

**THE VALUING THEORY OF PERSONAL IDENTITY OVER TIME: PHILOSOPHICAL
ANALYSIS AND PRACTICAL APPLICATIONS**

Nena Therese Davis

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy.

Chapel Hill
2012

Approved By:

Susan R. Wolf

Thomas E. Hill, Jr.

Jesse J. Prinz

C. D. C. Reeve

Bernard R. Boxill

© 2012
Nena Therese Davis
ALL RIGHTS RESERVED

ABSTRACT

NENA THERESE DAVIS: The Valuing Theory of Personal Identity over Time:
Philosophical Analysis and Practical Applications
(Under the direction of Susan R. Wolf)

Under what conditions and in virtue of what does a human being remain the same person over time, despite important changes in her psychology? Many ethical and policy issues are connected to this question. For example: Should we consider a patient in the middle or late stages of Alzheimer's disease to be the same person who authored an advance directive when that patient was legally competent? Should we be concerned about altering the identity of an individual when that individual receives psychoactive drugs or drugs that have psychoactive side effects?

In this dissertation, I propose a new theory of personal identity over time that avoids the pitfalls of previous theories. According to my theory, a human being is the same person over time if and only if that human being continues to value as that person did previously or what and how the human being values changes as a result of that person's reflection.

This project ends with some early applications of my theory, including considering whether changes in a human being's appetites, moods, and connections to reality sometimes transform that human being into another person.

FOR MY FAMILY

PREFACE

My interest in the problem of personal identity and the problem of the personal identities of dementia patients began in undergraduate philosophy courses at Macalester College. I ended my time at Macalester writing an honor's thesis on personal identity and dementia. As the culmination of my graduate education, this project reveals how the support of so many has enabled me to grow from a budding to a professional philosopher.

I owe many thanks to my committee members. First and foremost, I would like to thank my adviser, Susan Wolf, whose tireless mentorship, guidance, encouragement, and no-nonsense feedback enabled me to complete this project and to complete it while in absentia. This project would not be what it is – something of which I am exceedingly proud – without Susan's support. I must thank Tom Hill and Jesse Prinz for giving me helpful and encouraging feedback on drafts. I would like to thank David Reeve for important suggestions for improvements to a later draft. I must thank Bernie Boxill for his humor and encouraging words during the final days of the project's completion. I would like to thank all my committee members for earlier encouragement and guidance throughout my time at UNC – Chapel Hill, especially Tom and Bernie, as my adviser and second committee member for my Master's thesis, respectively. Finally, I must thank my committee for an invaluable parting gift from UNC: a most informative oral defense offering insightful thoughts that will help me as I further develop the ideas in this project.

I would like to thank Macalester College and my undergraduate professors at Macalester, Joy Laine and Martin Gunderson, who first introduced me to the problem of

personal identity and the problem of the personal identities of dementia patients, and who helped me to develop my earliest thoughts on these problems. I would like to thank UNC – Chapel Hill’s Department of Philosophy and UNC’s Graduate School for creating an excellent environment in which to develop into a professional.

I must thank my father-in-law Anthony, a professional philosopher, for his informal mentorship, encouragement, and advice. I would like to thank my friend Nicoletta Orlandi, a member of my cohort, for her thoughts and encouragement on later portions of this project. I must thank Jan Boxill for her support, encouragement, and mentorship throughout my time at UNC – Chapel Hill.

I would like to thank the many people who helped me in discovering errors in my document, especially my adviser, Susan. I must also thank Heather McAlpin and Suann Wentworth for their essential editing assistance in the final months before submission. Finally, I must thank many of my family and friends who helped with editing in the last days prior to submission. Despite all of their help, I am certain that I have missed errors and such errors are entirely my own.

I must thank my family and friends for their loving support and encouragement. Thanks to my father, Nathaniel, his sister, Frances, and my brother, Chris. My father, who earned a MS in library science, was especially encouraging as I pursued my PhD. I must also thank my three close friends – Jared, Deanna, and Kathy– for their loving support and encouragement.

I would like to thank my mother, Sandra, and her parents, Aubrey DeCosta Matthews, Sr. and Mattie McGuire Matthews, who together raised me and always encouraged my academic pursuits. Thanks to my mother for her love and for showing

me the importance of being a woman with a post-graduate degree through having a career as a librarian with a MS in library science. Aubrey Sr. and Mattie were individuals that valued higher education but as African-Americans born in the 1920s, they did not have the opportunities to pursue fully such an education. Though they are no longer here, I am happy that both saw me enter graduate school and made it clear to me that they never doubted that I would be the first in both my paternal and maternal families to earn a PhD. I would like to thank my mother's siblings – Aubrey Jr., Toney, and Jeffery – and each of their families for their unwavering love and support. I would not have been able to accomplish this personal and family milestone without the love of the entire Matthews clan.

I owe a special thanks to my husband Anthony Graybosch for his love and for reading portions of this project during various stages and offering his invaluable insights. I also must thank him for being a loving father to our now four-year-old daughter, Josephine Cassandra.

Finally, thank you Josephine for being here and being the greatest loving light in my life.

TABLE OF CONTENTS

INTRODUCTION	1
SECTION I	7
CHAPTER 1: PSYCHOLOGICAL THEORIES OF PERSONAL IDENTITY OVER TIME – PARFIT AND BEYOND	7
Introduction.....	7
Derek Parfit’s Neo-Lockean Theory of Personal Identity over Time.....	9
Parfit’s theory of personal identity	9
Critique of Parfit’s theory of personal identity	14
Marya Schechtman: Personal Identity and Narrative	21
Schechtman’s theory of personal identity	21
Critique of Schechtman’s theory of personal identity	23
Christine Korsgaard’s Work on Identity	30
Carol Rovane: Personal Identity and a Rational Point of View.....	34
Rovane’s theory of personal identity over time	34
Critique of Rovane’s theory of personal identity	40
Conclusion	44
CHAPTER 2: A NON-METAPHYSICAL APPROACH TO PERSONAL IDENTITY	47
Introduction.....	47
Metaphysics, most generally.....	47
The Metaphysics of Persons	49
Shifting the Questions from the Metaphysical to the Practical and Ethics	57

Conclusion	62
SECTION I - CONCLUSION	63
The Questions	63
Working Definition of ‘Person’	64
Criteria for a Successful Project	64
SECTION II: THE VALUING THEORY OF PERSONAL IDENTITY OVER TIME	67
Introduction of Preliminary Version	67
CHAPTER 3: THE NATURE OF VALUING	74
Introduction	74
Capacities: Emotion, Reason, Valuing	76
The Capacity to have Emotions	76
The Capacity to Reason	77
The Capacity to Value	79
Conclusion: Valuing and the Valuing Theory	82
The Products of Valuing	83
Charge: Valuing, Disvaluing, and Neutrality	83
To VaND & VND attitudes	84
Reasons for both Emotions and Actions	85
Possible Emotions Necessarily, Possible Actions Likely	86
Emotions may facilitate actions	87
Subjective but not private	88
Degrees of Awareness	88
Conclusion	89
Different ways of VaNDing v	89
Conclusion	91

CHAPTER 4 – VALUING AND THE SYNCHRONIC UNITY OF A PERSON	93
Introduction.....	93
A Glimpse of Synchronic Unity through VaNDing	97
Direct valuing connections	101
Direct reason-based valuing connections.....	102
Direct Event-based Valuing connections: Clashes and Synergies.....	105
Direct Valuing Connections: Had with varying degrees of awareness....	111
The Synchronic Unity of a Person	113
Synchronic Unity – Not Merely of a Moment	115
Conclusion – From Synchronic to Diachronic Unity	117
CHAPTER 5: DIACHRONIC UNITY – CHANGING WAYS OF VALUING.....	119
Final Version of the Theory	119
Sally: An Illustration.....	125
Variation in how persons change.....	130
Self-Narrating Optional: Inclusive of Episodics.....	131
Transitioning on a Consciousness Spectrum	135
Indirect to Direct or Vice Versa.....	137
Meticulously v “Quick & Loose”	141
With or Without Surprise.....	142
Pleasures	144
Conclusion	146
SECTION III: EARLY APPLICATIONS	148
CHAPTER 6: VALUING THEORY OF PERSONAL IDENTITY	
OVER TIME – EARLY APPLICATIONS	148
Introduction.....	148
An Objection.....	148

Plan for a Response.....	149
Akrasia and Moods Involve Emotions.....	149
Akrasia: Internal Conflicts.....	151
Moods: Akrasia and More	164
Moods – a tentative analysis	168
Two further points on moods.....	171
Conclusion & Depression, Mania, Psychosis, and Shifting Appetites	172
CONCLUSION – PROJECT SUMMARY AND CONCLUDING REMARKS	178
Summary	178
Strengths over relevant views discussed in Section I	179
Lingering Objections and Replies.....	182
Rationality is sufficient, Part i	182
Rationality is sufficient, Part ii	184
Emotions are sufficient for valuing.....	185
Memory.....	186
In Closing – The original practical questions	188
APPENDIX A: THE CAPACITY TO HAVE EMOTIONS.....	193
Introduction.....	193
Feelings: Bodily Sensations.....	193
Emotions partly Bodily Sensations & Partly Judgments	195
Judgments alone are not enough.....	196
Judgments account for Intentionality & Appropriateness of Emotions.....	196
Partly Judgments and Partly Bodily Sensations together account for Families of Emotions	198
How Non-Human Animals are Judging via Emotions	198
Optional Acceptance of Composite View.....	199

BIBLIOGRAPHY	202
---------------------------	------------

INTRODUCTION

I want to understand the identities of patients who experience various psychological changes due to disease, defect, injury, or exposure to various substances. The most vivid case for me concerns those with Alzheimer's disease – who are these patients as the illness progresses? Consider the case of Margo.

Margo (an alias) was a fifty-five year old patient with early-onset Alzheimer's disease when Andrew Ferlik, a medical student, met her. Ferlik began to visit and observe her on a daily basis. In an article, he described his observations of Margo.¹

Presumably, due to increasing problems with her memory, Margo led a very repetitive life. She listened to the same songs repeatedly, and seemed to Ferlik to enjoy the songs as if the last time she heard them was several years rather than several minutes ago. She told Ferlik that she liked to read mystery novels but Ferlik “noticed that her place in the book jump[ed] randomly from day-to-day; dozens of pages [were] dog-eared at any given moment.” Ferlik wondered if “for Margo reading [itself was] always a mystery.”² While Margo always greeted him, she never called him by name. Ferlik was never quite sure if she really knew him or if she was simply being polite. Margo also attended an art class for Alzheimer's patients. Her teacher told Ferlik that Margo had

¹ Andrew Ferlik, “Margo's Logo,” *JAMA - Journal of the American Medical Association*, (Vol. 265, No. 2 (January 9, 1991)), 201

² Ferlik: 201.

been painting the same abstract picture “ ‘religiously every day, exactly the same way, for at least the last five years.’ ”³

Ferlik observed that Margo seemed to get simple pleasure from many things. In addition to enjoying music, “reading”, and painting, Margo enjoyed peanut butter and jelly on white bread. Ferlik often would bring her such sandwiches because she loved them so much. “She can devour two in one sitting, especially if milk is at hand.”⁴ Her enjoyment of life confused Ferlik:

Despite her illness, or maybe somehow because of it, Margo is undeniably one of the happiest people I have ever known. There is something graceful about the degeneration her mind is undergoing, leaving her carefree, always cheerful. Do her problems, whatever she may perceive them to be, simply fail to make it to the worry centers of her brain? How does Margo maintain her sense of self? When a person can no longer accumulate new memories as the old rapidly fade, what remains? Who is Margo?⁵

In *Life’s Dominion*, Ronald Dworkin supposes that Margo left a Do Not Resuscitate order (DNR order) while she was still legally competent after her doctors diagnosed her with Alzheimer’s disease. In this hypothetical DNR order, she declared that if she ever reached a state similar to that in which Ferlik observed her, she would not want to be resuscitated. I believe that whether the patient Margo, a human being, persists as the same person is central to helping us decide whether we should honor such a DNR order.

We honor DNR orders for a given human being because we wish to respect the person who formed that DNR order. There are cases when it is clear that the human being who could be resuscitated is the same person who signed the DNR order. These

³ Ferlik: 201.

⁴ Ferlik: 201.

⁵ Ferlik: 201.

are cases when the patient's medical problems do not compromise his mental capacities. One example is a patient suffering through the final stages of a terminal illness who no longer wishes to suffer further. Unless one is against honoring DNR orders in all or most circumstances this seems to me to be clear cases when a DNR order should be honored.

Often DNR orders are carried out after the human being will no longer have any type of experience and lacks personhood, e.g., the human being is in a persistent vegetative state, permanent coma, or the end stage of Alzheimer's disease when the patient is bedridden and unresponsive. In these cases, since the patient is no longer a person and has no experiences, we have reason not to resuscitate: arguably,⁶ she no longer has a life worth living and neither the person that the human being once was nor the non-person that the human being is now has any interest in the resuscitation of this human being. In other cases, though, deciding what to do partly does turn on whether the person persists.

In contrast to cases when the person who signed the DNR order is suffering from a painful terminal illness or the patient lacks both personhood and the capacity to have any experiences, Margo suffers from a non-terminal disease that compromises her mental faculties and she is a human being who has a pleasant existence during the later stages of that disease. Understanding whether she is the same person is important to deciding what to do. Generally, unless someone is properly choosing on the behalf of the patient, we hold that it is immoral to choose not to resuscitate one person because a *different* person ordered it. No one will honor or no one holds that we should honor Phillip's order not to

⁶ For such an argument see Peter Singer, "Is the Sanctity of Life Ethic Terminally Ill?", *Bioethics*, Vol. 9, Nos. 3/4 (1995), 327-343.

resuscitate Roger because Phillip thinks that Roger has shallow projects which make his life lack meaning.

In order to help address the moral question of when we should honor and when we should not honor the DNR orders of Alzheimer's patients like Margo, in this project I develop and defend a new theory of personal identity over time. Given my interest in understanding the identities of patients, most broadly, I am interested in understanding the persistence of the persons that we interact with everyday – human beings who are persons. So, take any particular human being who *does* persist as the same person, I want to know the necessary and sufficient conditions for that human being persisting as the same person.

This theory does not provide a metaphysical analysis, which provides an analysis of personal identity that has the logical form of an identity relation. In Section I: Defining the Project, I discuss the work of other philosophers and then define the questions that my project will address. In Chapter 1, I consider several theories of personal identity and explain why they are problematic. In Chapter 2, I discuss the metaphysical questions that many philosophers address and clearly define the practical questions of this project.

In this project I argue that, given the same human being, that human being is also the same person if and only if P.

P: that human being continues to value as he did previously or changes how he values through his own reflection, a type of reflection which involves using both the capacity to reason and the capacity to have emotion.

I recognize that defending a necessary and sufficient condition of personal identity over time is ambitious. In all honesty, I did not wish to be ambitious and considered making P only a sufficient condition or explaining why P needed to occur not always but “usually” – which would itself need to be defined – in order for a human being to persist as the same person. I defend P as a necessary and sufficient condition because I could not find reasons to make P only a sufficient condition. For a number of kinds of cases, different people might have the P-is-not-necessary Intuition:

P-is-not-necessary Intuition: the intuition that a case under consideration is a case in which a human being persisted as the same person though how he valued did not change due to his own reflection.

When I thought through several kinds of cases where someone might have the P-is-not-necessary Intuition, I would conclude that, for that kind of case, the P-is-not-necessary Intuition did not survive close scrutiny of the case: either the person had changed himself through his own reflection or the original person was displaced by another person. I realize that I am limited by my own imagination and may have simply overlooked some example(s) that would call into question the necessity of P for the persistence of persons. I am certain that I still need to form additional arguments in order to fully defend P as a necessary and sufficient condition of personal identity over time, and I may ultimately conclude that P is only a sufficient condition. In this project, though, my working hypothesis is that P is the necessary and sufficient condition for a human being to persist as the same person.

I develop and defend my theory in Section II: The Valuing Theory of Personal Identity over Time, which consists of Chapters 3 through 5. Chapter 3 defines the nature

of valuing, Chapter 4 provides an analysis of the synchronic unity of persons, and Chapter 5 builds upon synchronic unity to define diachronic unity, i.e., personal identity over time. The defense of the theory continues into the last part of the project, Section III: Early Applications, which consists solely of Chapter 6. Chapter 6 focuses on how, on the valuing theory, personal identity relates to akrasia and moods. In this chapter I also discuss the valuing theory's implications regarding the personal identities of human beings who, due to any number of causes, experience changes in their moods, their appetites, or their connections to reality (going to or from psychosis). With akrasia as well as changing moods, appetites, and connection to reality, I argue that, on the valuing theory, the same person would persist. The aim of much of Chapter 5 and all of Chapter 6 is to provide arguments and defenses against potential or actual P-is-not-necessary Intuitions.

This project ends by considering what I can say about Margo's identity using the valuing theory and discusses areas for future work.

SECTION I

CHAPTER 1: PSYCHOLOGICAL THEORIES OF PERSONAL IDENTITY OVER TIME – PARFIT AND BEYOND

Introduction

What makes someone one and the same person over time? In *An Essay Concerning Human Understanding*, John Locke argued that the identities of persons could not be equated with the identities of human beings and that personal identity was neither a matter of bodily identity nor of the identity across time of any other substance, such as immaterial souls. Instead, according to Locke, a person's identity over time consists in aspects of her psychology, specifically, her consciousness and experience-memories.¹

In 1984, Derek Parfit published his seminal work *Reasons and Persons* in which he argued for a psychological theory of personal identity over time. Parfit takes a Neo-Lockean position regarding personal identity over time, expanding Locke's focus on experience-memory to include all aspects of human psychology.

In this chapter, I will discuss several psychological theories of personal identity over time, beginning with Parfit's theory. In the first section of this chapter, I will briefly discuss Parfit's theory and the criticism of his theory that it fails to take seriously that fact that persons are agents. Then, in the following three sections, I will discuss three post-Parfitian theories of personal identity over time, theories which emphasize the agency of

¹ John Locke, *An Essay concerning Human Understanding*, ed. Nidditich (Oxford: Carendon Press, 1979), 339.

persons. I will argue that, while those theories provide valuable insights into persons and personal identity, none provides a satisfactory view of personal identity over time.

Before turning to Parfit, I would like to clarify two philosophical terms. By ‘psychological theories of personal identity over time’ I do not mean theories of personal identity which are defended by psychologists. Rather, I mean philosophical theories which hold that the identity of a person over time consists in some aspect(s) of her psychology. In chapter two, I will discuss physical theories of personal identity over time. By ‘physical theories of personal identity’ I do not mean theories presented by physical scientists or medical doctors. Instead, physical theories of personal identity over time are philosophical theories which hold that personal identity over time consists in something physical, such as the human body, or brain. In this chapter, I will only focus on Parfit and several post-Parfitian theories which are all psychological theories of personal identity over time. The purpose of this chapter is to discuss Parfit and several theories that try to move beyond Parfit by making agency central to their theories of personal identity over time. This will lay the groundwork for the introduction of my own theory, which also makes agency central to personal identity over time.

*Derek Parfit's Neo-Lockean Theory of Personal Identity over Time*²

Parfit's theory of personal identity

At the beginning of the section on personal identity in *Reasons and Persons*,

Derek Parfit writes:

There is much philosophical debate about the nature both of persons and of personal identity over time. It will help to distinguish these questions:

- (1) What is the nature of persons?
- (2) What is it that makes a person at two different times one and the same person?
What is necessarily involved in the continued existence of each person over time?

The answers to (2) can take this form: 'X today is one and the same person as y at some past time *if and only if* . . .' Such an answer states the *necessary and sufficient conditions* for personal identity over time.^{3 4}

Parfit provides us with several definitions before formulating his own theory of personal identity. He starts by holding that the **psychological features** of our lives, such as memories, beliefs, desires, and intentions are central in our lives as persons. This is an important expansion of Locke's theory, which focuses only on experience-memories.⁵

For Parfit, when two psychological features at two different times are qualitatively identical or when the earlier psychological feature accounts for the existence

² Parfit formulates what he holds is the best theory of personal identity in order to later reject the importance that we place on personal identity, and by extension, survival. In this section, I simply present Parfit's theory of personal identity over time. I leave out his arguments in defense of the theory and his later arguments that, upon fully understanding personal identity, we should not see it as important as we do. See Footnotes 11 and 14 for some details on Parfit's claim that personal identity is less important than we might think.

³ Derek Parfit, *Reasons and Persons*, (New York: Oxford University Press, 1984), 202.

⁴ Parfit also says: "We can also ask: (3) what is in fact involved in the continued existence of each person over time?" (*Ibid.*, 202.) He does not want to answer this question because its answer would include features that are not necessary. "For example, having the same heart and the same character are not necessary to our continued existence, but they are usually part of what this existence involves" (*Ibid.*, p 202). He is only concerned with "*what this identity necessarily involves, or consists in*" (*Ibid.*, 202.) Ultimately, this is what I am interested in as well.

⁵ *Ibid.*, 205.

of the latter psychological feature, then there is a **direct psychological connection** between a person of the past and a person of the present. An example of having the same psychological feature at two points in time would be having the same belief at two points in time. If the forty-year old James holds the belief that having a solid career is important, and he held the same belief when he was twenty, he has a direct psychological connection with himself at age twenty. An example of a direct psychological connection which is formed because an earlier psychological feature accounts for a later psychological feature is between having and carrying out an intention. For example, someone can make the intention today to carry out some action tomorrow. If the person carries out the intention and performs that action tomorrow, there is a direct psychological connection between the intention today and the thoughts involved in actually carrying out that action tomorrow.⁶

“**Psychological connectedness** is the holding of particular direct psychological connections.”⁷ In Parfit’s view, psychological connectedness is the foundation of personal identity over time. However, just any level of psychological connectedness does not seem to be enough.

Suppose there are only 10 out of 1,000 possible direct psychological connections between a human being named James on his twentieth birthday and the same human being the next day. In this case, we would not be inclined to say that the human being James is the same person on his birthday as on the day after his birthday. What we

⁶ Note that for both, these psychological connections can occur *between* persons. Two different people can hold the same belief. Moreover, there can be a direct psychological connection between one person’s intention and other person’s later thoughts for carrying out that intention, e.g., if I compel my daughter to do her homework, then there is a direct psychological connection between *my intention* that my daughter completes her homework and *her thoughts* involved in doing that homework.

⁷ *Ibid.*, 206, emphasis mine.

expect is strong psychological connectedness from day to day. However, we do not expect as much connectedness between two days that are separated by a long period of time.

There might be only 10 of 1,000 direct connections between James on his twentieth birthday and James on his fortieth birthday. However, we might well hold that James at age twenty and James at age forty is one person persisting over time. There can be much change in an individual over twenty years. We do not usually conclude that only human beings who stay the same over a twenty-year period are the same persons while those who have undergone much psychological change are different persons.

To accommodate the intuition that connectedness is important from day to day but not between days separated by a longer period of time, Parfit defines strong connectedness. There is **strong connectedness**, when “there [are] enough direct connections.”⁸ What counts as just “enough” direct connections to entail strong connectedness has no precise non-arbitrary specification. As Parfit explains, “Since connectedness is a matter of degree, we cannot plausibly define precisely what counts as enough.”⁹

Parfit explains that defining strong connectedness is like the *Sorities Problem*, which asks how many grains are in a heap. In the *Sorities Problem* we suppose that we have n grains of sand, which, lying together, most people would agree is a heap. Then we take away one grain of sand at a time. When we have one grain of sand remaining, everyone would agree that this is *not* a heap. But, when did the group of remaining

⁸ *Ibid.*, 206.

⁹ *Ibid.*, 206.

grains go from a heap to not a heap? There is no clear answer to this question, though it is clear that when we began, we had a heap and when we ended, we did not. Somewhere in the middle, it is unclear what to say about these grains of sand. It is not that we lack all of the facts, but that there is a fuzzy border between being a heap and simply being several grains of sand. We may choose a cut-off point that provides us with an artificial “yes” or “no” answer, but that cut-off is arbitrary.

Given that defining strong connectedness is similar to defining a heap, Parfit chooses an arbitrary cut-off to define strong connectedness. “Since connectedness is a matter of degree, we cannot plausibly define precisely what counts as enough. But we can claim that there is enough connectedness if the number of direct connections, over any day, is at least half the number that holds, over every day, in the lives of nearly every actual person.”^{10, 11}

With a definition of strong connectedness, Parfit goes on to define **psychological continuity** and makes strong connectedness a necessary condition for personal identity over time. He calls his theory of personal identity over time the Psychological Criterion of Personal Identity:

The Psychological Criterion: (1) There is *psychological continuity* if and only if there are overlapping chains of strong connectedness. *X* today is one and the same person as *Y* at some past time if and only if (2) *X* is psychologically continuous with *Y*, (3) this continuity has the right kind of cause¹² ... Personal identity over time just consists in the holding of facts like [these.]^{13, 14}

¹⁰ *Ibid.*, 206.

¹¹ Parfit goes on to argue that, given that the cut-off is arbitrary, the resulting definition of a heap would not be particularly deep. Since what counts as strong connectedness is based on an arbitrary cut-off and is therefore not particularly deep, Parfit holds that personal identity is also not particularly deep.

¹² An example of the wrong kind of cause for experience-memory continuity *might* be someone believing that she directly recalls a real past event of her life when this “recollection” is based on someone else’s description of the past event. For instance, she may believe that she recalls almost drowning when

This definition requires that there be strong connectedness from day to day but not between one's twentieth and fortieth birthdays. Later, Parfit argues that the right kind of cause can be any cause, as opposed to only some normal cause, such as the normal functioning of the human brain. For instance, Parfit argues that personal identity can hold after a case of teletransportation in which the human being that Parfit is now is scanned, destroyed, and then duplicated on Mars. Parfit argues that this physical duplicate would be him. For Parfit, psychological continuity due to any kind of cause can satisfy the first three necessary conditions for personal identity over time.

she was a month old because her mother has recounted the story to her many times (Parfit, *Reasons and Persons*, 207).

¹³ *Ibid.*, 207.

¹⁴ The final necessary condition for personal identity over time is that the psychological continuity is a one-to-one relation. In order for p to be numerically identical with q , there can be no r distinct from q to which p is also identical. There are two types of identity, qualitative and numerical. Having the same qualities or properties is necessary for qualitative identity, whereas numerical identity holds between an object and itself, and may hold even if the qualities of the object have changed:

[T]wo white billiard balls are not numerically identical but may be qualitatively identical. If I paint one of these two balls red, it will cease to be qualitatively identical with itself as it was. But the red ball that I later see and the white ball that I painted red are numerically identical. They are one and the same ball (*Ibid.*, 201).

