three Pillars of Radicalization: Needs, Narratives, and Networks*. Oxford University Press.

Kukla, Q., & Lance, M. (2023). Telling Gender: The Pragmatics and Ethics of Gender Ascriptions. *Ergo an Open Access Journal of Philosophy*, *9*(0). https://doi.org/10.3998/ergo.2911

Kukla, R. (2014). Performative Force, Convention, and Discursive Injustice. *Hypatia*, *29*(2), 440–457. https://doi.org/10.1111/j.1527-2001.2012.01316.x

Kukla, R., & Lance, M. (2009). *'Yo!' and 'Lo!': The Pragmatic Topography of the Space of Reasons*. Harvard University Press.

Laloyaux, J., Fantini, C., Lemaire, M., Luminet, O., & Larøi, F. (2015). Evidence of Contrasting Patterns for Suppression and Reappraisal Emotion Regulation Strategies in Alexithymia. *The Journal of Nervous and Mental Disease*, *203*(9), 709. https://doi.org/10.1097/NMD.0000000000000353

Lance, M., & Kukla, R. (2013). Leave the Gun; Take the Cannoli! The Pragmatic Topography of Second-Person Calls. *Ethics*, *123*(3), 456–478. https://doi.org/10.1086/669565

Langton, R. (1993). Speech Acts and Unspeakable Acts. *Philosophy & Public Affairs*, *22*(4), 293–330.

Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *American Psychologist*, *46*(8), 819–834. https://doi.org/10.1037/0003-066X.46.8.819

Lear, J. (1998). *Love and Its Place in Nature: A Philosophical Interpretation of Freudian Psychoanalysis*. Yale University Press.

LeJeune, J., & Luoma, J. B. (2019). *Values in Therapy: A Clinician's Guide to Helping Clients Explore Values, Increase Psychological Flexibility, and Live a More Meaningful Life*. New Harbinger Publications.

Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General*, *135*(4), 501–512. https://doi.org/10.1037/0096-3445.135.4.501

Levinson, J. (1996). *The pleasures of aesthetics*. Cornell University Press.

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, *22*(6), 1027–1038. https://doi.org/10.1016/j.conb.2012.06.001

Leweke, F., Leichsenring, F., Kruse, J., & Hermes, S. (2011). Is Alexithymia Associated with Specific Mental Disorders. *Psychopathology*, *45*(1), 22–28.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, *8*(1), 339–359. https://doi.org/10.1007/BF00258436

Lieberman, M. D., Inagaki, T. K., Tabibnia, G., & Crockett, M. J. (2011). Subjective responses to emotional stimuli during labeling, reappraisal, and distraction. *Emotion*, *11*(3), 468–480. https://doi.org/10.1037/a0023503

Lillis, J., & Bond, D. S. (2019). Values-based and acceptance-based intervention to promote adoption and maintenance of habitual physical activity among inactive adults with

overweight/obesity: A study protocol for an open trial. *BMJ Open*, *9*(1), e025115. https://doi.org/10.1136/bmjopen-2018-025115

Lindemann, H. (2014). *Holding and letting go: The social practice of personal identities*. Oxford University Press USA - OSO.

Lopes, D. (2018). *Being for beauty: Agency and aesthetic value*. Oxford University Press.

Lundgren, T., Luoma, J. B., Dahl, J., Strosahl, K., & Melin, L. (2012). The Bull's-Eye Values Survey: A Psychometric Evaluation. *Cognitive and Behavioral Practice, Special Series: Acceptance and Commitment Therapy*, *19*(4), 518–526. https://doi.org/10.1016/j.cbpra.2012.01.004

Luoma, J. B., Hayes, S. C., & Walser, R. D. (2017). *Learning ACT: An Acceptance and Commitment Therapy Skills Training Manual for Therapists*. New Harbinger Publications.

Mackenzie, C. (2008). Imagination, identity, and self-transformation. In K. Atkins & C. Mackenzie (Eds.), *Practical identity and narrative agency* (pp. 121–145). Routledge.

Macpherson, F. (2012). Cognitive Penetration of Colour Experience: Rethinking the Issue in Light of an Indirect Mechanism. *Philosophy and Phenomenological Research*, *84*(1), 24–62.

Maitra, I. (2009). Silencing Speech. *Canadian Journal of Philosophy*, *39*(2), 309–338. https://doi.org/10.1353/cjp.0.0050

Mason, E. (2019). *Ways to be Blameworthy: Rightness, Wrongness, and Responsibility*. Oxford University Press.