Normally, all cases of psychological continuity involve one-to-one numerical identity relations. Normally, if X of t_1 is psychologically continuous with Y of t_2 , there is not some z at t_2 who is also psychologically continuous with X . However, Parfit asks us to consider what it means if there were two or more people at t_2 who were psychologically continuous with X . In such a case, since there would be a one-to-many relation, this would be a case in which X is not identical to Y , though he is psychologically continuous with Y . For X cannot be identical to more than one person. Parfit argues that X not surviving is not as important as we might think. According to Parfit, it is psychological continuity that matters, not survival or personal identity *per se*.

The claim that personal identity and survival are less important than we might think is what Parfit calls the Extreme Claim. Two of the three philosophers I discuss as alternatives to Parfit, Schechtman and Korsgaard, each make very different arguments to defend the position that one can avoid the Extreme Claim by not claiming that personal identity must have the logical form of an identity relation. See Schechtman, *The Constitution of Selves*, (Ithaca, NY: Cornell University Press, 1996), Chapter 2 and Korsgaard, "Personal Identity and the Unity of Agency: A Kantian Response to Parfit," *Philosophy and Public Affairs*. Vol. 28, No. 2 (1989), entire essay.

Critique of Parfit's theory of personal identity

At first this analysis of personal identity over time might seem plausible. It requires that we are very connected from day to day, but there is no such requirement regarding the level of connectedness between days which are separated by much larger spans of time. However, Christine Korsgaard takes issue with Parfit's psychological continuity theory. Her criticism is that, in his theory, persons are too passive.¹⁵ Someone is *not* the same person simply because her "bundle"¹⁶ of psychological features changes at a relatively slow pace.

Consider the type of fantastical thought experiments, which Parfit favors, in which a person acquires new or loses old psychological features through surgical intervention. If a surgeon removed three-fourth of the previous psychological features and added different ones, then it would not be the same person who awoke after the surgery. Parfit would say that this is due to a lack of strong connectedness between the psychological features before and after surgery. To these types of thought experiments, Korsgaard responds:

It is, I think, significant that writers on personal identity often tell stories about mad surgeons who make changes in our memories or characters. These writers usually emphasize the fact that after the surgical intervention we are altered, we have changed. But surely part of what creates the sense of lost identity is that the person is changed by intervention, from outside. The stories might affect us differently if we imagined the changes initiated by the person herself, as a result of her own choice. You are not a different person just because you are very different. Authorial psychological connectedness is consistent with drastic changes, provided those changes are the result of actions by the person herself or reactions for which she is responsible.¹⁷

¹⁵ Korsgaard, "Personal Identity and the Unity of Agency," 101 – 132.

¹⁶ *Ibid.*, 101 – 2.

¹⁷ Korsgaard, "Personal Identity and the Unity of Agency," 122 – 123.

One possible way to expand on Korsgaard's point here is by more closely considering Parfit's definition of psychological continuity and strong connectedness.

I believe that Parfit runs into serious difficulties when trying to define strong connectedness using a percentage, when he says that, "we can claim that there is enough connectedness *if the number of direct connections, over any day, is at least half the number* that holds, over every day, in the lives of nearly every actual person."¹⁸ I believe that there are three possible interpretations of how Parfit defines strong connectedness. On one interpretation of Parfit's definition of strong connectedness, that definition is in danger of including too many as persisting persons *and* it is circular. On the second possible interpretation, his definition of strong connectedness is wholly arbitrary, rather than merely having an arbitrary cut-off. On the final interpretation, his definition of strong connectedness problematically uses our intuitions to define strong connectedness. On this third interpretation, strong connectedness is again in danger of including too many as persisting persons *and* strong connectedness could not be used to form a necessary condition for personal identity over time.

On one interpretation of Parfit's definition of strong connectedness, the definition of psychological continuity looks too liberal (defining too many as persisting persons) and hopelessly circular. Parfit's words, taken literally, say that strong connectedness is at least half the number of direct connections the nearly every actual person has. (These are direct connections that nearly every person has "over every day," which I take to mean from one day to the next.) If nearly every actual person has n number of connections from day to day, then why would 50% of n qualify as *strong* connectedness? Half seems

¹⁸ Parfit, *Reasons and Persons*, 206; emphasis mine.

a bad place to define *strong* connectedness. Using 50% of n , isn't strong connectedness in danger of including too many as qualifying as the same person? Wouldn't, say, 90% or greater of n ¹⁹ be better places to define *strong* connectedness?

Second, how can we pick out "every actual person" in order to find out the number of connections that nearly all of them have from day to day? Don't we need the definition of psychological continuity to pick out those who qualify as persons who persist from day to day? Yet this is telling us that before we can define strong connectedness and so before we can define the psychological criterion we already need to know who meets the psychological criterion and so who is a persisting person.

To avoid interpreting Parfit as making a circular argument, we could turn to a second interpretation of Parfit. We could understand him as saying simply that strong connectedness is when (A) at least 50% of the psychological features of a human being on Day 1 are connected to psychological features of that human being on Day 2, or (B) at least 50% of the psychological features of a human being on Day 2 are connected to psychological features of that human being on Day 1.^{20 21}

It seems to me, though, that this interpretation provides an analysis of strong connectedness that is not terribly appealing. On what grounds would 50% connectedness qualify as *strong* connectedness? Why would any percentage qualify as strong connectedness? Why not define (A) and (B) using at least 10% or at least 90%?

¹⁹ Or perhaps using some percentage formed by using a statistically significant number related to n . For example, define the percentage for strong connectedness as n minus two standard deviations of the mean all divided by n .

²⁰ Mathematically, (A) and (B) do not result in the same number of connections if Day 1 and Day 2 do not have the exact same number of total psychological features.

²¹ Chapter 4 will give another reason why looking at the connections from day to day of a human being is problematic: it assumes that there can be at most one person within one human being.

Whatever percentage we could pick to define strong connectedness looks *wholly* arbitrary. Parfit himself emphasizes how our cut-off for defining strong connectedness is arbitrary to some degree, e.g., if we choose having at least 50% connections from day to day to define strong connectedness, why isn't at least 49% enough? However, my point is that, on this second interpretation, strong connectedness cannot plausibly be defined using *any* percentage because we have no way to define what would count as a plausible, though perhaps arbitrary, cut-off.

This second interpretation is the least likely to be the correct interpretation because Parfit focuses on defining the percentage using "every actual person". Suppose we do not interpret him as being circular as we did under the first possible interpretation. Again, Parfit says "we can claim that there is enough connectedness if the number of direct connections, over any day, is at least half the number that holds, over every day, in the lives of nearly every actual person." By "every actual person," we can understand Parfit as referring to those who match our intuitive sense of those who are persisting persons from day-to-day. So take all of those human beings who, intuitively, we think persisted as the same person from one day to the next, and see how many connections, m , they have from day to day. Connectedness from day to day counts as strong if there is at least 50% of m .

The first problem with this way of defining persons is, like with the first interpretation, that it would appear to define personal identity over time too liberally. If we find that those who we already think of as the same persons over time have m connections from day to day, why would half of an average of m plausibly qualify as

strong connectedness? Something like at least 90% of m seems more appropriate for *strong* connectedness.

More importantly, on this third interpretation, Parfit does not define strong connectedness such that it could be used to build a necessary condition for the persistence of persons. On this third interpretation, Parfit appears to be using reflective equilibrium when defining strong connectedness. His definition of strong connectedness and psychological continuity is formed by looking at “every actual person” we encounter in our lives, those whom we consider persisting persons pre-theoretically.²² I, too, will use the methodology of reflective equilibrium, so I do not take issue with reflective equilibrium *per se*. However, this is an inappropriate use of that methodology.

The methodology of reflective equilibrium is used to check parts of a theory or an entire theory. Suppose we have a theory which claims to provide the necessary and sufficient conditions to be an F; call it The Theory of F. Using the methodology of reflective equilibrium, we compare the implications of the Theory of F against our intuitions of what qualifies as an F to see if the theory is neither too liberal nor too conservative. It is too liberal if what we intuitively think of as clear cases of not Fs, under the theory, problematically qualify as Fs. It is too conservative if what we intuitively think of as clear cases of Fs, under that theory, problematically do *not* qualify as Fs. If The Theory of F so poorly matches our intuitions of what qualifies as Fs, there is a severe reflective *disequilibrium* and perhaps the proponents of The Theory of F need to re-think some or all of that theory.

²² Interestingly, this means that we would not define strong connectedness using fantastical thought experiment persons.

In contrast to using the methodology of reflective equilibrium to test his theory, on the third interpretation, Parfit would be using our pre-theoretical intuitions of who counts as persisting persons in his theory of persons, rather than using those intuitions to test the final theory. This would be similar to The Theory of Fs, which is supposed to provide the necessary and sufficient conditions for being an F, answering that one necessary condition of being an F is that it includes at least those who we already say are Fs. Parfit is saying that one necessary condition of *X* being the same person as *Y* is that, between the two there, is the same degree of connectedness as found in those we intuitively think of as the same persons.²³ Parfit does importantly emphasize the *psychological* in strong psychological connectedness, but *strong* is defined in a way that seems a non-starter. What are missing are the necessary and sufficient conditions that explain *why* it is the case that, if we were to look at those who intuitively are the same person, that *m* connections on average occur from day to day. Such an analysis would enable Parfit to define strong connectedness without problematically using our intuitions in the theory itself.

How could Parfit define strong connectedness without being circular, completely arbitrary, or problematically using our intuitions in the definition itself? His theory cannot stand without an adequate definition of strong connectedness: psychological continuity is defined by strong connectedness. I think these worries about how to define strong connectedness reflect the fact that Parfit's analysis of strong connectedness and so his final psychological criterion fails to emphasize what is important about persons. Parfit's analysis is promising because it focuses on psychology but fails to take seriously

²³ Well he is saying that half of that degree of connectedness defines strong connectedness, which again, seems too liberal.

the fact that some kinds of psychological features can come and go and it may be completely irrelevant to whether a person persisted, while other psychological features only exist as a result of agency and, if a person is to persist, those features can change only through that person's agency. So the criteria of personal identity should not define strong connectedness by a percentage of connections from day-to-day, but by which changes in psychological features properly occur through our agency. If we were to find that, on average, those we already think of as persons pre-theoretically have *m* connections from day to day, the analysis of personal identity should explain *why m* is the average much of the time, but sometimes the number of connections is much lower and it is still the same person. If someone were to realize that his world and those within it are radically different than he had assumed, most of his psychological features might change in one day – but through his agency and with good reason of course.

This is Korsgaard's point. Persons are most importantly agents who change themselves, change parts of their own psychologies. I contend that *the way* some of a person's psychological features change is more important than simply how much or how quickly her psychological features change. Even if some form of psychological continuity is a necessary condition for personal identity, this criticism rejects Parfit's first and third requirement that psychological continuity be defined using a problematic percentage analysis of strong connectedness and that strong connectedness can have any kind of cause. Instead, agency must be central in any satisfactory analysis of personal identity: for a person to persist, it is necessary that some, most or all changes of some kinds of psychological features are the product of the person's agency.

Since Korsgaard's and other criticisms of Parfit's view of personal identity have been published, a few philosophers have presented their own alternative psychological theories of personal identity which emphasize the agency of persons. In the rest of this chapter, I will discuss three alternative theories which focus on agency, those of Marya Schechtman, Christine Korsgaard, and Carol Rovane. The purpose of this discussion is to show that while Schechtman, Korsgaard, and Rovane move the discourse of personal identity in the right direction by recognizing the importance of agency, they, too, ultimately fail to provide a satisfactory theory of personal identity over time.

Marya Schechtman: Personal Identity and Narrative

Schechtman's theory of personal identity

In *The Constitution of Selves*, Schechtman provides another psychological theory of personal identity. She calls her view the narrative self-constitution view:

The narrative self-constitution view starts from the relatively uncontroversial assumption that not all sentient creatures are persons. Personhood – at least insofar as it acts as the underpinning for moral agency, compensation, self-interested concern, and, I contend, survival – involves, more than rudimentary consciousness. ...According to the narrative self-constitution view, the difference between persons and other individuals (I use the word "individual" to refer to any sentient creature) lies in how they organize their experience, and hence their lives. At the core of this view is the assertion that individuals constitute themselves as persons by coming to think of themselves as persisting subjects who have had experience in the past and will continue to have experience in the future, taking certain experiences as theirs. Some, but not all, individuals weave stories of their lives, and it is their doing so which makes them persons. On this view a person's *identity* ... is constituted by the content of her self-narrative, and the traits, actions, and experiences included in it are, by virtue of that inclusion, hers.²⁴

Later she writes:

Most broadly put, this means that constituting an identity requires that an individual conceive of his life as having the form and the logic of a story – more specifically, the

²⁴ Schechtman, *Constitution of Selves*, 94.

story of a person's life – where story is understood as a conventional, linear narrative.²⁵

26

To the extent that one does have a self-narrative, one's self-narrative does not self-constitute if it does not line up with reality. For Schechtman, a deranged person who thinks himself Napoleon fails to have an identity-constituting narrative.

On Schechtman's view, a person does not necessarily create one ongoing narrative; as the person's life progresses, his narrative can change. At age twenty, how one might discuss one's life – what one emphasizes, overlooks, and ignores in creating a self-narrative – would be different than how one self-narrates at age eighty. At twenty, Tim might see a certain experience as the worst thing that could ever happen to him, but at eighty he might see great value in that experience, or, alternatively, at eighty, Tim might fail to put that experience in his self-narrative at all.

Schechtman recognizes the fact that few are constantly creating a self-narrative. Instead, her narrative self-constitution view requires that people be able to articulate a narrative locally, or be able to “explain why he does what he does, believes what he believes, and feels what he feels.”²⁷

²⁵ *Ibid.*, 96.

²⁶ Initially, this is how Schechtman describes her view, but she goes on to add more dimensions to her view to avoid excluding individuals from other cultures.

Recognizing dominant feminist and other critiques that suggest that the linear form of narrative is ‘masculine’ or particular to ‘dominate western culture,’ Schechtman allows some degree of non-linearity in the forms of narrative which can constitute a person's identity. What she wants to exclude is a self-conception which is markedly lacking in any degree of linearity, such as the “self conception of an extreme psychotic”. Besides the issue of linearity, she also notes that very different cultures might have persons, but their conception of persons is extremely different from ours. For example, in some cultures, one's conception of one's identity might extend back through ancestral lines, and thereby produce alternative contours for responsibility, self-concern, survival, and compensation. In such cultures, one may be viewed as and think of one's self as more responsible for the wrongs of ancestors than in a Western culture such as our own. See *Ibid.*, 103.

²⁷ *Ibid.*, 114.

Schechtman allows for the fact that *sometimes* individuals may be at a loss for a self-explanation or be unable to explain themselves accurately. She holds that what an individual needs in order to be a person is an implicit self-narrative. This implicit self-narrative is at play when a person deceives herself about certain parts of herself or the motivations for some of her actions.

The implicit self-narrative is understood as the psychological organization from which his experience and actions are actually flowing. ...

I call a person's underlying psychological organization a *self*-narrative because it is not simply a static set of facts about him, but rather a dynamic set of organizing principles, a basic orientation through which, with or without conscious awareness, an individual understands himself and his world. These implicit organizing principles are not simply a collection of features, but a continually developing interpretation of the course of one's trajectory through the world. In this way it is legitimate to think of what I am calling the implicit self-narrative as a *self*-conception, even though it contains elements that the person explicitly denies.²⁸

Later, Schechtman goes on to add:

Granted, there is a deeply significant difference between those unstated elements which are easily available to the subject and those which are not – a difference that has serious implications for personal identity. Still, I wish to maintain that insofar as both kinds of elements shape the way a person approaches the world they are both parts of her self-narrative.²⁹

Critique of Schechtman's theory of personal identity

I hold that Schechtman's theory of personal identity faces a two-horned dilemma. On the first horn is that Schechtman seems to provide a view that is too exclusive: it excludes too many normal adult human beings from having personal identities. Sometimes she suggests that in order to be a person one must have a relatively explicit narrative. If her theory requires that in order to be a person one must engage in some degree of self-narrative creation, it might be too exclusive. On the second horn of the dilemma, her theory would avoid the worry of exclusivity and be fairly inclusive.

²⁸ *Ibid.*, 115-6.

²⁹ *Ibid.*, 116-7.

However, it would do so at the costs of not being a self-narrative-based theory of personal identity and of being problematically underdeveloped. Consider the first horn of the dilemma that her view may be too exclusive.

In “Against Narrativity,” Galen Strawson argues that many philosophers including Schechtman hold both a ‘*psychological Narrativity thesis*’ and a ‘*normative Narrativity thesis*’, and both these are wrong.³⁰ The psychological Narrativity thesis holds that normal human beings create their identities through creating narratives of their lives. The normative Narrativity thesis holds that we need to create narratives of our lives in order to have good lives, or be good persons. As evidence against the psychological Narrativity thesis, Strawson says that he does not create a narrative that includes his whole life and that there are others whose writings clearly indicate that they also live lives without a narrative. Strawson holds that he can give a story about his further past, but that is the story of Strawson the human being and not a story about Strawson himself. He has no narrative which connects Strawson the self today with Strawson the self who lived in the further past. If we believe that Strawson is correct and honest about his own case³¹ and assume that how he lives his life is not unique to him, then these cases of individuals who lack narrativity provide counter-examples to and thereby disprove the psychological Narrativity thesis.³²

In arguing against the normative Narrativity thesis Strawson suggests that, for some individuals living a narrative life may be necessary for living a good life, but for

³⁰ Galen Strawson, “Against Narrativity,” *Ratio*. Vol. 17, No. 4 (December 2004):428-29.

³¹ Strawson addresses the worry that he may be mistaken about the lack of narrativity in the creation of his identity. For a brief discussion on this point, see Footnote 36.

³² *Ibid.*, 433.

others it can be detrimental.³³ He holds that this is simply a difference between human beings: there are some human beings who need self-narratives and others who are better off without self-narratives. He argues that those who do not lead narrative lives can live good lives. Strawson holds that he is not interested in questions such as, “What has [he] made of his life?” nor does he need to form a narrative in order to properly structure his life.

[Instead] I'm living [my life] and this sort of thinking about it is no part of it. This does not mean that I am in any way irresponsible. It is just that what I care about, in so far as I care about myself and my life, is how I am now. The way I am now is profoundly shaped by my past, but it is only the present shaping consequences of the past that matter, not the past as such.³⁴

Moreover, he holds that those who do not lead narrative lives can still have good lives which include acting responsibly and forming strong friendships. He uses himself as an example but also cites Michel de Montaigne, someone who seemed to lack a narrative, and Montaigne's friend, Etienne de la Boétie, who viewed Montaigne as a great friend.³⁵

36

³³ *Ibid.*, 447.

³⁴ *Ibid.* 438.

³⁵ *Ibid.* 438, 449-450.

³⁶ Strawson addresses the worry that he may be mistaken about the lack of narrativity in the creation of his identity. Someone might claim that Strawson, despite his statements to the contrary, engages in self-narrative creation, and so is a person under Schechtman's theory of personal identity. He tries to allay this skepticism by pointing out that it might be difficult for someone to understand or appreciate how someone else could live without narrative if that someone does create or needs to create narrative to structure her life (*Ibid.*, 437). I would contend that Strawson's own self-reports seem more persuasive as evidence against the psychological and normative Narrativity theses than the position that he must be engaging in self-narrative creation from someone who upholds the psychological and normative Narrativity theses but neither experiences the inner life of, nor a personal relationship with, Strawson. As summarized in some of Strawson statements regarding himself, Strawson tries to bring home how he does not engage in , and, in order to have a good life or be a good person, need not engage in self-narrative creation. However, if someone continues to be skeptical and think that Strawson actually does engage in self-narrative creation, there is no way to argue against this skeptical position. All Strawson can do is insist that, at the very least, he is correct – and honest – about his own self-assessment.

I share Strawson's concerns with Schechtman's focus on narrative and think that creating a self-narrative may not be unessential to having an identity. I have no doubt that creating a self-narrative can help form the identities of some persons. For example, a person suffering from an identity crisis could create a self-narrative, and that self-narrative might help her decide what type of person she wants to be or clarify what type of person she has been. Moreover, I recognize that the abilities necessary for creating a self-narrative – such as the ability to reason in order to be able to explain one's self – are abilities necessary for personhood. However, it is unclear that “the creating of a self-narrative” *per se* is necessary for having a personal identity, as Strawson and others he discusses seem to be clear counter-examples.

Can Schechtman take some middle road which would enable her to avoid my dilemma that her view is either too exclusive or fairly inclusive but problematically underdeveloped? I hold there is no room for Schechtman to take a middle road. Strawson asserts that he lacks a narrative and he does not use one to explicitly or implicitly shape his life, and I take him at his word. Though he does not create a narrative to shape his own life, he seems a person to me.³⁷ If any degree of narrativity is required on Schechtman's theory, this excludes Episodics as well as both non-narrating and only intermittently narrating Diachronics and likely others³⁸ and this puts Schechtman's view on the first horn of my dilemma.

³⁷ I have not met Galen Strawson but the fact that he wrote various philosophical manuscripts is evidence enough for me that he is a person. A non-person did not write these manuscripts.

³⁸ Besides Episodics, Strawson suggests that there could be non-narrating Diachronics. I would add that narrating Diachronics do not narrate *all* of the time. Who are these individuals when they are making decisions but are not narrating at the time? (Schechtman anticipates my concern about Diachronics when they are not narrating but her response leads to the second horn of the dilemma for Schechtman.)

Of course the concerns that Strawson and I have with Narrativity might be misdirected towards Schechtman. In other parts of her text, she suggests that creating a narrative is not necessary to being a person with an identity.³⁹ However, this leads to the second horn of the dilemma that I argue Schechtman faces: that rather than being too exclusive, her theory is inclusive but problematically underdeveloped.

While Schechtman talks about narrative as a life story, she also makes several points that move away from narrativity. She notes that a person can fail to be able to fully articulate a narrative, rarely narrate, and have implicit parts of her narrative that are inaccessible consciously.⁴⁰ In fact, as I noted earlier, at one point she says in order to satisfy her narrative self-constitution view, a person simply needs to be able to “explain why he does what he does, believes what he believes, and feels what he feels.”⁴¹ With all this said, it becomes very difficult to understand how this is a *narrative* view of personal identity. It becomes unclear exactly what type of view it is. Rather than the creation of a ‘self-constituting narrative’ *per se*, she simply seems to be talking about the entire psychological life of a person, which includes using several important abilities and actively engaging in life. This suggests that, on her view, the psychological lives of persons can lack creation of a self-narrative.

Finally, I would add that besides Episodics and non-narrating or intermittently narrating Diachronics, Schechtman’s view excludes normal healthy human beings as possibly persons if they are a part of a culture that has no sense of time. I want to argue that such individuals could exist over time as persons, even though they certainly do not view their own lives in this way.

³⁹ Strawson might say that then Schechtman’s view would be a trivial version of the psychological Narrativity thesis. See Strawson, “Against Narrativity”, 438-9.

⁴⁰ Schechtman, *Constitution of Selves*, 115-117.

⁴¹ *Ibid.*, 114.

Schechtman seems to recognize the potential worry that she is suggesting that her so-called narrative self-constitution view of personal identity does not focus on narrative. When she argues that explicit and implicit aspects of one's narrative are both part of the narrative, even though, by definition, a person is unable to narrate about the implicit parts of her narrative, she writes the words below. I have highlighted in bold what seems problematic to me:

I realize that this is a somewhat unusual way to think about "self-narrative," and that I may not have satisfied everyone that it is a legitimate one. In the end, however, very little really turns on this choice of words. **It does not matter much whether we say that identity is determined by a person's self-narrative or by his psychical organization,** so long as it is understood that the psychological forces constituting identity are dynamic and active – things a person *does* – rather than static and passive features she *has*. **I use the term "self-narrative," even though it is somewhat controversial here, to underscore these features of a person's psychological life.**⁴²

If this were the beginning of a shift in her book to focusing on something other than narrative, then her point about semantics is well taken. However, later, in defending her theory, narrative remains central to her analysis of personal identity. For example:

Both the subjectivity and the capacities definitive of living the life of a person are, as I have already shown, tied directly to the possession of an identity-constituting narrative of the form defined in the narrative self-constitution view. Failure to create such a narrative entails the failure to become a person, and the loss of a narrative by an individual who has formed one constitutes that person's demise. This conclusion does not, moreover, rely on a mere technicality of definition. To see this we need only recognize examples of humans who fail to have such narratives.⁴³

She then goes on to talk about infants and those suffering from the later stages of dementia as examples of individuals who lack such narratives. While infants and those suffering from late-stage dementia lack such narratives and, I will assume for the sake of argument, both lack personal identities, it seems false that all who lack such narratives,

⁴² *Ibid.*, 117, bold emphasis mine.

⁴³ *Ibid.*, 150.

including Strawson, also lack a personal identity.⁴⁴ Moreover, unlike Strawson, infants and those suffering from dementia also lack other psychological capacities which are arguably necessary to have personal identities, which makes it unclear whether the lack of narrativity is what explains why infants and those suffering from late-stage dementia lack personal identities, rather than the absence of, say, the capacity to reason. If, on the other hand, Strawson can be included under her view, and so having a self-narrative is not necessary for constituting one's self, then exactly what *is* necessary to being a person and having an identity? What is a person's "psychical organization" or what are "the psychological forces constituting identity [which] are dynamic and active"? Without an analysis of this, this theory is not only surprisingly *not* about self-narrative or the creation thereof, it is problematically underdeveloped.

Rather than narrativity, both Christine Korsgaard and Carol Rovane offer analyses of persons and personal identity each which focuses on the capacity to reason, the capacity to engage in rational thought. While Korsgaard never fully develops her analysis of persons and personal identity over time – when she returns to the question of the identity of persons, she is focused on addressing a slightly different question – I will argue that Rovane's analysis, a possible development of Korsgaard's early thoughts, is problematic. The problem is that rationality is inadequate to unify the identities of everyday persons, persons who are less than fully rational.

⁴⁴ Schechtman might be right to claim that those suffering from the late stages of dementia lack personal identities. However, since I disagree that narrativity is necessary for personal identity, my reasons for concluding those in the last stages of dementia lack personal identities would differ from her reasons for that conclusion.

Christine Korsgaard's Work on Identity

In "Personal Identity and the Unity of Agency: A Kantian Response to Parfit,"⁴⁵

⁴⁶ Christine Korsgaard argues that Parfit's view is problematic because it fails to take seriously that persons are agents and not merely the subjects of experiences. Korsgaard says that the proper analysis of personal identity is one which avoids treating persons merely as subjects of experiences. Providing such an analysis gives a practical, rather than a metaphysical, analysis of persons.