Matejka, M., Kazzer, P., Seehausen, M., Bajbouj, M., Klann-Delius, G., Menninghaus, W., Jacobs, A. M., Heekeren, H. R., & Prehn, K. (2013). Talking about Emotion: Prosody

and Skin Conductance Indicate Emotion Regulation. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00260

Matherne, S. (2020). Edith Landmann-Kalischer on Aesthetic Demarcation and Normativity. *The British Journal of Aesthetics*, *60*(3), 315–334. https://doi.org/10.1093/aesthj/ayaa007

Matthen, M. (2017). The Pleasure of Art. *Australasian Philosophical Review*, *1*(1), 6–28. https://doi.org/10.1080/24740500.2017.1287034

Mcgowan, M. K. (2004). Conversational Exercitives: Something Else We do with Our Words. *Linguistics and Philosophy*, *27*(1), 93–111. https://doi.org/10.1023/B:LING.0000010803.47264.f0

McGowan, M. K. (2009). Oppressive Speech. *Australasian Journal of Philosophy*, *87*(3), 389–407. https://doi.org/10.1080/00048400802370334

Mendolia, M., & Kleck, R. E. (1993). Effects of talking about a stressful event on arousal: Does what we talk about make a difference? *Journal of Personality and Social Psychology*, *64*(2), 283–292. https://doi.org/10.1037/0022-3514.64.2.283

Merleau-Ponty, M. (2002). *The structure of behavior* (C. Smith, Tran.). Routledge and Kegan Paul.

Miller, C. B. (2013). *Moral Character: An Empirical Theory*. OUP Oxford.

Millgram, E. (1997). *Practical Induction*. Harvard University Press.

Mitchell, J. (2019). The intentionality and intelligibility of moods. *European Journal of Philosophy*, *27*(1), 118–135. https://doi.org/10.1111/ejop.12385

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal Theories of Emotion: State of the Art and Future Development. *Emotion Review*, *5*(2), 119–124. https://doi.org/10.1177/1754073912468165

Moran, R. (2001). *Authority and estrangement: An essay on self-knowledge*. Princeton University Press.

Murdoch, I. (2001). The idea of perfection. In *The sovereignty of good*. Routledge.

Nehamas, A. (1983). How One Becomes What One Is. *The Philosophical Review*, *92*(3), 385–417. https://doi.org/10.2307/2184481

Nguyen, C. T. (2019). Games and the Art of Agency. *The Philosophical Review*, *128*(4), 423–462. https://doi.org/10.1215/00318108-7697863

Nguyen, C. T. (2020). Autonomy and Aesthetic Engagement. *Mind*, *129*(516), 1127–1156. https://doi.org/10.1093/mind/fzz054

Nguyen, C. T. (2023). The Engagement Account of Aesthetic Value. *The Journal of Aesthetics and Art Criticism*, *81*(1), 91–93. https://doi.org/10.1093/jaac/kpac067

Nguyen, C. T. (2024). Value Capture. *Journal of Ethics and Social Philosophy*, *27*(3), 469–504.

Niles, A. N., Craske, M. G., Lieberman, M. D., & Hur, C. (2015). Affect labeling enhances exposure effectiveness for public speaking anxiety. *Behaviour Research and Therapy*, *68*, 27–36. https://doi.org/10.1016/j.brat.2015.03.004

Nussbaum, M. C. (2003). *Upheavals of Thought: The Intelligence of Emotions*. Cambridge University Press.

Onur, E., Alkın, T., Sheridan, M. J., & Wise, T. N. (2013). Alexithymia and Emotional Intelligence in Patients with Panic Disorder, Generalized Anxiety Disorder and Major Depressive Disorder. *Psychiatric Quarterly*, *84*(3), 303–311. https://doi.org/10.1007/s11126-012-9246-y

Oprah Daily. (2022, January 6). *Oprah Talks with Brené Brown About the Power of Being Able to Name Your Emotions*. Oprah Daily. https://www.oprahdaily.com/entertainment/books/a38569758/brene-brown-o-talks/

Orehek, E., & Kruglanski, A. W. (2018). Personal failure makes society seem fonder: An inquiry into the roots of social interdependence. *PLOS ONE*, *13*(8), e0201361. https://doi.org/10.1371/journal.pone.0201361

Orehek, E., Sasota, J. A., Kruglanski, A. W., Dechesne, M., & Ridgeway, L. (2014). Interdependent self-construals mitigate the fear of death and augment the willingness to become a martyr. *Journal of Personality and Social Psychology*, *107*(2), 265–275. https://doi.org/10.1037/a0036675

Owens, D. (2012). *Shaping the Normative Landscape*. Oxford University Press.

Pacherie, E. (2002). Emotion and Action. *European Review of Philosophy*, *5*, 55.

Pacherie, E., & Mylopoulos, M. (2021). Beyond Automaticity: The Psychological Complexity of Skill. *Topoi*, *40*(3), 649–662. https://doi.org/10.1007/s11245-020-09715-0

Páez-Blarrina, M., Luciano, C., Gutiérrez-Martínez, O., Valdivia, S., Ortega, J., & Rodríguez-Valverde, M. (2008). The role of values with personal examples in altering the functions of pain: Comparison between acceptance-based and cognitive-control-based protocols. *Behaviour Research and Therapy*, *46*(1), 84–97. https://doi.org/10.1016/j.brat.2007.10.008

Paul, L. A. (2014). *Transformative experience*. OUP Oxford.

Paul, L. A. (2015). What you can't expect when you're expecting. *Res Philosophica*, *92*(2), 1–23.

Pennebaker, J. W. (1997). Writing About Emotional Experiences as a Therapeutic Process. *Psychological Science*, *8*(3), 162–166. https://doi.org/10.1111/j.1467-9280.1997.tb00403.x

Perls, F. S., Hefferline, R., & Goodman, P. (1994). *Gestalt Therapy Excitement And Growth In The Human Personality*. The Gestalt Journal Press. http://archive.org/details/gestalt-therapy-excitement-and-growth-in-the-human-personality-by-perls-frederic

Pettigrew, R. (2015). Transformative Experience and Decision Theory. *Philosophy and Phenomenological Research*, *91*(3), 766–774.

Pickard, H. (2017). Responsibility without Blame for Addiction. *Neuroethics*, *10*(1), 169–180. https://doi.org/10.1007/s12152-016-9295-2

Pond, R. S. Jr., Kashdan, T. B., DeWall, C. N., Savostyanova, A., Lambert, N. M., & Fincham, F. D. (2012). Emotion differentiation moderates aggressive tendencies in angry people: A daily diary analysis. *Emotion*, *12*(2), 326–337. https://doi.org/10.1037/a0025762

Portner, P. (2004). The Semantics of Imperatives within a Theory of Clause Types. *Semantics and Linguistic Theory*, 235–252. https://doi.org/10.3765/salt.v14i0.2907

Preece, D. A., Mehta, A., Becerra, R., Chen, W., Allan, A., Robinson, K., Boyes, M., Hasking, P., & Gross, J. J. (2022). Why is alexithymia a risk factor for affective disorder symptoms? The role of emotion regulation. *Journal of Affective Disorders*, *296*, 337–341. https://doi.org/10.1016/j.jad.2021.09.085

Preece, D. A., Mehta, A., Petrova, K., Sikka, P., Bjureberg, J., Becerra, R., & Gross, J. J. (2023). Alexithymia and emotion regulation. *Journal of Affective Disorders*, *324*, 232–238. https://doi.org/10.1016/j.jad.2022.12.065

Prinz, J. (2003). Emotion, Psychosemantics, and Embodied Appraisals. *Royal Institute of Philosophy Supplements*, *52*, 69–86. https://doi.org/10.1017/S135824610000789X

Proverbio, A. M., Adorni, R., & D'Aniello, G. E. (2011). 250ms to code for action affordance during observation of manipulable objects. *Neuropsychologia*, *49*(9), 2711–2717. https://doi.org/10.1016/j.neuropsychologia.2011.05.019

Railton, P. (2017). At the Core of Our Capacity to Act for a Reason: The Affective System and Evaluative Model-Based Learning and Control. *Emotion Review*, *9*(4), 335–342. https://doi.org/10.1177/1754073916670021

Ramirez, G., & Beilock, S. L. (2011). Writing About Testing Worries Boosts Exam Performance in the Classroom. *Science*, *331*(6014), 211–213. https://doi.org/10.1126/science.1199427

Ransom, M., Fazelpour, S., Markovic, J., Kryklywy, J., Thompson, E. T., & Todd, R. M. (2020). Affect-biased attention and predictive processing. *Cognition*, *203*, 104370. https://doi.org/10.1016/j.cognition.2020.104370

Renzo, M. (2022). Defective Normative Powers: The Case of Consent. *Journal of Practical Ethics*, *10*(1). https://doi.org/10.3998/jpe.2382

Reuter, K., & Messerli, M. (2018). Transformative Decisions. *The Journal of Philosophy*, *115*(6), 313–335.