It is important to see what Korsgaard means by distinguishing the metaphysical from the practical. Parfit provides a metaphysical analysis of personal identity over time. Parfit's analysis seeks to explain what makes a person one and the same over time, how one is numerically identical to oneself over time. Numerical identity is a logical relation, which, among other things, is an all-or-nothing relation (A either *is* or *is not* identical to B, A cannot be more or less identical to B), and is transitive (If A is identical to B, and B is identical to C, then A is necessarily identical to C.). This analysis emphasizes how persons are like any other object, (though we believe that persons, as subjects of experiences, are special objects that are of moral concern). While we can understand persons as any other objects that are the subjects of experiences, Korsgaard's point is that we can also understand persons from a practical standpoint, as agents who must decide what to do.

⁴⁵ Korsgaard, Christine M. "Personal Identity and the Unity of Agency: A Kantian Response to Parfit." *Philosophy and Public Affairs*. Vol. 28, No. 2 (1989): 101 – 132.

⁴⁶ For example of a later work with a similar criticism, see Korsgaard, Christine M. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press, 2009.

Korsgaard provides a practical analysis of personal identity.⁴⁷ She asks how an individual is a person and not merely a Humean bundle, a bundle of “loves, interests, ambitions, virtues, vices, and plans.”⁴⁸ Her answer is that this bundle is that of a unified person because, through her agency, that person decides what to do. In having to act, there is the practical necessity of making a decision. When faced with a conflict of desires, for example, one must come to a resolution between one’s conflicting desires in order to decide what to do next. Moreover, and most importantly, when deciding what to do, one sees one’s self as something over and above one’s desires. One is “not simply waiting to see which desire wins”⁴⁹ out. In cases of conflict, instead, one has reasons for choosing one way or another which are more than simply desires.

And it is these reasons, rather than the desires themselves, which are expressive of your will. The strength of a desire may be counted *by you* as a reason for acting on it; but this is different from its simply winning. This means that there is some principle or way of choosing that you regard as expressive of *yourself*, and that provides reasons that regulate your choices among your desires. To identify with such a principle or way of choosing is to be “a law to yourself,” and to be unified as such. This does not require that your agency be located in a separately existing entity or involve a deep metaphysical fact. Instead, it is a practical necessity imposed upon you by the nature of the deliberative standpoint.⁵⁰

Moreover, Korsgaard claims, there is a practical necessity of viewing one’s self as extending over time:

Some of the things we do are intelligible only in the context of projects that extend over long periods. This is especially true of the pursuit of our ultimate ends. In choosing our careers, and pursuing our friendships and family lives, we both presuppose and construct

⁴⁷ Schechtman also finds a metaphysical analysis of persons problematic, though on different but related grounds, and provides a normative analysis of persons as well. In contrast, Rovane provides a metaphysical analysis, but one that is grounded in normative claims about the nature of persons. Given this focus on a normative analysis of persons, Rovane’s project could be understood as a practical *and* metaphysical project. Hence, Rovane, Schechtman, and Korsgaard all provide practical analyses of personal identity while Parfit’s analysis is purely metaphysical.

⁴⁸ Korsgaard, “Personal Identity and the Unity of Agency: A Kantian Response to Parfit”, 109.

⁴⁹ *Ibid.*, 111.

⁵⁰ *Ibid.*, 111.

a continuity of identity and of agency. On a more mundane level, the habitual actions we perform for the sake of our health presuppose ongoing identity. It is also true that we think of our activities and pursuits as interconnected in various ways; we think that we are carrying out plans of life. In order to carry out a rational plan of life, you need to be one continuing person. You normally think you lead one continuing life because you are one person, but according to this argument the truth is the reverse. You are one continuing person because you have one life to lead.^{51 52}

I believe that Korsgaard's comments here are interesting, but are only preliminary. She does not develop these ideas into a full analysis of personal identity at or over time. To be more precise, in later work, Korsgaard does develop an analysis of personal identity but the focus of that analysis has shifted in an important way.

In *Self-Constitution* and other works⁵³, Korsgaard no longer focuses on providing an analysis of the unity of the identities of persons. Consider her most recent work on personal identity, *Self-Constitution*. In this manuscript, Korsgaard argues that persons constitute themselves – unify themselves – by acting on formal rational principles, such as the Categorical Imperative, choosing to act on maxims that can be willed as universal law. Korsgaard certainly knows that almost no human being *always* acts on the Categorical Imperative, but she is not claiming that almost no human being is a unified person at or over time. Instead, she argues that when persons fail to act on the Categorical Imperative, those persons are unified *badly*. She explains how this is similar to how biological organisms can constitute themselves as such but do so badly.

Take a living giraffe. As a biological organism it constitutes itself continuously as a giraffe, maintaining its biological functions that enable it to continue to live.

⁵¹ *Ibid.*, 113.

⁵² Episodics do not view themselves as needing to lead one life. However, I believe that these points can be modified to accommodate Episodics, who clearly take on projects that they know are long-term projects.

⁵³ E.g., Korsgaard, *Sources of Normativity* (New York, NY: Cambridge University Press, 1996).

Already constituted – already alive – it can continue to constitute itself badly, e.g., for a couple of weeks by not eating enough or drinking enough despite available resources. The giraffe still exists due to properly continuing most of its self-constituting functions during those two weeks, but it was continuing to do its self-constituting badly, e.g., if the giraffe had encountered a pride of lions, those lions would have easily ended its existence because the giraffe was weak due to its failure to eat and drink adequately.⁵⁴

In a similar way, Korsgaard claims that persons fail to unify themselves *well*, when they fail to act on formal rational principles. Is this true? *Do* persons fail to unify themselves well when they fail to act on formal rational principles? I am skeptical but will not try to defend my skepticism here. What is important to note here is that personal identity theorists such as Parfit, Schechtman, and Rovane are not addressing the question of how persons can unify themselves *well*. Instead, most personal identity theorists want to know how persons are unified *at all*, though perhaps imperfectly, at or over time. So while Korsgaard's critique of Parfit and her preliminary thoughts about how persons unify themselves at and over time seem promising, subsequently she takes her work in another direction – albeit an extremely interesting and engaging one. Korsgaard never explains, given how most persons sometimes fail to engage in what she argues is good self-unification, how persons are unified, even when that unification is not done terribly well.

⁵⁴ This analogy might be especially apt: just as the giraffe will cease to exist if it fails to constitute itself well enough for long enough, a person might cease to exist if a so-called person *never* acts on universalizable maxims. This seems true if 'universalizable maxim' simply means that one acts on reasons that one thinks are reasons upon which to act in the very same circumstances: how is something being taken as a reason if one thinks it would not count as a reason for one in the *exact* same circumstances? How is that human being a person at all? The controversy, though, will come in with the idea that universalizability is not simply about *one* being in the exact same circumstances but *anyone* being in relevantly similar circumstances.

Carol Rovane's project seems to me to be one way to develop Korsgaard's preliminary comments into a full analysis of the identities of persons. Rovane develops her project independently and quite differently than Korsgaard⁵⁵ but, just as Korsgaard's preliminary comments on unification focus on rationality and the capacity to reason,

Carol Rovane: Personal Identity and a Rational Point of View

Rovane's theory of personal identity over time

In *The Bounds of Agency: An Essay in Revisionary Metaphysics*, Carol Rovane provides what she calls an ethical analysis of personal identity. She argues that our intuitions support both a physical and a psychological perspective on personal identity and holds that we cannot resolve the debate between physical and psychological personal identity theorists by showing that one type of theory is incoherent. Instead, we should look for cogent reasons to embrace either a physical or a psychological theory of personal identity. Rovane notes that by choosing one type of theory, the resulting theory will be revisionary. It is revisionary in that it is a revision of our common-sense intuitions regarding personal identity, for we have intuitions that favor both physical and psychological theories of personal identity. Rovane is revisionary and argues for a psychological theory of personal identity.

Rovane presents and defends a normative definition of 'person.' She defines persons as having properties that, for independent reasons, we hold are of moral

⁵⁵ Rovane herself clearly describes the important similarities and differences between her analysis in *The Bounds of Agency* and Korsgaard's discussion in her "Personal Identity and Unity of a Person". See Rovane, "Carol Rovane, *The Bounds of Agency: An Essay in Revisionary Metaphysics*, (Princeton: Princeton University Press, 1998), 161 footnote.

relevance.⁵⁶ Next, she argues for what she calls a metaphysical analysis of personal identity.⁵⁷ Like Parfit, she provides a necessary and sufficient condition for personal

⁵⁶ In this footnote I present some details about Rovane's understanding of personhood. While I make a few critical points, the goal of this footnote is neither to fully defend nor to fully argue against Rovane's conception of personhood. Rather, since she ties it to her understanding of personal identity, the goal is simply to introduce Rovane's view of personhood in case the reader wants some details about the view.

Rovane begins with an assumption about persons: persons are agents who "can engage in agent-regarding relations" (*Ibid.*, 72.) Agent-regarding relations are relations in which "one attempts to influence another and yet aim not to hinder the other's agency." (*Ibid.*, 72). In order for one's relation with another to be agent-regarding, generally one must engage in pure, open, and rational discourse, appealing to the other's rational point of view. Someone is a person if she has the ability to engage in agent-regarding relations. (Rovane notes that there are certainly exceptions to this desire that others engage in pure, open, rational discourse in particular cases, say, when another would simply push you out of the way of a moving car. In doing this, there is no direct appeal to your rational point of view, and instead the other person uses brute force. However, pushing you out of the way is an act that is acceptable in the circumstances from your rational point of view. Hence, in such cases, forcefully pushing you out of the way is likely an expression of full regard for you as an agent.)

In order to be able to engage in agent-regarding relations, individuals need additional abilities. First, they must be capable of social knowledge, which enables them to recognize whether others can engage in agent-regarding relations. Second, if they achieve social knowledge, they face an unavoidable choice concerning whether to engage in agent-regarding relations with others. Third, they know that "it matters to agents whether persons engage in agent-regarding relations with them, and furthermore, [by definition], persons recognize that this matters to other agents" (*Ibid.*, 105).

Rovane argues that her assumption about the nature of persons is acceptable if it meets three constraints:

[It] (1) does not beg the question in this dispute [about whether persons should be understood as a human animal or something with certain psychological abilities] (2) is sufficiently important that any account of the kind 'person' ought to accommodate it, and (3) is completely uncontroversial from an ethical point of view [i.e., no substantive ethical view would find this analysis of persons controversial] (*Ibid.*, 72).

Rovane argues that her assumption meets the first and second parts of the criteria. Other psychological identity theorists could accept her conception of persons. More importantly, even animalists, those who argue that our identity consists in being the same animal over time, could accept it as stating something important about persons and "then go on to argue that the only kind of thing that can possess [the] capacities [necessary to be a person] is a *suitably endowed animal*" (*Ibid.*, 101).

Rovane argues that her conception of persons is important and one that should not be left out of any substantive view about the nature of persons. "The importance of the ethical criterion of personhood lies in the fact that it helps to define and expose a particular form of prejudice against persons, which takes the form of a hypocritical denial of their personhood" (*Ibid.*, 123). An example would be when others deny the personhood of a racial or ethnic group in order to use force or manipulation and to avoid a need to address ethical questions, such as whether such treatment is acceptable towards persons. Rovane notes in this part of the discussion that even utilitarians must recognize that the choice concerning whether to engage in agent-regarding relations is an ethically important choice. This is because in order to do what is

identity. She argues that this necessary and sufficient condition works for both identity at a time (synchronic identity) and identity over time (diachronic identity).

Rovane's thoughts on personal identity begin with Locke. In Rovane's analysis of Locke, Locke holds that identity consists in the "phenomenological unity of consciousness,"⁵⁸ that one is the same person over time by being rational, reflective, and by having a self-aware, unified, first-person point of view. Rovane agrees with Locke's

conducive to overall happiness, they must recognize that the happiness of persons is tied to recognizing persons as such and showing full regard for their agency. In other words, a person desires that others who wish to influence him do so without hindering his agency, and, instead, appeal to his rational point of view rather than doing things such as manipulate, coerce, or brainwash him.

Rovane argues that this conception of persons is neutral between different substantive ethical views and does not favor Kantianism or rights-based theories over utilitarianism. I have some reservations about whether her view is neutral but I will not discuss those reservations here. Here, I will assume for the sake of argument that her analysis of 'person' is neutral.

However, even if her view of persons is neutral between substantive ethical views, one might wonder whether her view of persons is simply too focused on the nature of persons as ethical beings. While it does seem essential to persons, as opposed to simple human beings, that they have the ability to engage in agent-regarding relations, that ability does not seem to sum up what we essentially are. Morality is a part of who we are but not the sum of who we are, as Susan Wolf suggests in "Moral Saints". While the focus of Rovane's arguments might overlook this, her view can accommodate this point. The ability to be able to decide to engage in agent-regarding relations does presuppose other abilities which are not strictly speaking moral, such as the ability to engage in rational reflection. These abilities would seem to be necessary to have even an amoral life. Moreover, once Rovane moves into her discussion of personal identity, the ability to decide whether to engage in agent-regarding relations is not her focus, rather it is rationality and choosing projects.

I will end my discussion of Rovane's analysis of personhood here. Again, the point of this footnote was neither to fully defend nor fully argue against Rovane's conception of personhood. Rather, since she ties her conception of personhood to her understanding of personal identity, the goal mainly was to introduce Rovane's view of personhood in case the reader wanted some details about the view.

⁵⁷ By providing a normative analysis of persons, Rovane's project seems more practical than metaphysical. It starts by assuming that persons have the qualities that, when possessed by a being, we hold that ought to take that being into practical consideration. It is metaphysical in attempting to provide an analysis of personal identity such that personal identity has the logical form of an identity relation. (Rovane, *The Bounds of Agency*, 53-59) However, her entire manuscript focuses on the ethical analysis of persons and explains how rationality can provide unification. To provide an analysis of personal identity that satisfies the logical form of an identity relation, she adopts a metaphysical analysis – four-dimensionalism – that, actually, could be wed to metaphysical analyses of what we are which compete with Rovane's own analysis. For a description of four-dimensionalism, as well as its strengths and weaknesses, see Eric T. Olson, *What are We? A Study in Personal Ontology*. (Oxford, UK: Oxford University Press, 2007), Chapter 5.

⁵⁸ Carol Rovane, *The Bounds of Agency: An Essay in Revisionary Metaphysics*, 19.

emphasis on unity and rationality, but argues that this unity need not be phenomenological. Instead, she argues that personal identity stems from having a rational point of view.

A rational point of view is the point of view from which a person deliberates. The proper goal of deliberation is to arrive at, and also to act upon, all-things-considered judgments about what it would be best to think and do in the light of everything in the deliberator's rational point of view.⁵⁹

Personal identity stems from having a rational point of view. In Rovane's account, one's rational point of view comes from having unifying projects. The rational point of view that comes from unifying projects presupposes a commitment to overall rational unity. Personal identity both at a specific time and over time consists in having this commitment to overall rational unity.

By a 'unifying project', Rovane does not mean a project to unify one's life. Rather, she means to emphasize that when someone has multiple projects, each creates a single rational point of view from which to carry out that project. However, one also takes on a project to coordinate between those particular projects. This coordination project is one's unifying project, which she argues, presupposes a commitment to overall rational unity.⁶⁰ As an illustration of her view, consider this example.

Suppose the human being Nancy takes on the project of being a good professor and the project of being a good mother. (Let's make Nancy a simple example of a human being who has no other fundamental projects. So while she might have the project of being healthy, she has that project for the sake of being a good mother or for the sake of being a good professor.) Consequently, for each project, Nancy forms or continues to

⁵⁹ *Ibid.*, 21.

⁶⁰ *Ibid.* 164-5.

form ideas about what values one should develop and what actions one should perform in order to meet these goals. From each of these projects, Nancy could form two distinct rational points of view, each point of view which focused on one of these two projects. However, suppose further that Nancy coordinates these two projects, e.g., going to see her son's Wednesday night recital instead of working late Wednesdays as usual, staying at the office on weekday afternoons to grade papers or write scholarship, rather than spending afternoons with her son immediately after he leaves school. In coordinating these two projects, Nancy has formed a unifying project between these two projects. Rovane argues that carrying out this unifying project involves creating a single rational point of view, namely a single rational point of view about being *both* a good professor *and* a good mother. In taking on the unifying project, in addition to creating a rational point of view, this presupposes a commitment to overall rational unity. To return to the example, in taking on the unifying project to be both a good professor and a good mother, there is a commitment to trying to have a non-conflicting point of view about how to be both a good professor and mother *at the same time*.

If achieved, overall rational unity would involve making all-things-considered judgments which take all one's projects and their finer details into account *and* acting on these all-things-considered judgments. Rovane makes it clear that, on her analysis of personal identity, in order to be a unified person, one does not have to succeed in achieving overall rational unity, but that, if unified, one must be committed to overall rational unity, which includes striving for that unity, though sometimes unsuccessfully.

Nancy is a human-sized person if she strives to achieve overall rational unity between the projects of good professor and good mother. Successfully achieving overall

rational unity between these two projects would involve making and acting upon all-things-considered judgments (judgments which consider those two projects). Rovane says that Nancy qualifies as striving for overall rational unity between these two projects if Nancy does two things. First Nancy must see that overall rational unity between her two projects would summarize “what it would be to be fully or ideally rational.”⁶¹ Second, Nancy sees her failure to achieve overall rational unity between these two projects as grounds for self-criticism and self-improvement.⁶²

Rovane’s theory leads to a revisionary understanding of what can qualify as a person. Our common-sense understanding is that most human beings are persons. Most persons are human-sized: where there is one person, there is one human being. However, her view entails that there can be group persons, a person composed of more than one human being, as well as multiple persons within one human being sharing one human body. Group persons are possible because groups of individual human beings can come together to work on shared project(s) and have a commitment to overall rational unity amongst those shared projects.⁶³

Multiple persons within one human body are also possible. One possible example may be human beings with multiple personality disorder. Suppose it is true that each personality qualifies as a distinct person because each takes on projects and has a commitment to overall rational unity among its own projects. However, if those

⁶¹ Carol Rovane, “Personal Identity and Choice,” in *Personal Identity and Fractured Selves*. Debra Matthews, Hilary Bok, and Peter Rabins, editors, (Baltimore, MD: John Hopkins University Press, 2009), 105.

⁶² *Ibid.*, 105.

⁶³ Perhaps the ideal marriage is this type of group person consisted of the two human beings in that couple.

personalities do not share a unifying project to unify all of the personalities' projects, this means that there is no commitment to overall rational unity for that human being as a whole. Hence, there would be more than one person within that human being.⁶⁴

For Rovane, personal identity over time is given by unifying projects that are long-term projects; the commitment to overall rational unity extends over the length of time during which one takes on the long-term project. Long-term projects thereby extend a person's identity through time.⁶⁵

Critique of Rovane's theory of personal identity

Rovane contends that her view, while revisionary, provides a perspective from which to embrace a psychological theory of personal identity over a physical theory of personal identity. My main concern with her view is that it is more revisionary than she recognizes, and it leaves too many everyday, human-sized people outside of the category of whole persons.

Speaking, pre-theoretically, people often go through long periods of time having various commitments that are in conflict. It is not simply that those projects are in conflict at various moments due to occasional time constraints, but that having this set of projects is simply incoherent, e.g., there are constant time limitations which makes holding two or more of those projects incoherent or the values behind several of those projects are contradictory. Often a whole person knowingly has conflicting projects, and sometimes he chooses to have those conflicting projects rather than to give up or alter

⁶⁴ Rovane, *The Bounds of Agency*, 132-3, 169-79. Rovane notes that Multiple Personality Disorder may not in fact offer a real life example of multiple persons sharing one body. She carefully makes her case *assuming* some of the personalities have the capacities necessary to qualify as persons but notes that they may in fact lack such capacities.

⁶⁵ This similar to Korsgaard's early comments on the unity of a person over time.

one or more of the projects. Moreover, and most importantly, such persons can lack a commitment to overall rational unity. There is no striving to be rationally unified. The person does not criticize himself for failing to drop or modify one or more projects: instead the person continues to insist on keeping that incoherent set of projects. During such times, it seems to me that the person is committed to those conflicting projects, not overall rational unity between those projects. When people knowingly have non-compatible projects, but do not take those incompatibilities as grounds for self-criticism, many human beings, though pre-theoretically whole persons, are multiple persons on Rovane's account.

For example, suppose some human being, Mary, has the project of being a good mother and the project of being a successful businesswoman, specifically becoming the CEO of her company.⁶⁶ However, the type of person Mary reasons she needs to be in order to become CEO might involve a willingness to be rather ruthless in an effort to make as much profit as possible. At the same time, she holds that in order to be a good mother she needs, among other things, to be a good example to her son by being kind and considerate of others. The ideals of these two projects are (at least potentially) in conflict.

Furthermore, suppose Mary envisions these projects such that they cannot both be successfully achieved by a single human being: suppose she thinks being a good mother consists of being with her son during all of his non-school hours, but being a successful businesswoman means working sixteen hour days every day. On her understanding of

⁶⁶ Rovane argues that in a human being with multiple personality disorder, the alternative identities could be aware of one another's consciousness and yet still be separate persons because the personalities do not have a commitment to overall rational unity through unifying projects.

these projects, these projects do not cohere with one another: if she would like the projects of good mother and successful businesswoman to cohere, she must modify one or both of them. Mary may very well reason that these two projects are incompatible but the only things that she criticizes herself for is not successfully completing both projects. She does not see the lack of overall rational unity as grounds for self-criticism.

Mary lacks a commitment to overall rational unity, even a presupposed commitment to overall rational unity, because Mary, upon reflection, decided to forgo overall rational unity, rather than strive for it. She may wish that such a unity were possible but is not committed to it, given that she realizes that she must alter or modify one of her projects in order to move closer to overall rational unity. It is not simply that she fails to be unified; rather, it is that she has no unifying project over a period of her life. Therefore, for Rovane, it seems that Mary might be more than one person, the businesswoman and the mother.

With this example, I do not want to suggest that it is impossible to be both a successful CEO and a good mother, nor do I want to suggest it is impossible to have a commitment to overall rational unity between such projects. What I *do* want to suggest is that there are people like Mary out there, people who wish for overall rational unity in their lives but who *decidedly* are not committed to it. Such people lack unifying projects that presuppose a commitment to attaining overall rational unity. However, I think we want to say that the Marys of the world are not multiple persons. Rather, we might want to say that they are simply not ideal whole persons, at least in the rational, if not also in the moral, sense. Mary is certainly fractured, but is one person while she strives to be the

successful businesswoman and the good mother, despite her lack of commitment to overall rational unity.

It may seem to some that Mary is unified on Rovane's account because Mary coordinates between these two projects. Mary does coordinate between the project of being a successful businesswoman and the project of being a good mother – she decides when to do which project. However, Rovane does not think mere coordination is enough, and rightfully so. Taylor may fear his roommate, Mike, and so coordinate his schedule around Mike's schedule so that Taylor avoids a violent confrontation with Mike. In Taylor's coordination around Mike, Taylor and Mike do not thereby become one person, not even a group person made up of two human-sized persons. Instead, Taylor is treating Mike and Mike's temper something which is not a part of him but something unpleasant to avoid.

One thought might be that Mary has a unifying project at any given moment, either the project of being a good mother or the project of being a good businesswoman. This may be true. However, the problem is that no unifying project with the required commitment to overall rational unity brings those two projects together, yet we want to say that Mary is one person. Lacking a unifying project over time creates a similar problem for personal identity over time.

Suppose Mary comes to have a project that unifies her projects during a later phase of her life. She keeps the project of businesswoman and the project of mother but adopts a higher-level project with the goal of resolving the conflicts between these two projects: she modifies her projects. In Rovane's view, upon Mary's new commitment to overall rational unity, we are unable to say that Mary is the same person who once led a

very conflicted life. To me, it seems that Mary is someone, *the same person as before*, who finally decided to try to get her life together. However, on Rovane's view, instead we must say there were two persons, who have now merged into one human-sized person. Putting Mary's story aside, many people have many long-term projects which they take on at different points in their lives and which they either complete or drop at various points in their lives. Without a long-term unifying project which would thereby unify one's projects over one's whole life, on Rovane's view many human beings may not be the same individual persons over time over most of their adult human lives.

Perhaps one can think, "I cannot relate to the story of the fragmented Mary." However, what seems true of almost every individual is that he or she changes his or her projects over time. Rovane's theories of both personal identity at a time and especially of personal identity over time are *too* revisionary. We do want to say people like Mary are each a single individual person and that most of us are the same individual persons over time even though our projects change.

Given her emphasis on having a rational point of view, overall rational unity, and choosing unifying projects, Rovane makes the ability to be rational and the agency of persons central in her view of personal identity. However, the structure of her view leads to conclusions which, I argued, are unacceptable.

Conclusion

In this chapter, I first presented Parfit's view of personal identity and the criticism that his view failed to take the agency of persons seriously. Parfit's analysis of personal identity is built on a problematic understanding of strong connectedness. Instead of defining strong connectedness as a percentage of total possible connections, we should

not treat all psychological features on a par rather than giving special importance to those psychological features that we create through our agency. Then I presented the ideas of three other philosophers – Schechtman, Korsgaard, and Rovane. Each of these philosophers made agency a central part of their ideas on personal identity over time. In Schechtman's analysis of personal identity, she focused on agency in how we might *create* ourselves, through narrating. Problematically, on one interpretation her view excludes Episodics who do not narrate while on the other interpretation the importance of narrativity itself is in question given that even Diachronics occasionally engage with the world without self-narration. In their ideas of personal identity both Rovane and Korsgaard emphasize rationality and the capacity to reason, with Rovane fully developing such a view, arguing that the unity of a person comes from having and coordinating projects, which bring with them a striving for overall rational unity between one's particular projects. However, intuitively, people can go through phases in life when they each remain a whole person though they are not committed to overall rational unity. Moreover, it is unclear how a coordinating project accounts for personal identity over time, since that coordinating project will change each time one's projects are changed.

An alternative analysis of personal identity would be one which makes not narrativity but the capacity to reason and rationality, central but not exclusive in the process of creating one's identity. On my analysis of personal identity, in order for a person to change and yet persist, that person is engaged in ongoing creation of his or her identity using the capacity to value, which, I hold importantly but only partly consists in the capacity to reason. The goal of this project is both to describe and defend this new

analysis, which I call the valuing theory of personal identity, as well as make clear the definition of persons that it implies.