Richard Ridderinkhof, K., Forstmann, B. U., Wylie, S. A., Burle, B., & van den Wildenberg, W. P. M. (2011). Neurocognitive mechanisms of action control: Resisting the call of the Sirens. *WIREs Cognitive Science*, *2*(2), 174–192. https://doi.org/10.1002/wcs.99

Ricks, L., Kitchens ,Sarah, Goodrich ,Tonia, & and Hancock, E. (2014). My Story: The Use of Narrative Therapy in Individual and Group Counseling. *Journal of Creativity in Mental Health*, *9*(1), 99–110. https://doi.org/10.1080/15401383.2013.870947

Ridge, M. (2022). Why So Serious? The Nature and Value of Play. *Philosophy and Phenomenological Research*, *105*(2), 406–434. https://doi.org/10.1111/phpr.12826

Rietveld, E. (2008). Situated Normativity: The Normative Aspect of Embodied Cognition in Unreflective Action. *Mind*, *117*(468), 973–1001. https://doi.org/10.1093/mind/fzn050

Riggle, N. (2022). Convergence, Community, and Force in Aesthetic Discourse. *Ergo an Open Access Journal of Philosophy*, *8*(0). https://doi.org/10.3998/ergo.2248

Riggle, N. (2024a). Aesthetic Value and the Practice of Aesthetic Valuing. *The Philosophical Review*, *133*(2), 113–149. https://doi.org/10.1215/00318108-11251345

Riggle, N. (2024b). Autonomy and aesthetic valuing. *Philosophy and Phenomenological Research*, *109*(1), 391–409. https://doi.org/10.1111/phpr.13045

Rilke, R. M. (1982). The Archaic Torso of Apollo (S. Mitchell, Tran.). In S. Mitchell (Ed.), *Selected poetry of Rainer Marie Rilke*. Random House. http://www.ekphrastic.net/1/post/2017/07/the-archaic-torso-of-apollo-by-rainer-maria-rilke-translated-by-william-ruleman.html

Risafi, N. (Producer). (2025). *How US Neo-Nazism Actually Works | Authorized Account | Insider* [Video recording]. https://www.youtube.com/watch?v=d-g3Z8IWsdU

Rossi, M. (2021). A perceptual theory of moods. *Synthese*, *198*(8), 7119–7147. https://doi.org/10.1007/s11229-019-02513-1

Rovane, C. A. (1998). *The bounds of agency: An essay in revisionary metaphysics*. Princeton University Press.

Russell, F. (2026). Moral psychology as soul-picture. *The Philosophical Quarterly*, *76*(1), 301–319.

Schechtman, M. (1996). *The constitution of selves*. Cornell university press.

Scheffler, S. (2010). Valuing. In *Equality and tradition: Questions of value in moral and political theory* (pp. 15–40). Oxford University Press.

Schiffer, S. (1972). *Meaning*. Oxford University Press.

Schopenhauer, A., & Payne, E. F. J. (1819). *The world as will and representation*. Dover.

Scoccia, D. (2018). The concept of paternalism. In *The Routledge Handbook of the Philosophy of Paternalism*. Routledge.

Seidman, J. (2009). Valuing and caring. *Theoria*, *75*, 272–303.

Seidman, J. (2016). The unity of caring and the rationality of emotion. *Philosophical Studies*, *173*, 2785–2801.

Seo, M., Kang, H. S., Lee, Y. J., & Chae, S. M. (2015). Narrative therapy with an emotional approach for people with depression: Improved symptom and cognitive-emotional outcomes. *Journal of Psychiatric and Mental Health Nursing*, *22*(6), 379–389. https://doi.org/10.1111/jpm.12200

Shiffrin, S. V. (2000). Paternalism, Unconscionability Doctrine, and Accommodation. *Philosophy & Public Affairs*, *29*(3), 205–250. https://doi.org/10.1111/j.1088-4963.2000.00205.x

Shoemaker, D. W. (2003). Caring, Identification, and Agency. *Ethics*, *114*(1), 88–118. https://doi.org/10.1086/376718

Silins, N. (2016). Cognitive Penetration and the Epistemology of Perception`. *Philosophy Compass*, *11*(1), 24–42.

Smidt, K. E., & Suvak, M. K. (2015). A brief, but nuanced, review of emotional granularity and emotion differentiation research. *Current Opinion in Psychology, Emotion Regulation*, *3*, 48–51. https://doi.org/10.1016/j.copsyc.2015.02.007

Solomon, R. C. (1993). *The Passions: Emotions and the Meaning of Life*. Hackett Publishing.

Stalnaker, R. (2002). Common Ground. *Linguistics and Philosophy*, *25*(5/6), 701–721.

Stang, N. F. (2012). Artworks Are Not Valuable for Their Own Sake. *The Journal of Aesthetics and Art Criticism*, *70*(3), 271–280. https://doi.org/10.1111/j.1540-6245.2012.01519.x

Stichter, M., & Fridland, E. (2025). Emotional Self-Knowledge: How Affective Skills Reveal Our Values, Goals, Cares, and Concerns. *Elements in Epistemology*. https://doi.org/10.1017/9781009307857

Suvak, M. K., Litz, B. T., Sloan, D. M., Zanarini, M. C., Barrett, L. F., & Hofmann, S. G. (2011). Emotional granularity and borderline personality disorder. *Journal of Abnormal Psychology*, *120*(2), 414–426. https://doi.org/10.1037/a0021808

Tappolet, C. (2016). *Emotions, Values, and Agency*. Oxford University Press.

Taylor, C. (1985a). *Philosophical papers I: Human agency and language*. Cambridge University Press.

Taylor, C. (1985b). What is human agency? In *Philosophical papers 1: Human agency and language* (pp. 15–44). Cambridge University Press.

Taylor, G. J., Bagby, R. M., & Parker, J. D. A. (1997). *Disorders of affect regulation: Alexithymia in medical and psychiatric illness* (pp. xxii, 359). Cambridge University Press. https://doi.org/10.1017/CBO9780511526831

Thompson, R. J., Springstein, T., & Boden, M. (2021). Gaining clarity about emotion differentiation. *Social and Personality Psychology Compass*, *15*(3), e12584. https://doi.org/10.1111/spc3.12584

Todd, R. M., Miskovic, V., Chikazoe, J., & Anderson, A. K. (2020). Emotional Objectivity: Neural Representations of Emotions and Their Interaction with Cognition. *Annual*

Review of Psychology, *71*(Volume 71, 2020), 25–48. https://doi.org/10.1146/annurev-psych-010419-051044

Torre, J. B., & Lieberman, M. D. (2018). Putting Feelings Into Words: Affect Labeling as Implicit Emotion Regulation. *Emotion Review*, *10*(2), 116–124. https://doi.org/10.1177/1754073917742706

Tsai, G. (2014). Rational Persuasion as Paternalism. *Philosophy & Public Affairs*, *42*(1), 78–112. https://doi.org/10.1111/papa.12026

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 830–846. https://doi.org/10.1037/0096-1523.24.3.830

Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, *116*(2), 185–203. https://doi.org/10.1016/j.actpsy.2004.01.004

Turyn, G. (2023). Gender and first-person authority. *Synthese*, *201*(4), 122. https://doi.org/10.1007/s11229-023-04125-2

Ullmann-Margalit, E. (2006). Big decisions: Opting, converting, drifting. *Royal Institute of Philosophy Supplements*, *58*, 157–172. https://doi.org/10.1017/S1358246106058085

Veage, S., Ciarrochi, J., Deane, F. P., Andresen, R., Oades, L. G., & Crowe, T. P. (2014). Value congruence, importance and success and in the workplace: Links with well-being and burnout amongst mental health practitioners. *Journal of Contextual Behavioral Science*, *3*(4), 258–264. https://doi.org/10.1016/j.jcbs.2014.06.004

Velleman, J. D. (2009). *How We Get Along*. Cambridge University Press.

Velleman, J. D. (2020a). Motivation by ideal. In *Self to self* (Second edition, pp. 415–436). Michigan Publishing.

Velleman, J. D. (2020b). The self as narrator. In *Self to self* (Second edition, pp. 277–303). Michigan Publishing.

Vlasenko, V. V., Rogers, E. G., & Waugh, C. E. (2021). Affect labelling increases the intensity of positive emotions. *Cognition and Emotion*, *35*(7), 1350–1364. https://doi.org/10.1080/02699931.2021.1959302

Vromans, L. P., & and Schweitzer, R. D. (2011). Narrative therapy for adults with major depressive disorder: Improved symptom and interpersonal outcomes. *Psychotherapy Research*, *21*(1), 4–15. https://doi.org/10.1080/10503301003591792

Walden, K. (n.d.). *The imperative of genius* [Unpublished manuscript].