CHAPTER 2: A NON-METAPHYSICAL APPROACH TO PERSONAL IDENTITY

Introduction

Parfit hoped to answer a metaphysical question about persons. By “person,” Parfit agreed with Locke: a person is a being that “must be self-conscious, aware of its identity and its continued existence over time.”¹ The question Parfit asked was:

“[For any person], what is it that makes a person at two different times one and the same person?”²

Parfit understood this question as a metaphysical question. Many other personal identity theorists, both psychological and physical theorists, also understood this question as a metaphysical question. However, this project addresses a similar question as a practical question. In this chapter, I will describe the metaphysical version of this question and then I define the practical questions that this project will address.

Metaphysics, most generally

Generally, in the area of metaphysics philosophers focus on discerning what exists, and, for the things that exist, they seek to obtain a description of the nature of those things. Take anything that does or at least appears to exist; call it *x*. Does *x* actually exist, or does *x* merely appear to exist, like a mirage of a pool of water in the desert? There appears to be a pool of water, but, in fact, there is no such pool: it is a sensory – visual – illusion. For example, suppose that *x* appears to be a triangular, red,

¹ Parfit *Reasons and Persons*, 202.

² *Ibid*; 202.

silk cloth. Is x actually there, or is x a mirage of a cloth? Once it is either shown or assumed that x does exist, then there are further metaphysical questions that one could ask.

Assuming that x does exist, what are the necessary and sufficient conditions for x 's numerical identity at a particular moment in time? Numerical identity is a logical relation which has three attributes: (1) reflexivity: x is numerically identical to itself, (2) symmetry: if x is numerically identical to y then y is numerically identical to x , and (3) transitivity: if x is numerically identical to y and y is numerically identical to z , then x is numerically identical to z .

In order to find the necessary and sufficient conditions for x 's numerical identity at a particular moment in time, one can ask, what is the nature of x ; what is x 's essence? While there may be many things that x is, what type of thing is x essentially, such that x could not exist without being that type of thing? Does x have one or more properties that define its nature and are necessary for its current and continued existence, while other properties of x are only accidental properties of x ? A philosopher interested in this question will focus on establishing that while x has numerous properties, most of those properties are accidental, but x is an F such that x cannot exist without being an F . A philosopher is likely to focus on this question when other philosophers are claiming that x can exist without being an F , but x cannot exist without being something else, say a G . x is surely an F , G , and J , but one philosopher defends the essentialist claim that x is an F essentially, and another defends the essentialist claim that x is a G essentially, each explaining why the alternative view(s) that x is essentially something else are false. For example, suppose that x is a red triangle made of silk cloth. Philosophers could disagree

about whether x 's essential property or properties are that of being red, being a triangle, being a silk cloth, or some other property of x , or a combination of two or more of these properties.

Assuming that x exists, another metaphysical question to ask is: How long might x exist? If x can persist for longer than an instant, then one can look for the necessary and sufficient conditions for x 's numerical identity over time. Another way to ask about the numerical identity over time of x is to ask, "What are the necessary and sufficient conditions for x 's persistence over time?" A philosopher focusing on numerical identity over time may simply assume that x is an F essentially, and then attempt to establish the necessary and sufficient conditions for x 's persistence (which may thereby establish that x can persist longer than an instant). This often involves some analysis of exactly what it takes to be not only an F but one and the same F . A philosopher is likely to focus on this question if several philosophers agree that x is an F essentially but there is some debate about the necessary and sufficient conditions for x 's persistence. Suppose several philosophers agree that x is essentially a silk cloth; these philosophers could still disagree about what is necessary for x to persist as one and the same silk cloth.

The Metaphysics of Persons

The traditional debate within analytic philosophy between psychological and physical personal identity theorists is a debate regarding the metaphysical nature of persons. Most assume that we exist, so there are two remaining closely-related metaphysical questions:

- (1) What are we, i.e., you and I, essentially?
- (2) For each of us, what are the necessary and sufficient conditions for each of our numerical identities over time? In other words, what are the necessary and sufficient conditions for the persistence of you or me?

One central issue amongst philosophers regarding the answer to (1) is whether we, i.e., you and I, are persons essentially. Assume by ‘person’ that we mean Parfit’s Lockean-inspired definition: “a person must be self-conscious, aware of its identity and its continued existence over time.”^{3,4} You and I are persons now, but is that merely an accidental property? Could we exist or persist as non-persons?

In response to the question of what we are essentially, the animalist argues that we are essentially animals and only persons accidentally. Animalists hold that I persist if and only if the biological organism that we call Nena T Davis persists. The animalist may go on to address the question of what is necessary for our persistence by describing which of the many properties of biological organisms must continue in order for me (or anything that is essentially a biological organism) to persist. For example, assuming that I am essentially a biological organism, then biological life – living – is necessary for my persistence, and the absence of biological life would be the end of me.

In contrast to animalism, one can ascribe to what Eric Olson calls ‘person essentialism’.⁵ A person essentialist holds that we are essentially persons (though perhaps we are essentially other things as well, such as immaterial substances).⁶ In

³ Parfit, *Reasons and Persons*, 202.

⁴ Locke wrote that a person is “a thinking intelligent being, that has reason and reflection, and can consider itself, the same thinking thing, in different times and places.” (John Locke, *An Essay concerning Human Understanding*, 335)

⁵ Eric T Olson, “Personal Identity”, *The Stanford Encyclopedia of Philosophy (Winter 2010 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2010/entries/identity-personal/>, Section entitled “Understanding the Persistence Question”.

⁶ A person essentialist could hold that we have other essential properties besides being persons; for instance, we essentially could also be animals, or immaterial substances (better known as souls), such that we would cease to exist if either we were no longer persons, or if we were no longer animals, or had souls, respectively. Aristotle held that we are essentially rational animals. If we translated this into the terminology of the contemporary debate, Aristotle appears to hold that we are both animals and persons essentially. ⁷ Some of those who embrace person essentialism hold that being a person is our only essential

contrast to animalism, the person essentialist holds that I would cease to exist if the biological organism called Nena T. Davis fell into a persistent vegetative state, a state permanently lacking the capacities of persons *qua* persons.⁷

Immaterialism is another view of what we are, compatible with person essentialism. Immaterialists claim that we are essentially souls (though perhaps essentially other things as well). Though immaterialism is actually relatively strong as a metaphysical thesis of what we are⁸, immaterialism has been rejected by most contemporary analytic philosophers on scientific grounds.⁹ While immaterialism is implausible scientifically, animalism cannot be rejected on scientific grounds. Olson argues that, problematically, animalism is often ignored as a viable answer to the first question and person essentialism is either implicitly or explicitly assumed by most

property. I will call this exclusive type of person essentialist, a solo person essentialist. A solo person essentialist would say that it is possible that I could persist without biological life, say, as a bionic (robotic) being. According to the solo person essentialist, it is possible that I could persist without all of my properties, except those which are necessary and sufficient for my persistence as the same person.

⁷ Some of those who embrace person essentialism hold that being a person is our only essential property. I will call this exclusive type of person essentialist, a solo person essentialist. A solo person essentialist would say that it is possible that I could persist without biological life, say, as a bionic (robotic) being. According to the solo person essentialist, it is possible that I could persist without all of my properties, except those which are necessary and sufficient for my persistence as the same person.

⁸ Olson argues that the immaterialism has some important strengths that his favored view, animalism, lacks, and vice versa. See his discussion of immaterialism, Eric T Olson, *What are We? A Study in Personal Ontology*. (Oxford, UK: Oxford University Press, 2007), Chapter 7.

⁹ For anything immaterial to affect or be affected by physical things would violate our current understanding of physical laws. In order for the soul to cause a physical object such as an animal to move around, as the soul would need to do since we are animals as well (though perhaps only accidentally), this would require that the soul added energy to an animal. Adding energy to the animal would involve adding energy to the physical universe, since the animal is part of the physical universe. However, according to our best understanding of physical laws, the total amount of energy of the physical universe is fixed. Hence, a soul could not be a causal force on the material plain. Moreover, even supposing we have souls seems to multiply causes unnecessarily: we can explain how we move around, as well as how we affect and are affected by the world, without supposing immaterial souls. These reasons support the general rejection that we are immaterial souls, let alone that one of our essential properties is having a soul.

psychological theorists, including Parfit and Rovane, who debate about personal identity over time.¹⁰

Olson points out that, surprisingly, many physical theorists in the debate regarding personal identity over time are also person essentialists. Many physical theorists¹¹ hold that the continuation of a relatively high-functioning brain¹² is necessary for one's persistence. Some argue that it is not only necessary but also sufficient for our persistence.¹³ Olson points out that there seems to be no explanation for preferring a relatively high-functioning brain to some minimally functioning brain stem except the fact that a relatively high functioning brain enables the continuation of the psychology that we see as definitive of persons. Hence, person essentialism includes among its members physical theorists who hold that our persistence requires the continuation of a relatively high-functioning brain, since that is necessary for our continued personal psychologies.

Olson explains how some personal identity theorists who are person essentialists appear not to realize their commitment to person essentialism. Others realize their commitment to person essentialism but fail to recognize that person essentialism is a

¹⁰ Eric T Olson, *What are We?*, 19. Here, Olson sites as examples: Parfit, *Reasons and Persons*, Part 3. Rovane, *The Bounds of Agency*, 212. For full list of examples, see Olson, *What are We?*, 19.

¹¹ *Ibid.*, 19. Olson sites as an example, Peter Unger, *Identity, Consciousness & Value*. (Oxford, UK: Oxford University Press, 1990).

¹² What is necessary is either an organic brain or the continuous physical realization of something that functions as a high-functioning brain, such as if one were gradually to take out atoms or molecules of an organic brain and replace it with a non-organic parts that can do the same high-level functions in concert with either organic or non-organic parts and ultimately end up with a bionic brain. Gradually would have to be slow enough that during the whole organic-to-inorganic transformation there continued to be high-functioning brain activity.

¹³ E.g., Peter Unger, *Identity, Consciousness & Value*. (Oxford, UK: Oxford University Press, 1990).

contentious assumption, rejecting anti-person essentialist positions, such as animalism, without debate. Unlike animalism and immaterialism, Olson argues that person essentialism is often assumed and not defended with a full metaphysical argument.^{14 15}

¹⁴ Rovane may be a four-dimensionalist as well as a person essentialist. (See Footnote 16 for a brief description of four-dimensionalism.) Olson thinks that Rovane merely states “in passing” towards the end of her book on Page 212 (Rovane, *The Bounds of Agency*) that we are “a set of intentional episodes.” (Olson, *What are We?*, 19). Actually, I believe that Rovane quickly explains on Page 53-56 (Rovane, *The Bounds of Agency*), that she agrees with four-dimensionalism, noting that her manuscript will give some independent reasons to adopt four-dimensionalism. Because she states this so briefly, and cryptically (e.g., she never comes out and says, ‘I am a four-dimensionalist.’) it is understandable that Olson thinks that she merely assumed person essentialism, rather than also favoring four-dimensionalism as the full metaphysical thesis which could accompany her view.

¹⁵ In contrast to those who assume person essentialism, Lynn Rudder Baker is an example of a theorist who provides a metaphysical argument to support her person essentialism. In contrast to animalism and immaterialism, Baker defends a constitutionalist thesis about our nature. Constitutionalism is a metaphysical theory that claims there is a ubiquitous relationship between numerous things with different essences – the relationship of constitution. (Lynn Rudder Baker, *Persons and Bodies: A Constitution View*. Cambridge, UK(Cambridge University Press, 2000), 21.) When x is constituted by y , x is neither numerically identical to nor numerically distinct from y . Constitution is presented as a relation that is an intermediate between numerical identity and numerical distinction. (Lynn Rudder Baker, “Reply to Olson,” http://host.uniroma3.it/progetti/kant/field/bakersymp_replytoolson.htm; Point 7.) The statue *David* by Michelangelo is constituted by the *David*-shaped lump of marble. Living organisms are constituted by individual cells. All macroscopic things are constituted by atoms, which, in turn, are constituted by sub-atomic particles. According to Baker, we are essentially persons (which she defines as something that has the capacity for a first-person perspective) who are constituted by human organisms.

Baker says that we, as human persons, are essentially persons who are necessarily constituted by “human beings of the species *Homo sapiens*” (e.g., Baker, *Persons and Bodies*, 93) that she sometimes calls “human organisms” (e.g., Baker, *Persons and Bodies*, 91). She makes it clear, though, that she holds that human beings of the species *Homo sapiens* can acquire non-organic parts and still be a member of the species *Homo sapiens*: “Although a human body starts out as entirely organic, it can acquire non-organic parts. An artificial leg that I think of as my own, and that I can move merely by intending to move it, becomes a part of my (still human) body. Exactly how much replacement of parts a human body may undergo and still remain a *human* body is somewhat vague, but as long as it continues to be sustained by DNA-based organic processes, it should be considered a human body, a member of the species *Homo sapiens*” (Baker, *Persons and Bodies*, 95). I wonder why her view does not allow what constitutes us to be transformed into something wholly inorganic, so long as the same first-person perspective persists throughout this transformation.

Someone who favors constitutionalism need not agree with Baker’s claim that we are essentially persons. Moreover, someone can favor or reject either constitutionalism or person essentialism and still reject Baker’s definition of person as a being with a first-person perspective. However, within the debate regarding our nature, constitutionalism is used as a way to support person essentialism in light of the fact that persons and human beings are closely related but not numerically identical: human beings began and can persist as non-persons.*

In contrast, animalists (e.g., Eric T. Olson, *The Human Animal: Personal Identity Without Psychology*. (Oxford, UK: Oxford University Press, 1997); David DeGrazia, *Human Identity and Bioethics* (New York, NY: Cambridge University Press, 2005)) reject both constitutionalism paired with person

Olson argues that, when one considers full metaphysical theories which attempt to answer (1) ‘What are we, i.e., you and I, essentially?’ then one sees that animalism and immaterialism, as well as several other metaphysical theories about our nature, each face daunting philosophical challenges.¹⁶ Olson acknowledges that either nihilism about us or

essentialism and, more broadly, constitutionalism of any type. As an argument against Baker’s constitutionalism, animalists claim that we have the pre-theoretical intuition that *we*, not simply the human beings which supposedly once constituted us, can exist and persist as non-persons (e.g., fetuses and human beings in a persistent vegetative state). This pre-theoretical intuition is correct, and we are only persons accidentally.

*Baker consistently holds that we could not persist in a persistent vegetative state. In later work, however, Baker holds that we began as fetuses because fetuses have what she calls a rudimentary first-person point of view. It is unclear to me how fetuses have anything that can be understood as a first-person point of view and this later emphasis on a rudimentary first-person point of view seems to me to undercut Baker’s original position that we are essentially beings with a first-person point of view. However, this inclusion of fetuses and exclusion of human beings in a persistent vegetative state might be more plausible if Baker meant to emphasize the *capacity* to have a first person point of view, which a normal fetus would have while a human being in a persistent vegetative state would lack. This understanding of Baker as alluding to the capacity for a first-person point of view, though, is at odds with her statements about what makes someone the same person over time: it is having the same first-person point of view, which one can recognize experientially from that first person point of view. Besides the basic problem that we can lack self-knowledge and engage in self-deceit – so why should we expect perfect perception here? – it is still unclear how there would be a first person point of view during one’s time as a fetus, of which one could be aware, when one would not even exercise a capacity to have a first person point of view during one’s so-called time as a fetus. The necessity of *having* a first-person point of view is not satisfied by merely having the *capacity* for a first-person point of view.

¹⁶ Olson describes seven alternatives in the literature which attempt to answer to (1). Olson of course argues that person essentialism, without some metaphysical defense, is not a legitimate answer to (1). Besides animalism and immaterialism, two other answers to (1) are constitutionalism and nihilism. One form of constitutionalism, which provides a metaphysical defense of person essentialism, is discussed in the footnote immediately preceding this one. Nihilism is discussed in the main body of the chapter in the paragraph following the paragraph connected to this footnote. The three remaining options are: brainism – the theory that we are the brains of animals, four-dimensionalism – the theory that we are temporal parts, and the bundle view – the theory that we are bundles of psychological features. See Olson, *What are We?*, Chapter 4, for Olson’s summary and critique of brainism.

Four-dimensionalism is the view that we are temporal parts, first introduced by David Lewis (David Lewis, “Survival and Identity,” in A. Rorty, ed., *The Identities of Persons*. (Berkeley: University of California Press, 1976)) to deal with branching cases involving fission or fusion. Lewis’ four-dimensionalism claims that besides the three dimensions of space, we have a fourth dimension, time. So consider a case in which the psychological features of one person are now had by two persons, each with their own stream of consciousness, due to split brain within one human being. Parfit’s account would hold that neither person after the split is numerically identical to the person who existed before the split. This is because numerical identity is a one-to-one relation and a transitive one. (If A is the original before the split, and B and C are the two persons after the split, then A could be not identical to B & C, while B & C are not identical to one another.) The four-dimensionalist would say a person is defined by psychology over his or her entire existence. A was actually hiding two persons: B and C sharing a psychology for a

global nihilism may be true and that there is no sort of thing that we are. According to nihilism about us, “[w]e don’t exist. Our personal pronouns refer to nothing, there being nothing there for them to refer to. Nothing thinks our thoughts. Nothing wrote these words and nothing is reading them.”¹⁷ According to nihilism about “us”, there is nothing that could be “you” or “I” that is numerically identical to itself at or over time, hence there is nothing that “we” are. “We” only appear to exist and persist like that pool of water in the desert, which is in fact only a mirage. (Global nihilism holds that this is true of everything: there is nothing that satisfies the logical form of an identity relation at or over time, hence nothing exists or persists.) Olson finds nihilism unappealing, but he recognizes that there is still work to do in defending animalism, or any other view, against it.

Olson considers the strengths and weaknesses of animalism, immaterialism and several other theories. In regard to some issues, though, he remains silent. He makes no claims about the nature of personhood or the persistence of persons as such. He simply assumes that the Lockean definition of a person to which Parfit subscribes is correct. But unlike Parfit, Rovane, and many others, Olson is unconcerned with what are the necessary and sufficient conditions for any person to remain the same *person*. Olson

while and then B and C parted ways after the split. So before the split there were two people sharing the same psychological life, but this only became apparent after the split, which revealed two persons.

The bundle view is a metaphysical defense of psychological theorists’ assumption of person essentialism offered by Sydney Shoemaker. Olson thinks the main problem with the bundle view is the idea that there are thoughts, intentions, desires – psychological features – without some thing that has those psychological features. He thinks it is implausible that, instead of something that has those psychological features, there could be unthought thoughts or thoughts that think themselves. But if there are things that have those psychological features, then aren’t *we* those things, rather than the bundle of psychological features that those things have? (Olson, *What are We?*, Chapter 6.) See Olsen for further details.

¹⁷ Olson, *What are We?*, 180.

makes it clear that he is interested in the metaphysical question about *our* nature and about *our* persistence, and, since he thinks *we* are not persons essentially, he is not interested in the nature of *our* personhood or the necessary and sufficient conditions of *our* persistence as the *same persons*.^{18, 19}

While Olson finds analyzing personhood and the persistence of persons as such uninteresting, it is of central importance to a personal identity theorist who is a person essentialist, such as Parfit or Rovane. Even a person essentialist who provides a full metaphysical defense of his person essentialism would appear to have an incomplete description of our nature if he failed to define what persons are: the plausibility of the view that we are persons essentially at least partially depends on exactly what persons are argued to be. Moreover, personal identity theorists who are person essentialists are not the only philosophers interested in the nature of persons and their persistence conditions. Moral philosophers and others interested in grappling with moral questions want to understand the nature of persons and their identities *at* and *over* time.²⁰

¹⁸ E.g., Olson, *What are We?*, 44.

¹⁹ Here, I feel that I must be honest. Upon reading Olson I realized that I too had been assuming person essentialism. I share the intuition of a large number of personal identity theorists that we are essentially persons. However, animalists Eric Olson and David DeGrazia helped me to realize that I may have jumped to this assumption, likely because being a person is of central importance to our lives. Yet, both Olson and DeGrazia each note that the fact that something is of central importance to our lives does not appear to be a fact that can support a metaphysical thesis regarding what we are. (Eric Olson, "Personal Identity", *The Stanford Encyclopedia of Philosophy (Winter 2010 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2010/entries/identity-personal/>, Section on Fission) and David DeGrazia, "Are We Essentially Persons? Olson, Baker, and a Reply," *The Philosophical Forum*, Vol. 33, No. 1 (Spring 2002), 113). We may essentially be something that is relatively unimportant to us, such as biological organisms, which can persist in persistent vegetative states, even if this persistence would likely be of no importance to most of us.

²⁰ Locke's project was begun to answer a moral quarry of: For what actions is someone responsible such that God will judge him for those – and no more or less – on Judgment Day? (See Uzgalis, William, "John Locke", *The Stanford Encyclopedia of Philosophy (Winter 2010 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2010/entries/locke/>)

Shifting the Questions from the Metaphysical to the Practical and Ethics

The questions of this project are practical ones. They are:

- (1a) For any human being who is a person, what are the necessary and sufficient conditions for its personhood?
- (2a) For any human being who is a particular person at one time, what are the necessary and sufficient conditions such that that human being is the same person at a different time?

As a philosopher interested in addressing moral questions, I want to answer these non-metaphysical questions about the persistence of the same persons. This project does not concern itself with which metaphysical thesis is true about *us*. Even global nihilism could be true and it would not undermine this project.

Even if, metaphysically speaking, *we* do not exist or persist because nothing exists that satisfies the logical form of an identity relation at or over time, the practical questions of concern in this project still arise. If global nihilism is correct, “persons” and “human beings” do not exist let alone persist. However, from the practical standpoint, “we” do exist and persist and “human beings” persist and sometimes persist as the same or a different “person”. Moving forward, I take the practical stance regarding myself, you, human beings, and persons. As a practical matter all of these things exist and we are faced with deciding how to treat them and, so I am dropping the double quotes around

Locke’s answer is, not all actions done by a particular human being or all actions done a particular immaterial substance, but those all actions done by the same person, which he argues is “a thinking intelligent being that can know itself as the same thinking thing in different times and places”. Olson’s discussion implies that it is problematic for Locke to assume that this practical question could be answered with metaphysical analysis that successfully provides some form of numerical identity. (Olson, “Personal Identity”) Schechtman and Korsgaard make similar points. Schechtman argues that it is problematic for Locke, and Neo-Lockeans such as Parfit, to assume that any metaphysical analysis which provides numerical identity can answer such practical questions. (Schechtman, *The Constitution of Selves*, Chapter 2) Korsgaard suggest that one cannot expect one’s metaphysical analysis which provides numerical identity to have practical importance. Without a successful argument defending combining the metaphysical and practical, one needs to address separately a metaphysical question of what we are (which provides numerical identity) and a practical analysis that is relevant to how persons ought to treat one another. Also see David Shoemaker arguing that a metaphysical analysis of personal identity is not relevant in bioethics in “The Insignificance of Personal Identity for Bioethics,” *Bioethics*. Vol. 24, No. 9, 481 – 489.

words such as ‘I’, ‘you’, ‘human beings’, and ‘persons,’ double quotes that global nihilism would demand. As a practical matter, we are faced with questions about how to treat human beings who are persons but may not continue to be the same persons. “How should we treat this human being or that person?” The question does not go away even if human beings or persons fail to satisfy the logical form of a numerical identity relation.

The practical questions of this project hope to help answer moral questions about how to treat a human being who may or may not persist as the same person. Hence, I restrict the scope of my questions to the persistence of persons given the same human being.²¹ As a practical matter, we do not have to face fantastical thought experiment cases such as body-switching, brain splitting possibly combined with brain transplantation, or duplication of a human being who is a person. The moral questions that we face regarding the same person are questions that are tied to whether *a given human being* persists as the same person. Margo is an example of when these questions arise.

The case of Margo, an Alzheimer’s disease patient, was introduced in full in the opening pages of this project.²² In an article, Andrew Ferlik, a medical student describes a time during her illness when her life is very repetitive, simple, and happy but her memory is minimal. He concludes his brief summary saying:

Despite her illness, or maybe somehow because of it, Margo is undeniably one of the happiest people I have ever known. There is something graceful about the degeneration her mind is undergoing, leaving her carefree, always cheerful. Do her problems, whatever she may perceive them to be, simply fail to make it to the worry centers of her

²¹ Kathleen Wilkes notes that these are the only persons that we actually know, human persons. Kathleen Wilkes, *Real People. Personal Identity with Thought Experiments* (Oxford: Oxford University Press, 1988), 22 – 23.

²² See opening paragraphs of the Introduction to the entire project.

brain? How does Margo maintain her sense of self? When a person can no longer accumulate new memories as the old rapidly fade, what remains? Who is Margo?²³

Ronald Dworkin supposes that Margo, while still legally competent, left a Do Not Resuscitate order (a DNR order) stating that she should not be resuscitated once she had the kind of existence that Ferlik observed. Dworkin argues that, if she had left such an order, we should honor it.

Dworkin uses narrativity to defend his position, suggesting that, because Margo would have made a DNR order referring to the period late in her illness as her own, that part of her life is part of her identity.²⁴ Hence, Margo's identity is such that she has stronger interests in not being resuscitated than in continuing to have these pleasant experiences. So even if Margo no longer has personhood during the time Ferlik observed her, her narrative, formed during personhood, unifies her as having years of personhood and non-personhood.

Dworkin's dependence on narrativity is problematic. Recall my concerns about self-narrativity, which I discussed in my critique of Mayra Schechtman's self-narrativity view in Chapter 1. A thoroughgoing narrativity view is flawed because (1) it fails to adequately include Episodics, who are non-narrators, as persons and (2) even Diachronics do not always narrate and fail to include within their narratives every aspect of who they are. (A Diachronic might never describe herself as conceited, though she is, or include in any self-narratives, her mundane drives home from work, though she made those drives.)

²³ Ferlik, "Margo's Logo," 201.

²⁴ Dworkin actually assumed that narratives cover entire human life times. However, this assumption is strange: what if a human being only had personhood for a few years, then at age 20, that human being lived the remainder of her existence, say another 60 years, in a persistent vegetative state due to an accident. It would be difficult to see how a narrative – which was never formed while the human being was person – would unify the 70 years of this human being's life.

In defense of Dworkin, though, someone could respond that even if a thoroughgoing narrative view has these problems, if Margo were to leave a DNR order, *she* was a narrator who took the time to include these non-narrating years in her self-narrative. If someone makes such request based on his or her self-narrative, shouldn't we honor it?

I respond, no, we should not honor the DNR order because that period of time was a part of her self-narrative. I think that if Margo were to have made such a self-narrative it would be inaccurate and therefore not self-constituting.

As Schechtman reasonably argues, in order to be self-constituting, a self-narrative must be accurate. For example, someone alive today fails to have a self-narrative that is self-constituting, if that self-narrative says that he is Napoléon and it is the 19th century. However, both Schechtman and Dworkin problematically assume that self-narratives are accurate so long as those self-narratives do not contradict things done or experienced by that human being within that human being's lifetime.²⁵ But why are self-narratives accurate if those narratives include periods of time when the human being persists but without personhood?²⁶ Moreover, even if Margo lacks personhood by the time Ferlik met her, it is unclear whether Margo is the same person during earlier stages of her illness.

²⁵ Schechtman of course allows for some degree of inaccuracy. Schechtman's example is of a person who narrates about how he told a funny joke at last weekend's party but the person did not tell this joke there. However, if the person is someone who generally tells such jokes – maybe he told that very joke a month ago – this inaccuracy in his self-narrative does not undermine whether that narrative can be self-constituting.

²⁶ One answer might be that such a self-narrative is self-constituting because that self-narrative conforms to the metaphysical facts, say, if animalism is true. However it is unclear why any metaphysical fact is morally relevant. Instead, we honor DNR orders to honor persons, so we need an analysis of persons that does not assume that the entire life of a human being can be included in defining persons.²⁷ She might form such an order in advance because many people in her family had Alzheimer's disease.

Suppose Margo had signed a different DNR order that takes effect immediately upon a diagnosis of Alzheimer's disease.²⁷ Call this the Comprehensive DNR order. In the Comprehensive DNR order, Margo explains that if she is diagnosed with Alzheimer's disease, starting the day of diagnosis, if resuscitation is necessary for her survival, she does not want to be resuscitated. If Margo does not have to be resuscitated for a number of years after signing the DNR order, it is not at all clear that Margo would be the same person. At some point in Alzheimer's disease, a human being is no longer a person: the patient is bed-ridden and, most importantly, nonresponsive. However, the mental decay due to Alzheimer's disease is progressive and can be slow; human beings can live as long as twenty years with the disease.²⁸ Could there be a stage in the disease where rather than being the same person as before the disease, the human being is a *different* person? The case of Margo with the Comprehensive DNR order highlights how any self-narrativity view is problematic if it merely *presumes* that the chronological borders of a human being can define the outer most chronological borders of a self-constituting narrative.

Rovane and other personal identity theorists,²⁹ take the position that we cannot assume that persons are human-size and persist the length of the human being's personhood. Similarly, I also form an analysis of personal identity over time that defines persons without problematically presuming that the boundaries of persons are defined by boundaries of human beings. The analysis assumes neither that the physical boundaries of human beings make it the case that a human being can be at most one person at a time

²⁷ She might form such an order in advance because many people in her family had Alzheimer's disease.

²⁸ Luc Jasmin, MD, PhD, VeriMed Healthcare Network, David Zieve, MD, MHA, "What is Alzheimer's Disease?" A.D.A.M., Inc. http://www.alz.org/alzheimers_disease_what_is_alzheimers.asp

²⁹ E.g., Wilkes, *Real People*, Chapter 4.

nor that the chronological boundaries of human beings make it the case that a human being can be at most one person *over time*.³⁰

Conclusion

In this chapter, I defined the questions of this project, which are practical rather than metaphysical. They are:

- (1a) For any human being who is a person, what are the necessary and sufficient conditions for its personhood?
- (2a) For any human being who is a particular person at one time, what are the necessary and sufficient conditions such that that human being is the same person at a different time?

These questions are different from the ones that many previous personal identity theorists addressed. Many previous personal identity theorists asked metaphysical questions regarding *our* nature in a way that assumed that *we* were essentially persons. In this project, I put aside these metaphysical questions. Instead, I focus on a practical analysis of personhood and the persistence of persons given the same human being. In answering these practical questions, I hope to shed light on the identity of Margo and more generally, on the nature of both human persons and their persistence.

³⁰ Rovane boldly provides such analysis.

SECTION I – CONCLUSION

The Questions

The two questions of this project are:

- (1a) For any human being who is a person, what are the necessary and sufficient conditions for its personhood?
- (2a) For any human being who is a particular person at one time, what are the necessary and sufficient conditions such that that human being is the same person at a different time?

I will answer these questions in the spirit of Locke's analysis of persons and personal identity. Locke's definition of a person was "a thinking intelligent being, that has reason and reflection, and can consider itself, the same thinking thing, in different times and places."¹ In addition he thought that personal identity consisted in the phenomenological unity of consciousness, whereby through one's consciousness, one perceives that one is the same self of the past.^{2, 3}

¹ Locke, *An Essay Concerning Human Understanding*, 335.

² *Ibid.*, 335.

³ Locke's exact words were:

This being premised, to find wherein personal identity consists, we must consider what person stands for; which, I think, is a thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing in different times and places; which it does only by that consciousness which is inseparable from thinking, and, as it seems to me, essential to it: It being impossible for any one to perceive, without perceiving that he does perceive. When we see, hear, smell, taste, feel, meditate, or will any thing, we know that we do so. Thus it is always as to our present sensations and perceptions: And by this every one is to himself that which he calls self ... For since consciousness always accompanies thinking, and it is that which makes every one to be what he calls self, and thereby distinguishes himself from all other thinking things; in this alone consists personal identity, i.e. the sameness of a rational being: And as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person; it is the same self now it

Locke's claim that persons are intelligent and have reason and reflection seems to me to be correct. However, in holding that personal identity consists in the phenomenological unity of consciousness, by which one experiences oneself as persisting over time, Locke excludes Episodics as persons and persons who persist over time, since Episodics can lack the phenomenological experience of persisting over time.

Working Definition of 'Person'

I agree with Parfit on his methodology of starting with a working definition of persons inspired by Locke, and focusing on formulating an analysis of personal identity over time, though the final analysis of personal identity over time might alter the definition persons.⁴ Unlike Parfit, whose working definition of persons seems to be in complete agreement with Locke, my working definition of person is a modified version of Locke's definition of person:

Working Definition of a Person:

Person: A thinking intelligent being that has reason and reflection and *may* consider itself the same thinking thing in different times and places.

Criteria for a Successful Project

In forming their respective theories of personal identity, Parfit, Schechtman, and Rovane were each inspired by Locke and sought to improve upon Locke's analysis. In Chapter 1, I discussed the views of Parfit, Schechtman, Rovane, and Korsgaard, and claimed that each offered important insights into the nature of personal identity and persons, but each were also unsatisfactory. Parfit's view was an important expansion of

was then; and it is by the same self with this present one that now reflects on it, that that action was done (*Ibid.*, 335).

⁴ Parfit, *Reasons and Persons*, 201.

Locke's view in that it considers various psychological features, not just experiential memories or some sort of conscious sense that one persists over time. In claiming that all of our psychological features are relevant to our persistence, rather than just any self-perception that one persists over time, Parfit's view does not exclude Episodics as Locke's view does. However, Korsgaard notes that Parfit's analysis problematically treats persons as a bundle of psychological features, minimizing rather than emphasizing that persons are agents. Schechtman's view focused on the agency of persons. However, in requiring some degree of self-narrativity which connects one to one's further past and one's further future, Schechtman's view fails to include those who do not use narrative to form their whole identities (e.g., Episodics such as Strawson and non-narrating or intermittently narrating Diachronics). Pre-theoretically these individuals do not appear to lack identities and are persons who have a past and a future, though self-narratives would not create these identities. While Korsgaard takes rational agency to be a central feature of persons, she does not explain how a person is unified, though perhaps not ideally, when he fails to act on formal rational principles, such as the Categorical Imperative. Rovane's view focuses on rational agency as well and I believe that having unifying projects and also striving to achieve overall rational unity probably unifies persons to some degree. However, many human beings are pre-theoretically human-sized persons that go through stages of life when they are committed to incoherent projects, and reject overall rational unity in favor of those projects. Hence, Rovane's view unduly treats as multiple persons those who, pre-theoretically, we view as whole human-sized persons. Though Rovane holds that an overall commitment to rational unity is required to be a unified person, we can imagine what appears to be a whole person – Mary, the conflicted

businesswoman and mother, who lacks a commitment to overall rational unity. Rovane's focus on rationality does not provide a comprehensive account of the identities of persons.

In light of the strengths and shortcomings of previous views, I will present and defend an analysis of persons and personal identity over time that is (1) in the spirit of Locke with an emphasis on reason and reflection, (2) makes agency central rather than peripheral, (3) will not unduly exclude Episodics and non-narrating Diachronics as persons and persons who persist over time, and (4) will not unduly label those who appear to be whole but internally conflicted human-sized persons as group persons.

SECTION II: THE VALUING THEORY OF PERSONAL IDENTITY OVER TIME

In this introduction to Section II, I will describe the preliminary version of my theory and argue that it is supported by some of our intuitions regarding the persistence of persons. The chapters of Section II will move us from this preliminary version of the theory to the final version of the theory of personal identity and also reveal my final definition of ‘person’.

Introduction of Preliminary Version

In the Preliminary Version of my theory of personal identity over time:

Take some time interval t which starts at time t_1 and goes up to and includes some future time, time t_2 . Take some human being x that exists during t and who is person q at time t_1 . X is q at t_2 if and only if during t how and what x values remains the same or differs as a result of q 's own reflection.

According to the Preliminary Version, there are two ways that a human being can persist as the same person over some period of time. The first is when what and how she values remains the same because, on her reflection, she continues to agree with what and how she valued previously.

In order to value exactly the same, one must value the same things. Moreover, how one values those same things – the way one values those same things – must also be the same. For example, Holly values nothing more than her child, Richard, at times t_1 , t_2 , and t_3 – each time point separated by one year. However, due to having what she felt

were special moments with him as he grew over the year, her valuing of Richard was deeper at t_2 than at t_1 . This means that her valuing of Richard is slightly different between t_1 and t_2 . Suppose that, by t_3 , she has another child, Jessica, who she values to the same degree but differently than she values Richard. She values Jessica this way for the same reasons that she already value Richard that way, such as family is important to her. By t_3 , Holly values nothing more than her two children. Given how she once valued Richard more than anything else but now she values Richard *and* Jessica this way, at t_3 , Holly values differently than she did at t_2 .¹ So, even though it is true that Holly values nothing more than Richard, at t_1 , t_2 , and t_3 , her valuing changes from t_1 to t_3 , first her valuing of Richard grew deeper and then she valued Jessica similarly to her valuing of Richard.

Suppose that, in every other respect, how and what Holly valued remains the same, between t_1 through t_3 . Then, the time period from t_1 until Holly begins to value Richard more deeply, what and how she valued did not change. Intuitively, given this lack of change in what and how she values, Holly is the same person during that time interval. On the valuing theory, during that time interval, Holly remains the same person.

Of course, the idea that a person persists if what and how that human being values remains exactly the same does not get us very far into the real issue at hand. The question of a person's persistence arises because we see psychological differences in human beings over time. In particular, what and how a human being values changes over time, sometimes relatively little but sometimes quite a lot. Intuitively, Holly from t_1 to t_3

¹ When Holly values both above all other things, valuing Richard will sometimes involve different actions and lead to different emotions than when Richard was an only child. Hence, she values Richard differently because of coming to value Jessica. These types of connections between different things valued are discussed in Chapters 4 and 5.

is one example of a human being who is the same person though what and how she valued changed. The valuing theory would agree with the intuition that Holly is the same person.

The valuing theory would say that Holly's changes in valuing were a matter of the reflection of one continuing person. First, given that she already valued Richard greatly, Holly's original valuing of Richard deepened from what she understood to be special moments with Richard. Second, she came to value Jessica for reasons – e.g., her valuing of family – which also account for her ongoing valuing Richard.

Now compare our intuitions against how the preliminary version of the theory would apply to two stories of Parfit's Russian.²

In the first story, there is the human being, Ivan. As in Parfit's original story, this Russian, while raised in the establishment, was a revolutionary in his twenties. Suppose that when we look into the life of Ivan twenty years later, we see that he supports the establishment. Over the years, Ivan reflected on new experiences and reconsidered the pro-establishment ideals taught to him in his youth. This shift towards the establishment was partly due to the love of his parents. He suffered a severe leg injury when he was thirty and his parents, despite his earlier distancing of himself from them for being pro-establishment, helped him to recover. In addition, around the same time, Ivan learned of a revolutionary uprising that led to the torture and killing of noble families, including children.³ Both his parent's love and these horrifying killings (murders in his mind) led him to shift from disvaluing towards valuing the establishment.

² Parfit, *Reasons and Persons*, 327 – 8.

³ Such events are wholly fictional on my part and possibly anachronistic. In the 2010s torture is a reoccurring theme in both suppressive regimes and the uprisings against such regimes.

Intuitively, Ivan is the same person during that twenty year stretch of time, even though at the beginning and end of that period of time he values very differently. Any talk of Ivan as a different person is only metaphorical. When we say he is a different person at forty than he was at twenty, we are simply pointing to the pronounced change in what he values.

How can we tell that this talk of Ivan as a different person is only metaphorical? Others treat him as the same person. Suppose that, when supporting the revolution, Ivan had an affair with Dmitri's wife Alena. Dmitri was one of his closest childhood friends, a member of the Establishment, and Ivan's affairs with Dmitri's wife helped to destroy Dmitri's marriage and family. Twenty years later, Dmitri rightly would hold Ivan responsible and feel great anger towards him for what he did. In addition, if Ivan made a promise to support the revolution for his entire life, his cohorts in the revolutionary cause would have good reason to see him as a traitor twenty years later. Those cohorts would rightfully hold Ivan accountable and obligated to keep past promises and commitments.

The valuing theory would agree with the intuition that Ivan was the same person. On the valuing theory, like Holly, Ivan's changes in what and how he values are the result of the reflection of a persisting person. The love of his parents and the killings that horrified him led him to rethink his anti-establishment ideals. Ivan is the same person over those twenty years, though what and how he valued was very different. In contrast, consider a second story, the story of Boris.

The story of the human being Boris begins and ends very much as Ivan's story does: Boris was raised in the establishment, was a revolutionary in his twenties who slept with the cuckolding Alena, and when we look into his life twenty years later, we see that

Boris supports the establishment. In contrast to Ivan, suppose the difference between what Boris valued as a young revolutionary and what he values later is explained by a brain injury at age thirty. This injury did not compromise many of his mental capacities, such as speech and the capacity to value moving forward. However, this injury caused this human being to no longer value anything as that human being once valued it.⁴ In other words, Boris does not support the revolution or value anything else unique to him⁵ as he did immediately before the injury. He knows that this human being once valued the revolutionary but he does not – he hasn't decided whether to value or disvalue the revolution. Boris moves on from this brain injury. The love and support of *his* parents as he must begin his life over again touches him and he, like Ivan, learns of the revolutionary uprising and finds it horrifying. This leads him to both disvalue the revolution and value the establishment.

Intuitively, Boris is not the same person from age twenty to age forty. The valuing theory agrees. On the valuing theory, the difference between Ivan and Boris is that Ivan transformed himself through his own reflection into someone who valued the establishment, while Boris was a different person before the brain injury than he was after the brain injury. Ivan valued the establishment at the end of the twenty year period,

⁴ Boris could be a case of permanent retrograde amnesia, in which all of a human being's past cannot be recalled – the cells in the part of the brain which stored the information died. Alternatively, this loss of what and how one valued could be caused by a brain aneurism, but that explanation is more controversial. I heard of a case in which a woman had a brain aneurism and awoke with all of her memories, including memories of what and how she once valued. She and daughter agreed that she was a changed woman after the aneurism. Their recalling of this discussion of this change made it clear that she made no conscious decision that she should value differently, e.g., she did not appear to consciously decide that how she valued has been wrong in many ways. Instead this woman and daughter describe her as simply valuing very differently after she awoke. Without more details, the valuing theory cannot begin to assess what happened here.

⁵ So, after the injury, Boris might quickly come to value satiating his hunger, disvalue pain, or disvalue the disorientation he is experiencing from not valuing much of anything. But these responses seem a likely response for anyone and so could not indicate that that former person was still there.

but his own agency drove those differences, while Boris lost the valuing of the revolution, not through reflection, but due to the brain injury. Later, Boris' reflection did lead him to reject the revolution, but this was the reflection of a different person than the twenty-year-old Boris.

In contrast to forty-year-old Ivan, Dmitri should not hold the forty-year-old Boris responsible for partly destroying his marriage and family.⁶ If Dmitri truly understands Boris' brain injury, he might still really hate the person who helped to destroy his marriage and family but also realize that Boris is no longer that person. If Dmitri is angry enough he might even think sarcastically, "A brain injury could not have happened to a nicer guy," and, with spitefulness, think, "Good riddance to him!" Dmitri might still have angry feelings when he sees Boris but also think, "I cannot hold this man responsible for what happened twenty years before."

Moreover, if, in his twenties, Boris promised to support the Revolution his whole life, and his cohorts truly understand his injury, they will not see him as a traitor: they know that Boris is no longer the person they knew. The former cohorts still have plenty of reason to dislike the forty-year-old Boris: Boris is actively engaging in activities which are contrary to their revolutionary ideals. But the reasons to dislike the forty-year old Boris are exactly the same as those in favor of disliking any pro-establishment person. There is no additional reason to dislike Boris on the basis of betrayal. The person that they knew did not decide to betray them: that person no longer exists.

⁶ Some examples don't work as well: Boris might still be charged for the crimes like murder twenty years later though Boris is now a different person. The law is concerned with preventing people from forming questionable defenses and getting away with heinous crimes. So the law might reject a defense of "I'm a new person thanks to a brain injury," if the person who committed the crime were of her right mind during the crime and of her right mind now. The law is interested in deterrence and protection as well as retributive punishment.

On the valuing theory, Ivan is the same person over those twenty years because what and how Ivan valued became different in light of what and how Ivan *had* valued and Ivan taking in new experiences and having new thoughts and feelings regarding those experiences. In contrast, on the valuing theory, Boris is not the same person because Boris did not change his valuing suddenly or gradually from anti- to pro-establishment. Instead, he awoke after his injury no longer valuing the revolution; he then went on to value the establishment. In the preliminary version of my theory, a human being is the same person (q) over time if and only if what she values remains the same or when what she values becomes different, those differences are a result of q 's own reflection.

From considering the personal identities of Holly, Ivan and Boris, I believe that I have shown that the preliminary version of the theory agrees with intuitions about personal identity. The next step in developing and defending my theory of personal identity over time involves defining the capacity of reflection that is central to the persistence of persons. On my theory, the capacity to reflect upon and alter what and how one values is the capacity to value. I hold that the capacity to value is an integration of the capacities to reason and feel. The next chapter focuses on describing the exact nature of the capacity to value.

CHAPTER 3: THE NATURE OF VALUING

Introduction

In this chapter, I discuss the nature of valuing, by which I mean to include both valuing and disvaluing. The capacity to value is the capacity for a type of reflection that includes a central emotional component and enables a person to persist over time.

Two distinctions must be made in regard to valuing. The first distinction is between valuing something and judging that something is valuable or that one should value it. The second distinction is between valuing something and one's values. These distinctions will make it clear that my theory of personal identity over time centers on valuing, not on merely one's value judgments or merely on one's values.

Distinction 1: One's valuing v rather than One judging that v is valuable or that one should value v

Actually valuing some v , some thing that one could value, is different from simply reasoning that v is of value. A teenager might reason that his mother is of value and, moreover, that he should value her, specifically, cherish her. Suppose he learns that his mother may perish, that she is diagnosed with brain cancer.¹ If neither (a) learning of,

¹ It is difficult to think of a case where perishing is uncontrovertibly an evil, which is the kind of example I am trying to create, an example where the mother is in jeopardy in the mind of the teenager himself. For example, if the teenager has certain religious beliefs, he might think that there is an afterlife, and it is better than this life, and so think that his mother will not truly perish and dying will be better than this life. So, he might not experience sadness upon learning of her bad prognosis or joy upon learning she is in remission, which is what one would expect if someone cherished another. Moreover, I tried to choose an example where the risk of perishing could not have been partly created by poor health choices. Brain cancer patients seem able to avoid the scorn of the harshest of critics, unlike smoking lung cancer patients, or sunbathing skin cancer patients. If the teenager thinks that she partly and knowingly risked the creation

(b) her living with, nor (c) the resolution of, via remission or death, his mother's brain cancer brings about *any* emotions in him, then he does not, on my understanding of valuing, value her.² Instead, at best, he merely thinks that she is valuable or that he ought to value her.

Distinction 2: One Values (verb) rather than One's Values (noun)

Rather than focusing on merely one's ideals, which we might call one's 'values,' the valuing theory focuses on the capacity to engage in an activity, the activity of valuing. The activity of valuing includes valuing and devaluing. Normally, persons certainly value ideals but there are many other things that persons value as well, e.g., "objects and their properties (such as beauty), persons, skills and talents, states of character, actions, [and] accomplishments."³

of her cancer, and the teenager did cherish her, he might be angry at his mother and lash out at her during her time with cancer. However, even if the mother caused her own cancer and death itself (from the cancer or anything else), the teenager would have to think and find this cancer pretty painless for his mother in order for him to cherish her and experience *no* emotions regarding his mother during her time with cancer.

² This is not to say that if he experiences any emotions regarding her cancer, it means that he *does* value her. Whether we would say that he values her would depend on the exact nature and contexts of those emotions. This project does not intend to settle what counts as true cases of cherishing, or more generally, true cases of valuing. This project focuses on the nature of valuing and its importance in the persistence of persons over time. The teenager might have an emotional reaction to his mother's cancer because he devalues her, rather than values her, so he is (horribly) happy at the news that she may die. My point here is simply that the son does not value or devalue his mother if he has *no* emotions at all regarding his mother's cancer. If he simply thinks that he ought to value and cherish her, but has no emotions regarding her even during her time with cancer, then he does not value or devalue her.

³ Thomas Scanlon, *What We Owe to Each Other*, (Cambridge, MA: Belknap Press, 1999), 95.

Capacities: Emotion, Reason, Valuing

The Capacity to have Emotions

Though the valuing theory is compatible with a somatic analysis of emotions,⁴ I am defining the capacity to have emotions using a composite analysis of emotions. A more detailed discussion of this composite analysis of emotions is available in Appendix A.

On the composite analysis, emotions are something had by non-human animals as well as human beings who are either persons or non-persons. On the composite analysis of emotions, emotions are a combination of judgments and feelings. By feelings, I mean clusters of bodily sensations.⁵ In everyday language, we might say, “He feels his feelings,” referring to his emotions as “feelings”. I would translate this as, “He has – or feels – the bodily sensations of which his emotions partly consist.”

Each particular emotion (e.g., anger, sadness) has a specific cluster of bodily sensations associated with it. For example, the emotion of anger consists of bodily sensations such as experiencing heightened awareness, having rapid heartbeat, and experiencing muscle tension.

On the composite analysis of emotions, emotions are bodily sensations and judgments. For example, one can grow indignant on behalf of present-day immigrants in

⁴ One might be skeptical of or simply reject a composite analysis of emotions, and favor a somatic analysis of emotions which holds that emotions are only the clusters of bodily sensations which are sometimes elicited by judgments, but those judgments themselves are not part of the emotions. Someone favoring a somatic analysis of emotions need not reject the valuing theory of personal identity. Instead they can simply translate my use of the term ‘emotion’ to mean a composite of bodily sensations and eliciting judgments. See Appendix A.

⁵ I found the use of the word ‘cluster,’ in discussion of the somatic theory of the emotions in Jesse Prinz, “Embodied Emotions” in *Thinking about Feeling: Contemporary Philosophers on Emotions*. (New York, NY: Oxford University Press, 2004).

response to listening to extreme political commentary which vilifies present-day immigrants. That indignation, as a form of anger, will be composed of (1) bodily sensations such as rapid heartbeat and muscle tension and (2) judgments, which elicit those bodily sensations, such as the judgment that the commentary is unfair in its statements regarding present-day immigrants.

The Capacity to Reason⁶

The judgments involved in valuing come from the exercise of the capacity to reason. As I am defining it, the capacity to reason is the capacity to engage in the activity of *seeing* and *finding reasons*, reasons to think, have certain emotions, or take certain acts. Hence, the noun, ‘reason’ and the activity of ‘reason’ – reasoning – are closely related: when we engage in reasoning – or valuing which I claim partly consists of reasoning – we find or see reasons.

As Scanlon notes, what it is to be a reason is undefinable:

Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favor of it. “Counts in favor how?” one might ask. “By providing a reason for it” seems to be the only answer.⁷

Seeing reasons and finding reasons are different. To see something as a reason is to see it as a starting point in one’s immediate deliberations or see it as an ultimate reason, a foundational reason that cannot be justified by deliberation.⁸

⁶ This analysis of the capacity to reason borrows heavily from Scanlon, *What We Owe to Each Other*, Chapter 1.

⁷ Scanlon, *What We Owe to Each Other*, 17.

⁸ Foundational reasons can be mutually grounded such that A justifies B, which justifies C, which justifies D, ..., which justifies Z, which justifies A, making A though Z foundational. Though it might always seem as if each of these has a further justification behind them, in fact each of these is supporting one another.

Between normative realists, anti-realists, and idealists, there is a debate about what seeing ultimate reasons involves. Is seeing ultimate reasons the product of some innate intuition, some developed attunement to objective normative truths, or is it, in fact, simply something that comes with being able to reflect about what to do, think, or feel? This debate, though, does not need to be resolved in order to acknowledge that the capacity to reason includes seeing reasons, seeing things as one's immediate starting point or as one's ultimate reasons.

In contrast to seeing reasons, finding reasons involves rational deliberation, though it is grounded in simply seeing some thing(s) as constituting reason(s) e.g., as a reason, as no reason, as an overriding reason, as an inadequate reason. Rational deliberation includes the perfect application of rules of logic on propositions which lack any ambiguity. However, rational deliberation also occurs when applying the rules of logic imperfectly or using heuristics such as rules of thumb or common sense. For example, one is engaged in rational deliberation when one generalizes rules of thumb or common sense, though one is well aware that neither rules of thumb nor common sense are always true. In addition, one uses rational deliberation when engaging in skills and practices, such as the practices of medicine and law, which have some rigid rules, as well as rules of thumb, but often require one to deliberate using casuistry. Casuistry involves trying to find reasons regarding particular cases that are often complex starting from paradigm cases, though those paradigm cases that are often more simple than the particular cases about which one needs to decide. Given the many imperfect forms of rational deliberation as well as the accidental misapplication of the rules of logic, then, one can engage in rational deliberation but reach the wrong answer. Finally, one can engage in

rational deliberation but ultimately have thoughts, emotions, or actions that are irrational, which I am defining narrowly as when one knowingly violates one's own internal standards of consistency.^{9, 10}

The Capacity to Value

I begin defining valuing by starting from Scanlon's book, *What We Owe to Each Other*:

We value many different kinds of things, including at least the following: objects and their properties (such as beauty), persons, skills and talents, states of character, actions, accomplishments, and ideals. To value something is to take oneself to have reasons for holding certain positive attitudes toward it and for acting in certain ways in regard to it. Exactly what these reasons are, and what actions and attitudes they support will be different in different cases. They generally include, as a common core, reasons for admiring the thing and for respecting it, although "respecting" can involve quite different things in different cases. Often, valuing something involves seeing reasons to preserve and protect it (as, for example, when I value a historic building); in other cases it involves reasons to be guided by the goals and standards that [an ideal, which I value] involves (as when I value loyalty); in some cases both may be involved (as when I value the U.S. Constitution).¹¹

I contend that the exercise of the capacity to value enables persons to have a wide range of attitudes, with greatly valuing some things at the positive end of this range, and strongly disvaluing other things at the negative end of this range. Altering Scanlon's description of valuing above, we get a definition of disvaluing:

We disvalue many different kinds of things, including at least the following: objects, persons, actions, character (such as maliciousness), failures, and misguided ideals.¹²To

⁹ This is based on Scanlon, though I offer some of my own examples of forms of reasoning. See Scanlon, *What We Owe to Each Other*, 25 – 30.