Wallace, R. J. (2018). *The View from Here: On Affirmation, Attachment, and the Limits of Regret*. Oxford University Press.

Walton, K. (1993). How marvelous!: Towards a theory of aesthetic value. *Journal of Aesthetics and Art Criticism*, *51*, 499–510.

Webber, D., Chernikova, M., Kruglanski, A. W., Gelfand, M. J., Hettiarachchi, M., Gunaratna, R., Lafreniere, M.-A., & Belanger, J. J. (2018). Deradicalizing Detained Terrorists. *Political Psychology*, *39*(3), 539–556. https://doi.org/10.1111/pops.12428

Westlund, A. C. (2023). Authors, narrators, and autonomous agents: The art of relational autobiography. *The Southern Journal of Philosophy*, *61*(S1), 50–61. https://doi.org/10.1111/sjp.12531

Westra, E. (2020). Getting to know you: Accuracy and error in judgments of character. *Mind & Language*, *35*(5), 583–600. https://doi.org/10.1111/mila.12258

Westra, E. (2022). In Defense of Ordinary Moral Character Judgment. *Erkenntnis*, *87*(4), 1461–1479. https://doi.org/10.1007/s10670-020-00257-w

White, M. (2011). *Narrative Practice: Continuing the Conversations*. W. W. Norton & Company.

White, M. (2024). *Maps of Narrative Practice*. W. W. Norton & Company.

White, M., & Epston, D. (1990). *Narrative Means To Therapeutic Ends*. W. W. Norton & Company.

Whiteley, E. K. (2023). "A Woman First and a Philosopher Second": Relative Attentional Surplus on the Wrong Property. *Ethics*, *133*(4), 497–528. https://doi.org/10.1086/724538

Wilson, K. G. (2009). *Mindfulness for Two: An Acceptance and Commitment Therapy Approach to Mindfulness in Psychotherapy*. New Harbinger Publications.

Wilson, K. G., & Sandoz, E. K. (2008). Mindfulness, values, and the therapeutic relationship in acceptance and commitment therapy. In S. Hick & T. Bien (Eds.), *Mindfulness and the therapeutic relationship* (pp. 89–106). Guilford Press.

Wilson, K. G., Sandoz, E. K., Kitchens, J., & Roberts, M. (2010). The Valued Living Questionnaire: Defining and Measuring Valued Action within a Behavioral Framework. *The Psychological Record*, *60*(2), 249–272. https://doi.org/10.1007/BF03395706

Wolf, S. (2000). The moral of moral lcuk. *Philosophical Exchange*, *31*, 4–19.

Wonderly, M. L. (2016). On being attached. *Philosophical Studies*, *173*(1), 223–242. https://doi.org/10.1007/s11098-015-0487-0

Woods, D. W., Wetterneck, C. T., & Flessner, C. A. (2006). A controlled evaluation of acceptance and commitment therapy plus habit reversal for trichotillomania. *Behaviour Research and Therapy*, *44*(5), 639–656. https://doi.org/10.1016/j.brat.2005.05.006

Wu, W. (2011). Confronting Many-Many Problems: Attention and Agentive Control. *Noûs*, *45*(1), 50–76. https://doi.org/10.1111/j.1468-0068.2010.00804.x

Wu, W. (2023). *Movements of the mind*. Oxford University PressOxford. https://doi.org/10.1093/oso/9780192866899.003.0001

Yao, V. (2023). Eros and anxiety. *Synthese*, *202*(6), 200. https://doi.org/10.1007/s11229-023-04391-0

Yeats, W. B. (1889). The Lake Isle of Innisfree. In *The Wanderings of Oisin and Other Poems*. Kegan Paul, Trench and Co.

Zaki, L. F., Coifman, K. G., Rafaeli, E., Berenson, K. R., & Downey, G. (2013). Emotion Differentiation as a Protective Factor Against Nonsuicidal Self-Injury in Borderline Personality Disorder. *Behavior Therapy*, *44*(3), 529–540. https://doi.org/10.1016/j.beth.2013.04.008

Zamariola, G., Vlemincx, E., Corneille, O., & Luminet, O. (2018). Relationship between interoceptive accuracy, interoceptive sensibility, and alexithymia. *Personality and Individual Differences*, *125*, 14–20. https://doi.org/10.1016/j.paid.2017.12.024