¹⁰ For example, a woman believes that the moral rules and circumstances dictate that she should treat a particular person with respect, but she knowingly fails to do so. She is irrational in this case, not because moral rules are objectively true and she fails to follow them, but because she herself accepts that x & y are the moral rules and z , namely, treating the person with respect is the implication of those rules, but she chooses not to act accordingly.

¹¹ Scanlon, *What We Owe to Each Other*, 95.

¹² I had some difficulty describing in general terms some of the things we disvalue in a completely neutral way, a way that does not already suggest a negative attitude regarding them. But this is no different than Scanlon's nice list of things that we value. For example, consider Scanlon's inclusion of

disvalue something is to take oneself to have reasons for holding certain negative attitudes towards it and for acting in certain ways in regard to it. Exactly what these reasons are, and what actions and attitudes they support will be different in different cases. In strong disvaluing, they generally include, as a common core, reasons for having negative attitudes such as contempt, disgust, and lack of respect (disrespect) for the thing, although “contempt,” “disgust,” and “disrespect” can involve quite different things in different cases. Often disvaluing something involves seeing reasons to undermine or destroy it (as, for example, when I disvalue a sharp pain in my leg); in other cases it involves reasons to reject the goals and standards that a misguided ideal that I disvalue involves (as when I disvalue moral egoism); in some cases both may be involved (as when I disvalue the ideology and plans of terrorists wanting to destroy the United States.).

When someone values or disvalues v , there is an attitude regarding, possible emotions surrounding, and possible actions taken regarding v . Consider two detailed examples of valuing, that of Drake and George.

Suppose Drake values an historic building. More specifically, he *treasures* the building. Drake could feel sadness or anger¹³ upon learning that the government allowed the historic building to be demolished, happiness upon learning that it will be restored, sadness upon learning that it was destroyed by a tornado, or fear upon learning that rising river waters might cause its ruin. His reasons for treasuring the building might also support the actions of working with others to challenge the government’s plans to demolish the building, voting or even campaigning for his local official who brought in funds to restore the building, and aiding in the efforts to find and collect any historically central parts of the building amongst its tornado-caused rubble. If Drake were to judge that he might save the building from flood waters, he would help volunteers form a sandbag barrier around the building.

accomplishments as something we value. To call something an accomplishment tends to suggest that its completion was good. E.g., we would not call the successful completion of a self-destructive course of action an accomplishment.

¹³ I am assuming that, by definition, anger comes from the judgment that someone will do or has done something wrong or that fate is unfair, while sadness comes from the judgment that something bad has happened or will happen, though that thing may not be caused by someone’s or fate’s “misdeeds”.

Suppose George values loyalty. More specifically, he respects loyalty in both himself and others. Supposing that George is loyal to three colleagues and assumes they are each loyal to him, George would feel anger upon learning that the first colleague betrayed him, happiness (and not, say, disappointment¹⁴) upon learning that the second colleague received an endowed chair, sadness (and not, say, schadenfreude) upon learning that the third colleague failed to get tenure. To George, loyalty, in part, means courage and performing actions such as giving support to whatever or whomever one is loyal, even when it is not easy to do so. In his valuing of loyalty, George may sometimes need to engage in rational deliberation about what his loyalties require in particular situations. In other words, since George values loyalty, often he finds reason to deliberate about exactly what he ought to do when someone or something to which he is loyal may require his sacrifice for their benefit. Say, after deliberating about the circumstances surrounding his colleague failing to get tenure, he finds reason to stand up for the colleague in an appeal regarding tenure. George finds reasons to take these steps despite his anxiety that, in taking these steps, he might lose some of his own political currency. However, in other occasions during which loyalty requires courage, George simply might see that his loyalty requires an act of courage and act accordingly, without deliberation. George, though, may have no occasion to act instantaneously on his loyalty. A classic example of simply *seeing* reasons for actions regarding something valued is when a Secret Service agent, loyal to the United States Presidency, shields the President

¹⁴ This assumes that his colleague getting the endowed chair does not undermine any other things that George values. If it does undermine other things he values, then whether George feels disappointment, happiness, or vacillates to some degree between pleasant and unpleasant emotions depends on exactly what else is affected by his colleague getting the endowed chair given the totality of what and how he values. Again, how valuing and disvaluing of various things affect one another is discussed more in Chapter 4.

from gun shots without deliberating at the time about whether to do so. This Secret Service agent simply sees that his loyalty requires that he physically shield the president and so acts.

Conclusion: Valuing and the Valuing Theory

On my theory of personal identity over time, it is a necessary and sufficient condition of personal identity over time that both what and how a human being values and disvalues as a particular person changes or stays the same as a result of the exercise of the capacity to value based on what and how that person valued previously. There are many complexities involved in changes that are person-preserving, such as the place of akrasia and moods. Discussion of akrasia and moods will occur in Chapter 6. Given the central role that memory has had in the debate regarding personal identity beginning with Locke himself, I will discuss the place of memory in the valuing theory in this project's conclusion.

The purpose of the remainder of this chapter is to present a level of detail regarding the nature of valuing which will aid my forthcoming discussion of personal identity at and over time in Chapters 4 and 5. In the next section I discuss aspects found in the valuing or disvaluing of (almost) anything. Then in the subsequent section, I discuss optional ways that one may value or disvalue something, such as valuing something intrinsically or extrinsically.

The Products of Valuing

Charge: Valuing, Disvaluing, and Neutrality

The exercise of the capacity to value involves having attitudes with what I will call a 'charge'. The possible charges of valuing attitudes are positive, negative, and neutral.

In valuing or disvaluing something, call that thing v , one sees or finds reasons to have one or more specific attitudes towards v , d , m , or n , each various things that one can value. If one values v , then one's exercise of the capacity to value produced an attitude that has a positive charge, such as admiring or respect v . If one disvalues d , then one's exercise of the capacity to value produced an attitude that has a negative charge, such as having disdain or contempt for d . On my analysis of the capacity to value, one cannot value v or disvalue d and yet lack any possible emotions regarding v or d , respectively. Moreover, given that one values v and disvalues d , one likely has any number of potential actions that one would take regarding v and regarding d , respectively. The role of emotions and actions in valuing and disvaluing is discussed shortly.

Besides valuing and disvaluing, there is neutrality. Given the various other things that one values or disvalues, one can have a neutral attitude regarding n , which I will call 'neutrality' as well as 'being neutral towards'. Neutrality towards n , which represents the third and last possible charge, mostly consists of a lack of emotions regarding n and a lack of action regarding n , though sometimes neutrality can involve emotions or actions regarding n . Neutrality is an intermediate between some of one's valuing and disvaluing. Neutrality towards n only expresses emotions or entails actions when n is in some

relation to other things which one values or disvalues. E.g., if one is neutral towards n , one would be happy that n benefited instead of d , but sad if n benefitted instead of v .¹⁵

Finally, it seems plausible to me that for everything that a person encounters or imagines, some of those are things he neither values, disvalues, nor is neutral towards. For some things, rather than having neutrality towards it, one may not exercise one's capacity to value in regard to it. I will call this "the non-valuing of" m .¹⁶

To VaND & VND attitudes

Sometimes for the purposes of discussion, it does not matter which exact attitude is had regarding v . Instead, within that discussion, it only matters that the exercise of the capacity to value has produced one of the three types of attitudes regarding v . When the exact charge of the attitude does not matter, in order to avoid the awkwardness of and bad parallelism of stating the three possible charges – valuing, disvaluing, and neutrality towards – I am coining a new verb, VaND, pronounced /vænd/ – rhymes with 'land'. To VaND v is to value, disvalue or have a neutral stance regarding v , with V standing for valuing, N for neutrality, and D for disvaluing. We can say one or more persons VaND v

¹⁵ On my analysis of valuing, the idea of seeing and finding reasons to take a neutral attitude towards *everything* and thus, have no possible emotions and actions regarding anything makes no sense. I can imagine a healthy human being who is also a person whose all-things-considered judgment is that everything is neutral, but I cannot imagine this individual actually neither seeing nor finding any reasons (though, perhaps, no all-things-considered reasons) to value or disvalue some things. I cannot fully defend this claim until the end of the project. I will take up the issue of universal neutrality in the last chapter. Also see next footnote.

¹⁶ Similarly to neutrality towards everything, it does make sense to me that a person could have, in regard to everything, no judgments which would support valuing, disvaluing, or neutrality. I hold that this would be a non-person. As with universal neutrality, I cannot fully defend this claim until the end of the project. I will take up the issue of universal neutrality in the last chapter.

As the project progresses, I will provide arguments for my claims that complete neutrality or complete non-valuing is impossible for persons. This connects to why I hold that the capacity to value, rather than simply the capacity to reason, is definitive of personhood.

or VaNDed v or is VaNDing v or, passively, that v is VaNDed or v was VaNDed.

Sometimes as an alternative to VaND in order to express that v is valued, disvalued, or viewed neutrally, I might use phrases such as “the VND attitude towards v ” or “ v is something towards which there is a VND attitude”.¹⁷

Reasons for both Emotions and Actions

The reasons which partially constitute VaNDing are like any other reasons in that a person can find them through deliberation or simply see them as reasons. Suppose one VaNDs v . Take any circumstance that one sees as regarding v such that one will both have emotion(s) regarding v and take action(s) regarding v . Reasons justify, to one’s self, one’s attitude towards v . Those same reasons partly constitute the emotions one will have regarding v and partly justify, to one’s self, the actions one takes regarding v .

Consider Drake’s treasuring of the building, fearing for the building, and helping to try to save it from floodwaters. In that situation, Drake’s fear for the building consisted of his reasons for treasuring the building along with the judgment that the building was in jeopardy from floodwaters. Moreover, his actions to help save the building were partly justified by the reasons that he treasures the building. If he only judged that the building was in jeopardy, that judgment alone would neither invoke fear in him nor lead him to try to save the building. Instead, his judgment that there are reasons to treasure the building both partly constitute his fear for the building and partly constitute his motivation to act to try to save the building.

¹⁷ I find it displeasing to have to create a new verb. However, it seems to me to be the lesser of evils. There are times when I need to speak in the passive voice, and I think that phrases such as “ v is something towards which there is a VND attitude” are more unpleasant to read than “VaNDing v ”.

Possible Emotions Necessarily, Possible Actions Likely

George is sad upon learning that his colleague failed to get tenure because he values loyalty, but he does not always have some emotional expression of his valuing of loyalty. Moreover, George will not always be in a situation that he thinks requires him to act from loyalty.

When the exercise of the capacity to value leads someone to VaNDs v , he will have emotional reactions in various situations surrounding v . If one VaNDs v , rather than always experiencing emotions or always acting in regard to v , one necessarily has an emotion-stance regarding v and one almost always has an action-stance regarding v .

To have an emotion-stance regarding v entails one will have one or more emotions in one or more circumstances regarding v . Similarly, to have an action-stance regarding v entails that one will take one or more actions in one or more circumstances regarding v . The set of circumstances which will lead one to have emotions regarding v and the set of circumstances which will lead one to act regarding v may be completely the same, overlap to some degree, or be completely different. When one of these circumstances occurs, then one's VaNDing of v is expressed. Expressing one's VaNDing of v means that one's current emotions or actions at least partly reflect one's VaNDing of v .

The fact that the emotions of VaNDing v are not always expressed can help account for why if one greatly values v one rarely experiences positive emotions regarding v . Perhaps one takes v for granted and so does not experience positive emotions regarding having v in one's life. However, we can say that one values v rather than one simply thinks that v is valuable if one would be deeply saddened by the loss of

v .¹⁸ Generally, even though one VaNDs v , one may never react emotionally to the presence or thought of v , and one may never see or find any reason to do anything regarding v , because circumstances never arise for the expression of VaNDing v in emotion or action. Though no circumstances may arise in which one's VaNDing of v will be expressed, VaNDing v necessarily involves an emotion-stance, and almost always involves an action-stance.¹⁹

Emotions may facilitate actions

Suppose one VaNDs v . There is often a special relationship between emotions and actions in a particular circumstance in which both emotions and actions expresses one's VaNDing of v . This special relationship is one in which the emotions regarding v facilitate the movements of the actions that one will take regarding v . The bodily sensations of fear such as heightened awareness, faster heartbeat, and pent-up energy caused by a release of adrenaline would enable Drake to help bag sand many hours past what would normally be his limits. In general, when one VaNDs v and a circumstance arises when one expresses that in both emotion and action, one's emotions may help facilitate one's actions regarding v .

¹⁸ This agrees with Martha Nussbaum's position that emotions point to what we value. See Martha Nussbaum, "Emotions as Judgments of Value and Importance" in *Thinking about Feeling: Contemporary Philosophers on Emotions*, 183 – 199 (Oxford, UK: Oxford University Press, 2004).

¹⁹ It would be rare that the valuing of v would not include one or more possible actions in regard to v . It is difficult to think of an example when there is valuing of v but that valuing includes no possible actions regarding v . The following could serve as such an example: Holly admires God, understood as an omniscient being in every possible sense. Though Holly admires God, it may be impossible for her to ever judge that there is a situation requiring actions regarding God. This would be tied to a number of beliefs: Sally thinks God is omniscient, so God does not need her to make choices to help bring about what God wants in the world; Holly thinks God does not need her to praise God at all; and Holly thinks that what she ought to do is solely determined by what is right, not by what God thinks she ought to do (though, of course, God wants her to do what is right).

Subjective but not private

There may be objective normative facts about what one ought to VaND. Even if there are objective facts about what one ought to VaND, what one VaNDs may or may not correspond to those objective facts. VaNDing is always a subjective activity and experience.

Even though this activity is subjective, this does not mean that others cannot recognize what and how one VaNDs, sometimes in fact better than one does. Tim may think he values loyalty, but others might recognize that Tim is wrong in this self-assessment. Others may see that the most accurate description of Tim's attitude surrounding loyalty is that Tim values loyalty in others *towards him*.

Degrees of Awareness

Not even the capacity to reason alone is always engaged on a fully conscious level. One can deliberate without full attention being paid to the fact that one is doing so – though if another pointed while one is deliberating, one will most likely agree that one is deliberating. If one is forgetful, though, after deliberation, one may not remember that one has deliberated even if another tries to remind one that one did so. Similarly, one can VaND v without being completely aware of so VaNDing, and one may be even less aware of what emotions or actions one may take regarding v . Even if one recognizes that one VaNDs v , one is unlikely to know *all* of the different circumstances which would lead one to express one's VaNDing of v or know exactly how those expressions would go.

Conclusion

This section discussed the necessary features of valuing. For anything that one VaNDs, call that v , (1) there is an attitude towards v with a charge, (2) reasons for the attitude which are also involved in the reasons for the possible emotion and possible actions that one might take regarding v , (3) there is necessarily an emotion-stance and likely an action-stance regarding v , (4) VaNDing v is subjective but not private and (5) one can do so with various degrees of awareness of one's VaNDing of v .

Different ways of VaNDing v

This final section discusses the various ways that one may VaND something.

One can VaND something in part or as a whole. When VaNDing some v in part, often the reason that the person VaNDs v at all is what we or the person herself would use to describe this VaNDing of v , as in, “Theresa values Henry *for his looks*.” One may VaND something as a whole when one VaNDs it for a property that one finds as reason to VaND with one charge, despite other properties it possesses that one generally VaNDs with the opposing charge. Ideal parental love may be an example of loving wholly: A person loves her child perhaps because that child is her own and despite qualities that child possesses which she disvalues.²⁰

One can both value and disvalue some v . This is at least partly because one can value some v in part, so one could value v for one part and disvalue v for another part. An unwilling cigarette addict might value cigarettes for the psychological and physical

²⁰ Perhaps if someone values some v for v 's essential properties, which, by the definition of “essential properties”, are properties that could not change without destroying v itself, then one wholly values v . Henry values being a bachelor wholly because a bachelor is an unmarried man. A Kantian respects persons wholly because persons have the capacity to reason.

equilibrium they restore. At the same time, she could disvalue cigarettes both as carcinogens and for their ability to create an artificial need by forming a psychological and physical disequilibrium – cravings.

Alternatively, a different unwilling cigarette addict might disvalue the cigarettes while valuing some of their traits. One can value (or disvalue) v , in spite of disvaluing (or valuing) one or more of v 's traits. One might value charisma, but greatly disvalue Hitler, though he may have been charismatic. One might disvalue spitefulness but value someone as a friend despite her spiteful streak. Valuing (or disvaluing) v , though v has traits that one disvalues (or values), is an example of a form of VaNDing v that can be distinct from both VaNDing v in whole or in part, or from both valuing and disvaluing v simultaneously.

For two or more things VaNDed, some things one might VaND some more than others while VaNDing some of those things equally. I value my vacation time on the beach more than drinking my tenth glass of water for the day, but I value each of my pairs of sunglasses equally. However, the valuing of some things might be incommensurate with the valuing of other things: Holly, the mother of two children, values each of her two children greatly. It is not true that she values one more than the other, or even that she values them equally. Instead, she values each child greatly, but differently.²¹ Similar ideas apply to disvaluing various things, one can disvalues some things more or less, others equally, and yet others incommensurately.

One might value some things no matter the situation, but value other (if not most) things only given some range of conditions. For example, in a situation of extreme thirst

²¹ There are, of course, both mothers who do in fact value one child more than another and mothers who value all of their children equally.

or in which I accidentally imbibe toxic chemicals, I might value my twentieth glass of water *and* value it over the beach vacation.

One might VaND something intrinsically, in and of itself, or extrinsically, for what use it has, or both. Holly could extrinsically value the family computer because of the homework it enables her children to complete, while valuing her children intrinsically.

Conclusion

In this last section, I briefly describe many possible features of VaNDing, and my goal was not to be certain to introduce *all* of the features of VaNDing but to try to show that there are many such features. I likely did not list all the features of how persons can VaND. Many of these features of VaNDing are reflected in the reasons one finds or sees as well as in the emotions one expresses in particular situations which one sees as related to something one values, e.g., one being frustrated that another glass of water is unavailable despite extreme thirst. These features are some of the many of the ways that one can value something. Having introduced these features will enable me to better describe how changes in what and how a human being VaNDs can be either person-preserving or person-destroying.

Conclusion

I hold that a human being remains the same person so long as what and how she values changes through use of the capacity to value and this change is based on what and how the human being valued previously. This chapter described the nature of the capacity to valuing, the exercise of which produces one's VaNDing of various things.

Exercise of the capacity to value involves the capacities to reason and to have emotions.²²

Exercise of the capacity to value, like exercise of the capacity to reason, can occur with less than full awareness. VaNDing v involves seeing or finding reasons to take an attitude regarding v , which necessarily includes an emotion-stance and likely include an action-stance regarding v , and thereby may lead to particular emotions and actions regarding v . The next chapter discusses how the VaNDing of various things creates synchronic unity of a person.

²² Though believing often involves reasoning and emotions, believing should be distinguished from valuing, though valuing is connected to believing. There will be both the belief in the truth of P which involves feelings of conviction regarding P , and, perhaps a VaNDing that P , which involves judgments to take a VND attitude regarding the fact that P is true, e.g., “I despise that P .” Moreover, the fact that one believes that P is also revealed in that fact that one despises that P – one takes P to be true and one does not simply think, “I would despise that P , if P were true,” but rather one actually despises that P because one takes P to be true.

CHAPTER 4 – VALUING AND THE SYNCHRONIC UNITY OF A PERSON

Introduction

This chapter focuses on explaining how an analysis of persons, based on the capacity to value, can provide a satisfactory analysis of the synchronic unity of persons, personal identity at a time. Parfit, Korsgaard, Rovane and Schechtman each provide problematic analyses of synchronic unity.

Parfit's discussion of persons does not analyze what unifies a person at a time.¹ Instead, Parfit seems to borrow from our pre-theoretical sense of what counts as unified persons. He asks us to consider whether X today is Y of some future point in time but never makes it clear how we distinguish X from all the other persons out there in the world today and Y from all the other persons out there in the world at Y's point in time. He appears to be simply assuming that there is, at most, one person within each human being. This would enable us to separate X from other persons during X's time and separate Y from other persons during Y's time. Parfit is assuming that persons are

¹ Parfit's discussion of divided minds (Parfit, *Reasons and Persons*, Chapter 12) suggests that he holds that unity of consciousness might be a necessary and sufficient condition for unity of a person on any given day. (Also see Parfit, where he states that there "are two unities to be explained: the unity of consciousness at any time, and the unity of a whole life." (Parfit, *Reasons and Persons*, p. 217)) However, given his inclusion of all psychological features as relevant in defining psychological connectedness and continuity, consciousness cannot be a necessary condition for the unity of a person: *all* of the psychological features of everyday persons are not *always* conscious. E.g., a person likes ice cream, but today he doesn't think about ice cream. His failure to think of it today does not mean that liking ice cream was not one of the person's psychological features today. Or, a person could hate his brother, but, being in denial about this attitude towards his brother, hatred of his brother never enters his conscious thoughts. (Brother example from Schechtman, *Constitution of Selves*, 115, though she is using the example for other purposes.) These examples show that unity of consciousness cannot be a necessary condition for the unity of a person.

human-sized. This assumption seems especially problematic for Parfit, since Parfit considers fantastical thought experiments to reveal important features about persons. I would contend that the coherence of tales of persons sharing one body, such as Robert Lewis Stevenson's tale about Dr. Jekyll and Mr. Hyde,² reveal that our understanding of persons allows for multiple persons possibly co-existing within a single human being.³ If it is coherent that two or more persons could exist at a time and persist over the same time within a single human being, then an analysis of persons at and over time must be such that it does not rely on the problematic assumption that a given human being is at most one person.

In Korsgaard's early thoughts on personal identity, she assumes that it is the having of one human body and one life that puts pressure on each human being to unify as one person at and over time.⁴ While having only one body and one life would seem to

² A less refined example is the 1984 movie, *All of Me*, a comedy in which a man and woman unhappily share one body due to a magic spell.

³ Stevenson was not a philosopher. The claim that our pre-theoretical concept of person allows for the possibility of multiple persons co-existing within one human being may get *greater* support from Stevenson's work than from the typical philosopher's fantastical thought experiment.

Several philosophers** have questioned the validity of using fantastical thought experiments to analyze our concepts. For example, Richard Wollheim argues that philosophers often under-develop those thought experiments in a way that supports their theory and unfairly – though perhaps unintentionally – biases the audience against alternative theories. (Richard Wollheim, *The Thread of Life*, (Cambridge, MA: Harvard University Press, 1984), 21 - 30)

In contrast to forming a thought experiment for philosophical purposes, Stevenson was simply writing a work of fiction that he likely assumed would resonate with readers – and it did. This suggests that the idea of more than one person within one human being is within the limits of our concept of 'person'.

** For others who question the use of fantastical thought experiments to do conceptual analysis, see Ludwig Wittgenstein (Ludwig Wittgenstein, *Zettel*, ed. By G. Anscombe and G. von Wright, and translated by G. Anscombe, Blackwell, 1967, Proposition 350); W. V. Quine (Review of Milton K. Munitz, ed., *Identity and Individualization*, *The Journal of Philosophy*, Vol. 69 (1972), 488 – 491, 490; Wilkes, *Real People*, Chapter 1. ⁴ See the section on Korsgaard in Chapter 1 of this project. Rovane also notes this assumption in Korsgaard's essay. Rovane, *The Bounds of Agency*, 161 footnote.

⁴ See the section on Korsgaard in Chapter 1 of this project. Rovane also notes this assumption in Korsgaard's essay. Rovane, *The Bounds of Agency*, 161 footnote.

put pressure on human beings, she ignores the fact that such pressures may not always be enough to unify a human as one person. There is clearly no practical necessity in unifying within one human being as one person in the same way that, as Korsgaard argues, there is a practical necessity in deciding what to do and be over and above one's competing desires, plans, and so forth. If some desires and plans of a human being are *not* those of a particular person within that human being, then that person would not be compelled to be something above those "foreign" desires and plans as that person decides what to do. This is similar to there being no necessity in my being something above both my own desires and projects *and* my neighbors' desires and projects as I decide what to do.

Rovane's analysis of synchronic unity depends solely upon the capacity to reason specifically striving for rational unity. A unified person emerges from trying to coordinate various projects and striving for overall rational unity. I agree that part of a person's unification does seem tied to striving for overall rational unity – trying to get one's projects to cohere together within one's constraints and often modifying them for that reason. However, sometimes overall rational unity in the coordination between projects is not possible, there is the recognition that these projects – at least in their present form – cannot be coherently pursued, and yet those conflicting projects continue to be pursued without self-criticism based on one's failure to meet the rational ideals of coherence.

Recall Mary, a human being that has the project of being a good businesswoman and the project of being a good mother such that these two projects do not go together, but she continues to have those conflicting projects. Rather than striving for overall

rational unity, she is committed to those projects instead of overall rational unity.

Rovane's analysis would entail that she is not one person.

In contrast, I hold that Mary is one person, though she rejects striving for overall rational unity and prefers being irrational⁵ in continuing to have both of these projects. On the valuing theory, Mary would be one person if she values both projects, which includes emotion-stances towards both projects. If Mary values both projects, while Mary is unwilling to adjust her projects to make those projects rationally cohere, she is one person whose unity is partly revealed in her ongoing misery – she is regularly frustrated, sad, depressed, and sometimes self-berating, when her job suffers from her time spent with her children or when her children suffer from her long hours at the office. Mary's valuing of both projects is seen when she has negative emotions because one project is getting short-changed *while she is doing the other project*. The valuing of both is what makes Mary one human-sized person with these two projects. Despite her forgoing overall rational unity for the sake of keeping her two projects, Mary is not composed of two persons, one with each project.

In contrast to Schechtman, I hold that Mary surely can be unified if she lacks a self-narrative in which she understands herself as the businesswoman and mother, though if she had such a self-narrative and it were in agreement with reality as Schechtman requires, I think that having such a self-narrative *could* indicate that Mary is one person.

However, if Mary lacked a self-narrative in which she saw herself as a businesswoman

⁵ Irrationality, narrowly construed, is irrationality based on Mary's own standards of what is rational and irrational. See definition of irrationality under the discussion of the capacity to reason, in Chapter 3. Moreover, as I noted when I introduced this example, Mary in the future, or someone else could view being a CEO and mother as compatible. The problem is: how is Mary a unified person if she conceives of these projects in an incompatible way – having both would require 36 hours a day – and so forgoes rather than strives for overall rational unity between these projects. For this full discussion, see Chapter 1, under critique of Rovane.

and a mother, her unity as a single person would again be revealed in her misery when these projects conflicted. Also in contrast to Schechtman, I would not assume that, if Mary, the human being, had a narrative then that narrative would be accurate because it only covered experiences had and actions performed by Mary. Instead, I would look to see whether Mary's projects were connected to one another on a rational or emotional level. Generally, if some of a human being's projects were not connected to the rest of what is VaNDed by that human being, then those projects unconnected to the rest would belong to someone else within that human being.⁶

A Glimpse of Synchronic Unity through VaNDing

This chapter explains how the capacity to value produces synchronic unity of persons. This analysis takes seriously the fact that it is not theoretically impossible for two or more persons to coexist within the same human being, and that persons, who are such in part due to the capacity to reason, are nevertheless often irrational.

Suppose we could gain perfect knowledge of other persons and we are proponents of the valuing theory. We have before us a given human being that VaNDs various things, including some projects. It is clear that this human being has personhood, but we are not going to assume that this human being is only one person. We've learned that this human being VaNDs hundreds of different things – how can we discern whether the VaNDing of each thing belong to only one person or several persons with that human being? We look for connections between these things VaNDed.

⁶ Suppose there were several projects not connected to an interconnected set of projects. Call them the loose projects. Whether the loose projects belong to only one or more than one persons would depend on whether there were connections between these other projects.

We are going to find that many connections between the things VaNDed that are based on reasons. An example of this connection based on reasons is between Holly's valuing of family and the valuing of each of her children – she values her children partly because she values family.⁷ Other connections between VaNDed things are going to be emotional, such as Mary's intermittent misery when her businesswoman project must suffer while she is working on her motherhood project and vice versa. If we find that with this human being each thing VaNDed has connections to the other things VaNDed, such that everything is connected either directly or indirectly, then all of these things VaNDed are those of one person.

This kind of a survey is not something that we, given our own epistemological limitations, could do in the real world with real persons⁸, but this hypothetical survey illustrates what, on the valuing theory, unifies person *and* defines a person. Those things VaNDed and connections between those things VaNDed within a human being provide for synchronic unity of a person, and it defined who that person is at a point in time.

⁷ Holly is an example from the Introduction to Section II.

⁸ In the real world, as a pragmatic matter, we are more likely to assume that a human being with personhood is unified as one person unless clues surface that indicate that though a human being VaNDs various things, the VaNDing of this human being is irregular. Unlike Mary, another human being, Martha, might have the projects of motherhood and businesswoman but Martha always seems to be tripping over herself. (In contrast, Mary is very organized.) One minute Martha is working on her business proposal and the next she wonders what she is doing and goes to care for her child. There is no coordination, no striving for rational unity, no emotions regarding one project while she is working on the other. Instead, there are times when she will say she does not care about being a businesswoman, she is not a businesswoman, and there are other times when she will say that she does not care about being a mother, she is not a mother. The self-narrative never includes both projects. Given the evidence, we might begin to suspect that there are two persons within Martha.

This project focuses on in what the unity of a person actually consists rather than how we would discern whether someone is the same person. How best to discern whether a given human being is unified at or over time is a question that I need to address in order to be able to make the valuing theory something that one can readily apply. However, this project focuses on formulating the valuing theory itself. This project does not provide a full analysis of how to apply the theory.

In this chapter, I will present a picture of synchronic unity based in the general exercise of the capacity to value, which results in VaNDing. Diachronic unity and the persistence of persons over time will then be built upon synchronic unity. Synchronic unity consists of (1) the things that one VaNDs as well as (2) the influences between VaNDed things. For some of what one VaNDs, the VND attitudes regarding those things are determined by one or more *other* things that one VaNDs. For example, today Joy values her good health and she sees aerobic fitness as partly constitutive of good health and this is why she values her aerobic fitness.

The way one VaNDs anything is interdependent on some or all of what else and how else one VaNDs. Recall that *how* one VaNDs v is defined by the reasons one has for the VND attitude towards v . The reasons for that VND attitude support an emotion-stance towards v , which will lead to emotions partly constituted by those reasons. The reasons for that VND attitude towards v also likely supports an action-stance towards v , which will lead to actions regarding v that are partly justified by those reasons. However, what reasons one sees or finds regarding v , what emotions one will experience regarding v , and what actions one will take regarding v , depend on what else one VaNDs. Which emotions one will have depends not just on how v is affected but on how any other things that are VaNDed are affected. Which actions one will take regarding v also depends on the impact of those actions to the other things that one VaNDs. Reconsider Drake, who was first introduced in Chapter 3 as an example of someone who treasures something.

In a complex person, treasuring v is not simply a matter of being angry whenever v is willfully destroyed and defending v whenever v is in danger. For example, Drake treasures an historic building. However, there are situations in which Drake will be

pleased, not angry, that the government will destroy the building, and Drake will not aid the building though it is in danger (of being destroyed by the government) even though Drake sees or finds the same reasons, in this situation, that he found before to treasure the building. What makes Drake pleased rather than angry with the government for planning to destroy the building and what prevents him from trying to stop the building's destruction is that, recently, the building has become a danger to other things that Drake values. In this particular situation, the historic building suffered an earthquake, and the government must destroy it to protect people from the building collapsing on them and their homes. Drake still treasures the building, but his lack of anger with the government and his unwillingness to aid the building are exceptions to the paradigmatic version of treasuring, e.g., being angry at the willing destruction of the treasured v and coming to v 's aid when it is in danger.⁹ However, his lack of anger and unwillingness to aid in this situation is a part of how Drake treasures the building: he does not treasure it so much and so deeply that he wants others put at risk for the building's preservation especially without their consent. In general, if one VaNDs more than one thing, then much of one's VaNDing involves many non-paradigmatic versions of the attitudes of valuing or

⁹ In saying that there are paradigmatic versions of valuing and disvaluing attitudes, I do not mean to suggest that there is not a range of options which fall within any particular paradigmatic valuing or disvaluing attitude. In defining direct valuing connections, I am actually more concerned about explaining what emotions or actions seem to clearly fall outside of paradigmatic versions of valuing or disvaluing attitudes. The only way to explain why Drake does not express something like anger and does not at least consider coming to the aid of the building is by looking to the rest of what and how he values and disvalue, and seeing that at least one thing he values, disvalues, or views neutrally modifies his treasuring of the building and exactly how that treasuring is expressed.

In suggesting paradigmatic versions of various valuing and disvaluing attitudes, I also am not trying to defend any claims about what generally falls under paradigmatic treasuring or paradigmatic versions of any other attitude, though my examples clearly suggest what I suspect falls within the paradigms of various attitudes.

disvaluing.¹⁰ This results in a type of interconnectedness in the what and the how of one's VaNDing.

I contend that the kind of interconnectedness seen in the Joy example and the kind of interconnectedness seen in the Drake example together constitute the synchronic unity of a person. This chapter discusses the nature of this interdependency and the synchronic unity that results.

Direct valuing connections

How Drake treasures the building is tempered by how Drake values the lives of others. When how one thing is VaNDed affects or is affected by how another thing is VaNDed, there is what I will call a **direct valuing connection** (sometimes shortened to “**direct connection**”) between these things that are VaNDed. There can be more than one direct valuing connection between two things that are VaNDed.

There are at least two general types of direct valuing connections, reason-based and event-based. Between some v and w , each one VaNDed, **direct reason-based valuing connections** (sometimes shortened to “**reason-based connections**”) occur when one or more of the reasons supporting that VaNDing of v are dependent upon one or more of the reasons for VaNDing w , or vice-versa. Between some v and w , each VaNDed, a **direct event-based valuing connection** (sometimes shortened to “**event-based connections**”) occur when an event creates one or more clashes or synergies between v and w . These clashes or synergies can occur in the reason, emotion, or action aspects of the VaNDing of v or w .

¹⁰ I am uncertain that there is a paradigmatic version of a neutral stance regarding something, since I hold that neutrality is parasitic on the valuing or disvaluing of other things.

Direct reason-based valuing connections

Direct reason-based valuing connections include the dependence of extrinsic VaNDing of v on the VaNDing of something else. More specifically, if v is valued solely for the extrinsic reason that it promotes w , then the reasons for valuing w are the reasons for valuing v . Suppose Joy values physical flexibility because she understands it to be part of good health. Suppose she values good health solely because it enables her to spend time with and provide support to her friends and family. Suppose further that she values yoga solely for its ability to promote flexibility. This means that Joy's reasons for her valuing of yoga *are* that it helps to enable her to spend time with and provide support to her friends and family. In other words, the reasons for the valuing of yoga are the reasons for her valuing of flexibility and good health. Similarly, something can be disvalued on the extrinsic grounds that it undermines something else that one values, or undermines something that one values. E.g., Joy disvalues trans-fats solely extrinsically because trans-fats undermine health, and so she disvalues trans-fats solely because she believes that trans-fats may decrease the time that she will have to spend with and provide support to her friends and family.

In addition, direct reason-based valuing connections also come in the form of mutually grounded VaNDing. Mutually grounded VaNDing occurs when v is VaNDed because of its relation to one or more other things, w and x , each which are VaNDed as well, but the VaNDing of w or x is based on their relation to v . One could value good communication in which individuals come to greater understanding of one another in part because one values close friendships. At the same time, one can value close friendships in part because close friendships involve good communication. Similarly, one could

disvalue rudeness because it promotes hostility, which is disvalued in part because it often involves rudeness.

Direct reason-based valuing connections also come in the form of shared grounded VaNDing. Shared grounded VaNDing occurs, for instance, when valuing v is linked both to valuing w and to disvaluing both x and y because v and w are seen as mutually promoting, while x and y are seen as undermining both v and w . One may value political freedom because one sees it as a human right. One then values freedom of speech because one holds that political freedom is partly constituted by freedom of speech, which means that the reasons to value freedom of speech include the reasons to value political freedom. In contrast, one may disvalue government propaganda because one holds that government propaganda will undermine political freedom. The reason one has to value political freedom, namely that it is a human right, is mirrored in the reasons one has to disvalue government propaganda, which one sees as undermining a human right.

Conclusion

There may be other forms of direct reason-based valuing connections besides those in extrinsic VaNDing, mutually grounded VaNDing, and shared grounded VaNDing. I am not trying to present an exhaustive list of the types of direct reason-based valuing connections. The purpose of this subsection on direct reason-based valuing connections is to describe the nature of direct reason-based valuing connections as well as illustrate the nature of these connections by describing a few of types of these connections.

While I do not hope to provide an exhaustive list of all forms of direct reason-based valuing connections, I want to make clear one way in which direct reason-based valuing connections do *not* occur. Direct reason-based valuing connections are *not* formed simply from the fact that the VND attitude for v and the VND attitude for w stem from what might be called “the same reasons”. One may value both a painting and a vase because each “is beautiful.” Nevertheless, how one values the painting and how one values the vase have no direct reason-based valuing connection, if one’s valuing of the painting because it is beautiful does not somehow depend on one’s valuing the vase because it is beautiful or vice versa. Note that both the valuing of the painting and the valuing of the vase could each be reason-based valuing connected to valuing beauty. However the direct valuing connection would then be between (1) the valuing of the painting and the valuing of beauty and (2) the valuing of the vase and the valuing of beauty.¹¹ (1) and (2) would still add no *direct* valuing connection between the painting and the vase. A direct valuing connection exists when how one thing is VaNDed is dependent upon how some other thing is VaNDed. A direct reason-based valuing connection exists when the reasons for VaNDing v partly depend on the reasons for VaNDing w , or vice versa. The mere fact that two things are VaNDed for “the same reasons” does not form a direct valuing connection.

¹¹ This may be an example of one of the aforementioned reason-based valuing connections, though perhaps it suggests another type of reason-based valuing connection, something to do with valuing or disvaluing of instantiations of ideals valued or disvalued. This would seem to be a case of when the instantiations of an ideal are valued or disvalued for reasons the ideal is valued or disvalued.

Direct Event-based Valuing connections: Clashes and Synergies

In addition to direct reason-based valuing connections, events can form direct valuing connections between two or more things that are VaNDed. These are event-based connections. An event forms a direct valuing connection between v and w , both VaNDed, by creating non-paradigmatic emotion-stances or action-stances towards v or w .

Recall the example of Drake and his treasuring of the historic building. Drake's treasuring of the historic building is an example of non-paradigmatic treasuring. Because he values other persons as well as the building, Drake's treasuring of the building does not always involve being angry at those who would destroy what he treasures or taking action to preventing its destruction. Instead, Drake is pleased with the government and its plans to demolish the building in order to protect the public.¹²

Events: Subjectively & Broadly Defined.

In a direct event-based valuing connection, an event is an actual or possible state of affairs. An event is whatever one sees or understands as a state of affairs that significantly affects one or more things VaNDs. The "event" of a direct event-based valuing connection is defined subjectively. Consider Drake and Paul, who both know that the building was demolished. Drake, though not angry about or willing to stop the demolition of the building, is saddened by the demolition of the building. The demolition of the building is an event for Drake: it is something that he sees as affecting something he values. Another person, Paul, knows that the building was demolished though he does not know why it was demolished. Moreover, Paul had no standing VND attitude towards

¹² It is still treasuring of the building, though, non-paradigmatic: Drake is saddened at the building's destruction, wishing that it could have been saved while keeping the public safe.

the building itself. Seeing or finding no reasons to take a VND attitude towards the building, the building's demolition is not an event for Paul. For Paul, the demolition is like countless changes in states of affairs each and every day that mean nothing to a person.

An event is *not* some actual or possible state of affairs objectively or inter-subjectively understood. So what someone might call "the same event or situation" will only sometimes be an event on my technical use of the term, "event". Consider the situation of "driving the last several meters up to your home". Consider (a) seeing your first home for the first time when you decided to buy it, (b) seeing it for the 5,236th time, and then (c) seeing it the 5,237th time burning to the ground.¹³ The first time, it is an event: that first time is when you came to value it or value it even more. The 5237th time was an event: you are watching the end of the house that you love. The 5,236th time and most other drives up to your home, will likely involve something slightly different from the other times driving up to your home and so there is a new state of affairs, e.g., the shrubs have grown since yesterday. However, that 5,356th time and many other times driving up to your home do not qualify as events for you, a state of affairs that you notice *and* that you see or find as significantly affecting something you VaND.

An event is an actual or possible state of affairs, which may or may not be a state of affairs *that one reasons to be relevantly caused by one's actions*. What can count as an event includes, but is not limited to: (a) one's actions, (b) what may occur through no causal powers of one's own, as well as (c) combinations of one's actions and their

¹³ Reference needed, but unknown at this time. I read an example about seeing your home for the first time verses seeing it for the *n*th time, where the *n*th was uneventful, but I cannot recall where I read that example.

consequences. A direct event-based valuing connection is formed between things each which are VaNDed and that connection is formed by an actual or possible state of affairs rather than by some type of dependency between the reasons of the things VaNDed as with reason-based valuing connections. A direct event-based valuing connection exists between two or more things VaNDed when some event or possible event creates clashes or synergies between the two or more things VaNDed.

Direct event-based valuing connections are ubiquitous. When one finds that one thing treasured ought to be harmed for another thing treasured; e.g., when morality (which one values) requires that one must make a sacrifice, this is a direct event-based valuing connection in the form of a valuing clash. Whenever one can promote many things treasured with one action, or whenever many things treasured are undermined by the same event, one experience direct event-based valuing connections in the form of valuing synergies.

Clashes

Direct event-based valuing clashes often bring about emotional turmoil as well as turmoil about what actions to take. Suppose v and w are valued, while x and y are disvalued. A clash can occur when an event causes (1) benefit to v or w but benefit to x or y as well, or (2) harm to v or w and harm to x or y as well. The clash can be emotional: one has mixed feelings surrounding the event, such as being both pleased and disappointed that the event occurred – pleased how the event affected some things valued but disappointed in how it affected other things valued, feeling satisfaction as well as discomfort with one's own actions.

In addition, the clash can take the form of indecisiveness. This can be indecisiveness about what to do when more than one of one's considered options seems to be good in some ways and bad in others. This indecisiveness can also take the form of indecisiveness about how to VaND the event itself: given how it both benefits things one values and harms things one disvalues, one might be unable to find satisfactory reasons to take a decisive VND attitude towards the event.

Synergies

Synergy occurs when an event compounds things in either a positive or negative way. The more an event benefits or promotes valued things or harms and destroys disvalued things, the more positive the synergy it causes. The more an event benefits or promotes disvalued things which also harm or destroy valued things, the more negative the synergy it causes. In contrast to the mixed emotions and indecisiveness that a clash event can bring, a synergy event can lead to heightened emotions and easier decision-making – heightened or easier in comparison to what would occur after an event that only affected one thing VaNDed. For example, suppose a President, who values the United States¹⁴, loses over 3,000 citizens at the hands of terrorists, who are known to have subsequently evaded U.S. military capture. This President will be *more* saddened, *more* angry, *more* grief-stricken, *more* depressed by this event (which he defines collectively as the terrorists' attack and their subsequent escape), than if the attack had occurred but the attackers had at least been captured.¹⁵

¹⁴ This example has this President valuing the United States, the lives of its citizens, the loved ones of lost citizens lost due to acts of terrorism, and justice.

¹⁵ Another way one might want to put this is to say something like, "The President will be just as upset about the victims even if the terrorists are captured but this will be offset by his satisfaction of guilty having been captured." I am not certain that this way of putting it is different than mine, though. In his

When I say that the President would have reacted with less sadness, depression, anger, and grief if the attackers had been captured, I am assuming that in their capture, there are not some unexpected immediate consequences that complicate matters, e.g., the President learns that, shockingly, the capture of the attackers leads other terrorists to kill recently captured hostages, so he is not less sad, depressed, angry, or grief-stricken with their capture because more of those he values lose their lives. In other words, when I say that the President would react less intensely if the attackers had been captured, I am assuming that, if the alternative event happened, the President would not find any additional welcome or unwelcome consequences that would further complicate the President's reaction to the attack followed by the capture of the attackers and a less intense emotional reaction would occur. In general, this discussion of clashes and synergies should be understood as having a built in *ceteris paribus* clause.

If the President can both comfort the families of the citizens lost and bring the terrorist to justice by a swift military reconnaissance mission, then authorizing that mission is an easier action to choose over actions that can only comfort family or only bring about justice but not do both. Moreover, possible actions that will both upset the families and be unjust are easier to dismiss as possible courses of action over alternatives that are only comforting to the families or only bringing about justice. When I say that deciding what action to take is easier in cases of synergy, I mean that the reasons behind

sadness and such being offset, he has less moments in which he feels those emotions as intensely than if those emotions had not been offset. The President does not find the loss any less an evil, but he will not feel as badly about it – as intensely or as often – because he will have this sense of calm about the capture of the terrorists. This is similar to when the families of missing persons who are likely murder victims find closure in the remains being recovered and buried or the criminals being brought to the justice that the family thinks the criminal deserves. Those families do not find those murders to be any less horrible but they have a few more moments of peace due to these different kinds of closure and have a different emotional experience – sometimes greatly different, sometimes slightly different – than they did before that closure.

the attitudes and actions regarding those things VaNDed are not in conflict and all those reasons support a particular course of action (or resoundingly oppose a particular course of action), and that, moreover, it be may emotionally pleasing to take those actions (or avoid other actions). For this President, it is comforting, in addition to other pleasant emotions, to choose the military reconnaissance mission given how it will benefit what is valued and harm what is disvalued.

Complex Clashes and Synergies

Both clashes and synergies can become parts of more complex clashes and synergies. This occurs, for example, when one values an anticipated a state of affairs because it will have a minor clash or positive synergy but then actual state of affairs ends up being an event which involves something worse than what one expected, such as a major clash or a negative synergy. This also occurs when one disvalues an anticipated clash or synergy but then an event occurs which involves something better than what one expected, such as no clash or a positive synergy. For example, suppose George anticipates that tomorrow's event will involve a positive synergy, and so he values that event for its anticipated positive synergy. He expects it to benefit three things valued, harm one thing disvalued, and leave everything else unaffected. When the event occurs but harms three things valued and one thing disvalued, and leaves everything else unaffected, George is *more* disappointed with this clash than if he had not already anticipated a more positive synergy instead.

Alternatively, suppose one is expecting a clash between two highly valued things. One can value or disvalue this anticipated clash. Then, instead of what one expects, an event occurs that eliminates this clash, by eliminating the choice between the two and

leaving one with only one thing that is highly valued. One may be more pleased by the lack of choice, due to the relief of avoiding the clash of deciding between two things valued highly, or one may be disappointed because one values having more options even though having more options comes with clashes.

Consider the examples of Tracy and Ben. Tracy and Ben are each applying for philosopher jobs, though in unrelated areas of philosophy, and each anticipates being offered positions by his top two choices and needing to decide between those two choices, the first choice valued only slightly more than the second. Each learns that his first choice school hires someone else instead, and for each, his option is to take the post at his second choice school. For months, Ben is sullen whenever he thinks of how he was passed over by his first choice. Unlike Ben, Tracy, after a brief period of disappointment, is happy that his first choice chooses another because he realizes that this rejection allowed him to avoid the gut-wrenching selection of his first choice over his second choice. The difference between these two is how each values the clash in combination with how each values being chosen by two top choices. Tracy strongly disvalues the clash that would occur in deciding between his two top choices, and, for him, this far outweighs his disappointment in being passed over by his first choice, so he feels brief sadness but then mainly relief and joy regarding selection by his second choice. In contrast, Ben strongly disvalues the clash because he greatly values being selected by his top choices; hence, his sullenness in the clash's absence.

Direct Valuing Connections: Had with varying degrees of awareness

Full awareness of the complexity of one's VaNDing is not necessary in order to so VaND. One need not be fully aware of direct valuing connections for those

connections to exist, just as one need not be fully aware of all that one VaNDs in order to VaND all that one VaNDs. One can be less than fully aware that one despises x , disrespects y , treasures v , loves¹⁶ w . One can walk around reacting to x in a despising manner, to y in a disrespectful manner, to v in a treasuring manner, and to w in a loving manner. If those reactions – those emotions and actions – are based on reasons that one sees or finds for those VND attitudes, then one values v and w , and disvalues x and y , though one may be less than fully aware of one's VND attitudes regarding these things, as well as unaware that those VND attitudes are thereby constitutive of VaNDing of those things.

Even more likely than being unaware of exactly *what* one VaNDs, one can be unaware of *exactly how* one VaNDs what one VaNDs because one is unaware of the potential clashes and synergies that may occur. This can happen for at least three reasons. First, again, one may not be fully aware of all that one VaNDs. Hence, one may not be fully aware of clashes or synergies surrounding things which one VaNDs with less than full awareness. Second, even if one is fully aware of how one VaNDs several things individually, as in when they seem be unconnected to anything else, yet one may be unaware of how one would react to clashes and synergies between these things. Why am I so sad after I got something I valued? The answer, of which I am not fully aware, is: Getting this came at the price of benefitting something I disvalued as well. Why am I so

¹⁶ At least one use of the word 'love' is to denote the positive attitude of valuing, rather than simple an emotion. When one loves something one does not simply have pleasant emotions in regards to it. Rather, one is happy when the object of love does well, sad when it is undermined and so on (though loving anything is much more complex than this due to the recently discussed direct event-based valuing connections – the love of one object stands in connection to and is partly defined by how one values and disvalues other things.) In other words, love involves a valuing of a beloved which includes an emotion stance and possible action stance towards the object of one's love. That emotion stance might include times during which one experiences what we might class "the emotion love," which I can only describe as "a warm happy glow" inside that forms when one's beloved comes to mind.

calm after something so horrid occurred? The answer, of which I am not fully aware, is: Something I disvalue even more was kept at bay. Third, even if one is aware of how one VaNDs some of the things that one VaNDs, both when those things are in isolation *and* when *expected* or known possible clashes or synergies would occur between them, one may be unaware of how one will VaND various things given *unexpected* clashes or synergies. Before experiencing such relief, Tracy is unaware of the relief that he will feel after being rejected by his first choice because he only anticipates that he will be saddened if he is rejected by his first choice.

The Synchronic Unity of a Person

Suppose Abraham values sunscreen in the summer and disvalues winter weather. There may be no reason or possible event which forms a direct valuing connection between these two valued things. The absence of direct reason-based and event-based connections between two VaNDed objects, still allows for **indirect valuing connections** (sometimes shortened to “**indirect connection**”) between them. The shortest indirect valuing connection involves three items; call the first two A and Z, which are indirectly connected through only one intermediary call B. A is directly connected to B, which in turn is directly connected to Z. For example, Abraham could value sunscreen in the summers for health, and disvalue winter weather because he is inept at outdoor winter physical activities, and so he is less able to promote his health in winter weather. The valuing of sunscreen is directly connected to the valuing of health, which is directly connected to the disvaluing of winter weather; hence, Abraham’s valuing of sunscreen has an indirect connection to his disvaluing of winter weather. Earlier I suggested another example of the shortest possible indirect valuing connection: If one values beauty

and also values a vase and a painting as instantiations of beauty, then the valuing of beauty acts as an intermediary for an indirect valuing connection between the valuing the vase and the valuing the painting.

Indirect connections between things VaNDed can be longer, in that there are more intermediaries between two things indirectly connected. An abstract example of a longer indirect connection between A and Z would be such that A is directly connected to the B, which is directly connected to C, and so forth, which is directly connected to Y, which is directly connected to Z. I suspect that my own valuing of my latest digital camera and my disvaluing of the texture of coconut are connected indirectly via several intermediaries.

If one were to visualize indirect valuing connections, indirect valuing connections are similar to the relation between atoms within a molecule that are not directly bonded to one another.

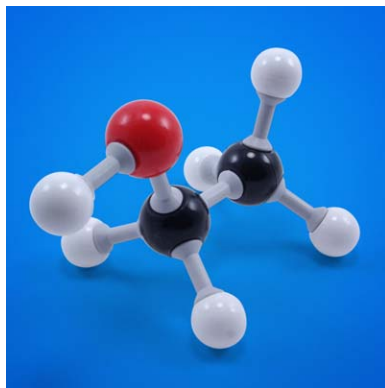


Figure 1: Model of the ethanol molecule. However, this model can be used as a visual illustration of a simple case of indirect v. direct valuing connections:

White balls = disvalued item
Black balls = valued items
Red balls = valued and disvalued
Gray links = direct valuing connections

In this example, the valuing or disvaluing of the red item is directly connected to the valuing of the first black item but indirectly connected to the valuing of the second black item.

Direct and indirect valuing connections together both define and separate persons at a time.¹⁷ Necessarily, valuing connections exist among all the things that a particular person VaNDs. Things VaNDed that are interconnected by direct and indirect valuing connections form a set, what I will call **a way of valuing**. Take all the things VaNDed by a given human being at time t_1 . If there are direct and indirect valuing connections connecting what or how all those things are VaNDed, then there is only *one* way of valuing and *one* person within that human being at that time.

In contrast, if there are n separate sets of interconnected VaNDed things, this means that there are n ways of valuing within that human being at that time. If there are n ways of valuing within that human being at that time, then there are n persons within that human being.

Synchronic Unity – Not Merely of a Moment

I contend that synchronic unity is incoherent if it requires (or even allows for) a way of valuing that exists only at an instant of time. Philosophers such as Korsgaard and Schechtman have criticized the idea of comparing a person at one point in time with a person at another point in time. Psychological features, whichever features one might emphasize, do not occur or exist in an instant. Narratives necessarily involve spans of

¹⁷ In defining persons using the combination of direct and indirect valuing connections, I recognize that the Universal Direct Connections Claim might be true:

Universal Direct Connections Claim: For all persons, everything VaNDed by the person is necessarily directly connected to each and every other thing VaNDed by the person.

If the Universal Direct Connections Claim is true then indirect connections would thereby be unnecessary to explain the synchronic unity of a person. However, at this time, I see no way to defend the Universal Direct Connections Claim, though I suspect the claim is true. For everything not connected with direct reason-based valuing connections would seem to be connected via direct event-based valuing connections. E.g., my valuing of my camera and my disvaluing of the texture of coconut would lead me to protect my camera over partaking in an opportunity to eat some coconut.

time over which experiences occur. VND attitudes, let alone entire ways of valuing, cannot be formed through seeing or finding reasons or expressed through emotions or actions in an instant. As Schechtman notes, “It is by no means obvious that the most essential part of a person’s experience at any time can be reduced in an independent time-slice, even if we imagine that slice containing all of the relevant forward- and backward-looking elements.”¹⁸ I think this skepticism is warranted.

My discussion of defining a way of valuing at a given instant does not presuppose that we must be able to coherently think of a way of valuing as existing for *only* an instant. Suppose that at time t_1 , there is only one way of valuing, one person, within a human being. A way of valuing can exist at time t_1 , yet that way of valuing extends either into the past, future or both, such that there is at least the minimum time necessary to exercise the capacity to value based on that way of valuing. The exercise of the capacity to value, given a particular way of valuing, involves expressing particular VND attitudes within a way of valuing. The expression of any VND attitude involves seeing or finding reasons to take that VND attitude, and possibly responding with emotions or actions regarding v according to that way of valuing. Seeing reasons can take less time than finding reasons, since finding reasons involves rational deliberation: seeing reasons can be as short as the time it takes to have a succinct thought, though even this is not an instant. However, the exercise of the capacity to value involves not only seeing and finding reasons, but experiencing emotions partly constituted by those reasons seen or found. Moreover, when one $VaNDs$ v , and so has a VND attitude regarding v , that VND attitude must be expressed such that the other things $VaNDed$ in fact shape that VND

¹⁸ Schechtman, *Constitution of Selves*, 144.

attitude regarding v .^{19 20} Hence, the minimum time necessary for a way of valuing to exist would seem to me to be longer than what is necessary for the expression of an independent VND attitude. Most importantly perhaps, a particular way of valuing can be expressed through the exercise of the capacity to value for hours, days, weeks, perhaps longer. Ways of valuing change, but not every few seconds of the day. Exactly how long any particular way of valuing exists will depend on exactly when that way of valuing emerges and when one sees or finds reasons to change; change is the subject of the next chapter.

Conclusion – From Synchronic to Diachronic Unity

What and how two or more things are VaNDed can have one or more direct valuing connections between them, which partly define what and how these things are valued. Suppose A through Z are each VaNDed by the same person. A and Z might be indirectly connected, with this indirect connection formed through direct valuing connections with other things VaNDed.

Note that indirect valuing connections are necessarily transitive, while direct valuing connections are not: If A is indirect connected to N and N is indirectly connected to Z, then A is indirectly connected to Z. However, if A is directly connected to M and

¹⁹ This does not mean that all the relevant VND attitudes are taken into account when one VND attitude is expressed.

²⁰ This is not to say that in order for a way of valuing to exist at time t_1 , it must express itself in emotion or action at or near time t_1 , but that a way of valuing cannot exist for shorter than the time it would take for a way of valuing to express itself. A way of valuing exists longer than any period of time that is so short that one could not, regarding things valued, see reasons to take attitudes, begin to experience an emotion from an attitude, or begin to take actions from reasons found for the attitude. I do not know the exact minimum amount of time any way of valuing needs in order to be expressive, but it seems to me that the amount of time is longer than any instant of time. Hence, when I say that a way of valuing exists at t_1 , this includes an understanding that this way of valuing exists for longer than t_1 .

M is directly connected to Z, A and Z are *indirectly* connected, but A and Z may lack any direct connections between them.

For example, consider the relationship between a vase valued for beauty and a painting valued for beauty. Suppose that they are *indirectly* connected through beauty itself which is valued. However, that vase and painting may lack any direct connections between them, though there is a direct connection between beauty itself and the vase as well as beauty itself and the painting.

When a human being has, through direct and indirect valuing connections, interconnectivity in all of what and how that human being values at t_1 , there is a way of valuing (call it Way_1), which defines and distinguishes the one and only person within that human being at t_1 . Then, suppose that some amount of time after t_1 but before t_2 , a qualitatively distinct way of valuing, call it Way_2 , comes into existence immediately after Way_1 disappears, and, then Way_2 exists past time t_2 . In the next chapter, I will explain how the person with and defined by Way_2 at t_2 could be the same person as the one with and defined by Way_1 at t_1 . I will argue that direct valuing connections, which form the synchronic unity a person, also enable diachronic unity, the persistence of persons over time.

CHAPTER 5: DIACHRONIC UNITY - CHANGING WAYS OF VALUING

Final Version of the Theory

On my account, any human being has personhood if and only if that human being has the capacity to value.¹ The capacity to value consists of integrated capacities to see and find reasons and to experience emotions. The exercise of the capacity to value produces VND attitudes, each of which in turn includes an emotion-stance and, likely, an action-stance. VND attitudes are shaped by one another through reason-based and event-based direct valuing connections. Synchronic unity for a person at some moment in time, m , is defined by the interconnected VND attitudes that form a way of valuing that exists over some period of time which includes m .

Few persons retain a single way of valuing over their entire existence. Persons change. The preliminary version of the valuing theory of personal identity over time holds that each person, as such, changes as a result of his or her own reflection:

Take some time interval t which starts at time t_1 and goes up to and includes some future time, time t_2 . Take some human being x that exists during t and who is person q at time t_1 . X is q at t_2 , if and only if during t how and what x values remains the same or differs as a result of q 's own reflection.

¹ Some implicit justification for this claim occurs in Chapters 3 – 6 through describing what the capacity to value involves and produces. See this project's Conclusion chapter where I address lingering challenges, especially against the claim that the capacity to reason alone is necessary and sufficient to give a human being personhood.

The focus of this chapter is to define and explain what counts as “one’s own reflection,” what makes it (1) *a matter of reflection* and (2) *one’s own reflection*.

I hold that the relevant capacity to reflect is itself built into the capacity to value. At this stage of the project, I believe that my claim that the capacity to value, as I have defined it, is *a capacity of reflection* is an uncontroversial claim. The capacity to value is a capacity to reflect because its exercise involves seeing and finding reasons, reasons to have the VND attitudes which make up a way of valuing.

The next question is, “What makes any exercise of a capacity to value specific to a particular person?” My answer begins by first recalling from Chapter 4 that no way of valuing exists for an instant of time, such as *m*, and any way of valuing exists for some period of time.² Hence, the person who formed and is defined by a particular way of valuing at *m* exists at least as long as that particular way of valuing exists.

For the person defined by a particular way of valuing for some period of time that includes *m*, after an event occurs, two general responses may occur. One response to an event is the one discussed in Chapters 3 and 4. The person will continue to see or find reasons to react with emotions or actions to the event as that way of valuing (call it *Way₀*) dictates. The person will have unique emotions and actions that are more or less paradigmatic given exactly how the event relates to things about which the person has VND attitudes. When a person’s exercise of the capacity to value goes as that person’s way of valuing dictates, that person is directing himself. *Way₀* was formed by the person: the person sees or finds reasons to hold the VND attitudes of the person’s own way of valuing.

² See Chapter 4 for full discussion.

The second way to respond to an event is that one may see or find new reasons to change – slightly or greatly – one or more of one’s VND attitudes in one’s current way of valuing – Way₀ – and this change results in a new way of valuing, call it Way₁. Events which lead to transitions are what I will call **transitional events**. When transitional events occur, rather than Way₀ dictating the mere expression of itself in emotions or actions, Way₀ dictates changing one’s set of VND attitudes. This change may be slight or great, but so long as there is any difference between Way₀ and Way₁, then a difference has occurred in one’s way of valuing.

Direct valuing connections secure synchronic unity of a person by creating interconnections between the VND attitudes that are had at a time. Similarly, there are other direct valuing connections that secure diachronic unity. Call these **transitional connections**. Transitions occur because there are direct valuing connections between things VaNDed in one’s current way of valuing (Way₀) and things VaNDed in a potential future way of valuing. Such transitional connections are both reason-based *and* event-based connections. It is an *event*-based connection because a transitional event initiates the transition from the old to the new way of valuing. Based on some of the VND attitudes within Way₀, the person sees or finds, within this transitional event, the grounds to reject, alter, or add VND attitudes and thereby transition from Way₀. It is a *reason*-based connection because it involves seeing or finding reasons, given reasons for several VND attitudes already had, to have one or more different VND attitudes. When a transitional event leads a person along a transitional connection, that person who was defined at some earlier stretch of time by a previous way of valuing (Way₀) will transition to a new way of valuing (Way₁). This transition makes it the case that the

person who once had Way₀ is the same person who later has Way₁. This pattern can repeat such that the person persists through many transitions, e.g., from Way₁ to Way₂ and later from Way₂ to Way₃, and so on.

Moving forward, I will use generic terms such as “change” or “difference” to denote any two ways of valuing that are distinct. In contrast, by **transition**, I am denoting when one way of valuing causes a new way of valuing to emerge from the previous way of valuing given a transitional event.

When there is a way of valuing that then leads to a transition to another way of valuing, a **stream of valuing** exists that consists of at least two ways of valuing and a transition between them. A person within a given human being is defined by a unique stream of valuing. A stream of valuing may actually consist of only one way of valuing with no transition. A one-way-of-valuing-long stream would define a person who did not change her VND attitudes during her total existence. Most often, though, persons transition to subsequent ways of valuing more than once over their existence, such that most streams of valuing consist of more than one way of valuing. When subsequent ways of valuing are part of the same stream, these subsequent ways of valuing are connected by direct valuing connections. A valuing connection constitutes each transition from a previous way of valuing to a later way of valuing. A stream of valuing that is larger than one way of valuing consists of one way of valuing (Way₀) transitioned to another way of valuing (Way₁), which may then be transitioned to yet another way of valuing (Way₂), and so on.

Streams of Valuing:

No transitions to another way of valuing, stream is only one way of valuing
Way₀

OR

Two or more ways of valuing connected via direct valuing connections, such as
Way₀ → Way₁ → Way₂ → Way₃ ... and so on

Key:

Way_n = a distinct way of valuing

→ = the direct valuing connections which constitutes the transition from one way to another way.

Ways of valuing are not necessarily consecutive

Note that streams of valuing longer than one way of valuing necessarily consist of ways of valuing with transitions between them. However, streams of valuing do not necessarily consist of *consecutive* ways of valuing. A stream of valuing that consisted of consecutive ways of valuing would have transitions between ways of valuing that existed in an uninterrupted chronological sequence. If a human being's total ways of valuing are such that each is transitioned to the next one immediately following it chronologically, then consecutive ways of valuing would be a part of one stream of valuing, and this human being, during its entire personhood, is one and only one person.

The valuing theory allows for the possibility of multiple persons within one human being. Given one human being, there is normally only one stream of valuing and therefore only one person over the life of that human being. However, it is possible for more than one stream of valuing to occur in the life of a human being, and this entails that there is more than one person over some span of time within a given human being. Since any stream of valuing does not necessarily involve consecutive ways of valuing, multiple persons within one human body could either coexist at the same time or exist during non-overlapping periods of time.

For example, suppose Dr. Jekyll and Mr. Hyde *are* two distinct persons (and the only distinct persons) within a given human being. Suppose the ways of valuing of that human being begins with Way₀ Way₁ Way₂ and so on until biological death at Way₅₀₀.

Dr. Jekyll could have the steam of valuing consisting of:

$$\text{Way}_0 \rightarrow \text{Way}_4 \rightarrow \text{Way}_8 \rightarrow \text{Way}_{12} \rightarrow \dots \text{Way}_{500}$$

And Mr. Hyde had the stream of valuing consisting of:

$$\text{Way}_1 \rightarrow \text{Way}_2 \rightarrow \text{Way}_3 \rightarrow \text{Way}_5 \rightarrow \text{Way}_6 \rightarrow \text{Way}_7 \rightarrow \text{Way}_9 \rightarrow \text{Way}_{10} \rightarrow \text{Way}_{11} \dots \rightarrow \text{Way}_{499}^3$$

Alternatively, two persons could exist within a human being at different times. The example of two subsequent persons existing within one human being is the example of the human being Boris in the introduction of Section II. If Boris had a ways of valuing, Way₀ through Way₅₀₀, then the two persons within that human being could have the streams of valuing consisting of the following:

The person that Boris is before head injury:

$$\text{Way}_1 \rightarrow \text{Way}_2 \rightarrow \text{Way}_3 \rightarrow \text{Way}_4 \rightarrow \text{Way}_5 \rightarrow \dots \rightarrow \text{Way}_{220}$$

The person that Boris is after the head injury:

$$\text{Way}_{221} \rightarrow \text{Way}_{222} \rightarrow \text{Way}_{223} \rightarrow \text{Way}_{224} \rightarrow \text{Way}_{225} \rightarrow \dots \dots \rightarrow \text{Way}_{500}$$

³ This example has Dr. Jekyll's and Mr. Hyde's transitions following a pattern: Dr. Jekyll has one transition and then Mr. Hyde has three and then Dr. Jekyll another one followed by Mr. Hyde having another three. If there were two or more persons coexisting within one human being no pattern may emerge in how many transition occur in one person compared to how many occur in the other persons. I put a pattern in the example so that I could use the "and so on" punctuation (...) and avoid listing out all of the ways of valuing in each stream.

At this point, we have enough terminology to introduce the final version of the valuing theory of personal identity over time:

Take some time interval t which starts at time t_1 and goes up to and includes some future time, time t_2 . Take some human being x that exists during t and who is person q with Way_0 at time t_1 . X is q at t_2 if and only if x has q 's stream of valuing at t_2 . X has q 's stream of valuing if either (1) or (2) is true: (1) x has Way_0 at t_2 or (2) x 's way of valuing at t_2 is the product of transitions from Way_0 .⁴

As an example, below I describe Sally, a person whose ways of valuing changed over the last year, and I explain how changes came about such that she is the same person. After describing Sally and her particular transitions, I discuss various ways that a person might transition from one way to another way of valuing. Much of the discussion regarding various ways that a person can transition is meant to show how many kinds of changes, which may appear to some to not involve transitions, in fact involve transitions.

Sally: An Illustration

One year ago, Sally valued her computer solely extrinsically because it enabled her to do her valued work and enjoy her valued hobbies. In addition, Sally strongly disvalued all religious things: she thought religions were dogmatic, authoritative, and lacking in any positive qualities, and she thought religious people were deeply flawed for having such beliefs. In fact, one of her hobbies was writing an anti-religious blog. Call

⁴ This definition does not explicitly address if r and q are co-existing persons or subsequently existing persons within X during some or all of t . Suppose x is q during t . That does not prevent there being a distinct stream of valuing at t_1 or t_2 that does not belong to q but rather to r :

X is r at t_2 if and only if x has r 's stream of valuing at t_2 . Suppose x is r and q at t_2 . X is two distinct persons, r and q , at t_2 if r 's stream is not the same as q 's stream.

Sally's way of valuing one year ago, which includes this particular valuing of the computer and disvaluing of all religious things "Way₀".

Eleven months ago, Sally's old computer broke down and was irreparable; she began to disvalue the computer. Call the new way of valuing that results "Way₁". A month later, she began to value some religious teachings and people. Call the next new way of valuing that results "Way₂". Over the rest of the year, Sally's valuing of religious teachings and people kept increasing. Call the subsequent ways of valuing that Sally has "Way₃," "Way₄," and so on; today she is at Way₂₀. Sally is a simplistic example: her transitions over the last year only number twenty and each somehow relate directly to a few things that she VaNDs, her computer and all religious things. While real people would likely have more transitions which are directly related to more than two things VaNDed, even given the simplistic case of Sally, how is Sally the same person over this last year?

Sally coming to disvalue her computer, and yet remaining the same person, may seem simple and unproblematic compared to her change regarding religious things. Since her valuing of the computer was solely on extrinsic grounds, it is not surprising that she no longer values it when it no longer supports anything else she values. She transitions to a disvaluing rather than a neutral attitude regarding the computer because she recognizes that the computer not only no longer supports those things that she continues to value (such as her work and hobbies), but it also is impeding her ability to engage in activities she values because it occupies needed space for a new computer. Given several VND attitudes in Way₀ about such things as her hobbies and work, she finds reasons to transition to Way₁ in which she disvalues the computer, is frustrated

when seeing it taking up valuable space, and acts to replace it at her earliest convenience. Now consider her gradual transition from disvaluing to valuing many religious things.

Sally had disdain and disrespect for all religious things, including religious people as such. She frowned, rolled her eyes, felt exasperation at the thought of anything religious, and would balk at the thought of attending any religious service. Her disvaluing attitude toward all religious things was so strong that she found reason to regularly decry all things religious on a blog. She found reason for this disdain and disrespect many years prior to one year ago, based on encounters with and facts about various religious things and people. Below are a few of the encounters and facts in which she found reasons to have this disdain and disrespect:

- She learned about the religious “cleansing” (e.g., massacres) throughout recorded human history.
- She grew up around religious people who wanted to exclude (what seemed to her to be) nice people because they did not subscribe to the dominant religion of her community.
- She was harshly rebuked and often punished the times in her childhood when she challenged (what she found to be) dubious claims in her family’s faith to her parents and family friends.
- At twenty years old (some twenty years ago), she became friends with someone who admitted that he was abused by a leader of faith, but who had nowhere to turn for help, as everyone else assumed that her friend was lying and that, as a religious leader, he must be innocent.

Someone might claim that these experiences and facts constitute reasons for anyone to disvalue all religious things. However, the valuing theory of personal identity does not address such normative claims. Rather, what is important in the valuing theory is that these experiences and facts constituted reasons *for Sally*.⁵ Having these and similar experiences and learning these and similar facts were events for Sally that led her

⁵ This discussion implies that Sally had previous ways of valuing and a period earlier in her life when she may have disvalued all religious things less deeply or perhaps even valued it, and then these events slowly led her to increasingly or more deeply disvalue it. Labeling her way of valuing one year ago ‘Way₀’ is not meant to suggest that Sally’s first way of valuing was Way₀, but rather the first one that we are considering.

to begin, later deepen, or simply reinforce a disvaluing of all religious things. Those events either (1) elicited emotions and actions which expressed the exact ongoing disvaluing at the time or (2) brought about a new way of valuing with a deeper disvaluing of all religious things. Today, she no longer remembers many of these exact experiences or facts. However, those experiences and facts, in addition to others, formed a standing belief in her that religion is highly dogmatic and authoritative. Sally had this standing belief for many years prior to one year ago. One year ago, as for several years, Sally saw these qualities of religion as constituting reasons for her disdain and disrespect of religious things.

About ten months ago, Sally read a journal article about a religion and its associated members, which and whom seemed neither dogmatic nor authoritative. In this article, she found reasons to value this religion. She read that the religion and its participants are inclusive and leave theological views very much up to the individual participants, with the minister offering general guidelines through one's spiritual journey which participants are free to utilize or reject. This religion and its members seemed to focus on challenging their own tendencies to view anyone as 'the Other,' fighting injustice in societies because they exclude others, and helping those who are in economic need. The religion and its members also seemed interested in forming bonds of friendship with one another and fostering a community with a diversity of members and ideas. Neither the religion nor its members appeared to show dogmatic or authoritative tendencies. Based on information in this article, Sally found reasons to value this particular religion as it appears to be in agreement with many of her VND attitudes,

including her valuing of ideals such as openness, diversity, equality, charity, community, and justice.

The rest of her transitions in ways of valuing can be summarized this way: Sally takes action to learn more about this particular religious group by attending a service. In the service, she notices attributes of the service which give her reasons to be even more admiring and respectful of the religion and its members; her valuing of the religion and its members slightly increases. She sees all of these new reasons to admire the religion as reasons to temporarily stop writing the anti-religious blog. Subsequently, to her surprise, at an after-service coffee, she runs into several friendly acquaintances who are active members of this group. Given her respect for these individuals, her valuing of this religion and its members deepens. While distancing herself from people who valued the anti-religious blog and its views, she keeps coming to services week after week, and becomes more involved in the group by becoming a member. She learns through interacting with many members that many of their admirable traits were instilled in them during childhood while they were members of other religious groups. As in Way₀, Sally would never join these other religious groups since they adhere to several ideals which she continues to reject. However, she comes to see how many of her fellow members were instilled with many admirable ideals through their childhood religions and sees that fact as reason to respect rather than disvalue many religions. She never returns to the anti-religious blog, loses friendships connected to that blog, and gains many friendships through her new religion. Today, she values many religious things at least in part, and strongly values one particular religion and its members.

To summarize, the story of Sally provides a window into the last year of a person who is defined by a stream of valuing which consists of at least twenty-one ways of valuing with direct valuing connections which enables a transition between each subsequent way of valuing. This stream might in fact have many additional ways of valuing, which were formed before one year ago and may be formed after today, all connected to this year's portion of that stream with additional transitions.

Not all transitions, either in Sally or in others, share all the particular traits of the transitions I have just described. For example, persons can transition with a much lower degree of consciousness than was involved in Sally's transitions from disvaluing to increasingly valuing many religious things. The remainder of the chapter discusses several of the various ways that a person can transition between ways of valuing such that the person persists. The next chapter will discuss early applications of the valuing theory, and explain how various phenomena do not in and of themselves change that human being into another person. The phenomena discussed are akrasia and changes in a human being's moods, appetites, or connection to reality. My intention for the discussion in both Chapters 5 and 6 is to make it highly plausible that the valuing theory agrees with *and accounts for* our pre-theoretical intuitions regarding when a given human being persists – or fails to persist – as the same person.

Variation in how persons change

In Chapters 3 and 4, I discussed two sources of variety in how one may VaND anything, call that thing v . In Chapter 3 I discussed how one might VaND v in whole or in part, intrinsically or extrinsically, and so on. In Chapter 4, I discussed how the VaNDing of v can be shaped in different ways by VaNDing of other things, namely

through reason-based and event-based valuing connections between v and other things VaNDed. There is also variety in how one might transition from one way of valuing to another.

This section discusses several ways to transition from one way of valuing to another. Just as with the Chapter 3 discussion of the variety in the shape and scope of VND attitudes, this discussion on the variety in how one might change one's way of valuing is not meant to present an exhaustive list of that variety.

Self-Narrating Optional: Inclusive of Episodics

Sally's story lends itself to the possibility that a great deal of self-narrating could be involved in some of the changes to new ways of valuing. Yet, one does not need to engage in self-narrating in order to change even a very central part of one's way of valuing. (By very central, I mean a part of one's way of valuing from which many other VND attitudes get their existence and shape.) While self-narrating could occur, one may in fact always skip self-narrating, even when one is fairly conscious that one's way of valuing is changing or has recently changed. As an example of someone who does not require self-narrating, Sally is someone who simply sees new reasons to transition from disvaluing to valuing many religious things. In contrast, consider Kim, someone who had similar disvaluing of all religious things like Sally one year ago, but who, unlike Sally, engages in a great deal of self-narrating. Self-narrating, thinking about who one's is or ought to be, simply a way of thinking about what reason(s) one has to have or revise one or more VND attitudes. However, a person can also think about what reason(s) one has to have or revise one or more VND attitudes without thinking about it as reasons to have the same or a different self-identity.

I can imagine both Kim and Tim to be much like Sally, both who greatly disvalued all religious things, because each finds that all religious things go against certain ideals, such as inclusivity and equality, whose long-standing self-narratives include the statement, “I am a person who despises all religious things.” Each having read the same article that Sally did and given their ideals behind despising all religious things, those ideals support each of them moving towards increasingly valuing many religious things as Sally did. However, Tim’s ongoing narrative will help Tim move towards valuing many religious things, while Kim’s ongoing narrative will make it more difficult for Kim to move towards valuing many religious things. Consider first how Kim’s self-narrating is an impediment.

If Kim reads the article and starts making plans to go to a service of this religion, making such plans would contradict the part of her long standing self-narrative which says that she despises religion. She might be disgusted with herself and say to herself something like, “I am someone who despises all religious things. How could I even consider going to a religious service?” Her answer to this question will be found in considering whether there are more reasons to continue to disvalue all religious things or to begin to value one religion because this religion reflects the very ideals that led her originally to despise all religious things. In fact, subconsciously Kim already values the religion slightly but quickly notices how her new action-stance (e.g., attend a service) goes against her long-standing self-narrative. This tension between her long-standing self-narrative and her subconsciously valuing one religion slightly exists because she values the long-standing self-narrative: there is a clash between valuing this long-standing self-narrative and this one religion. Initially, she sides with the long-standing

self-narrative and so she is disgusted with herself. If she is to get past that disgust, she will discern, through the capacity to value, what to modify: her long-standing self-narrative or her new valuing of one religion. She may realize that she has reason to disvalue her long-standing self-narrative and adopt and value a slightly different self-narrative that says simply that she is someone who upholds certain ideals which explain why she now values a particular religion.

Kim's story speaks against narrativity. Kim has more reasons, from her own subjective way of valuing, to value this one religion. However, her valuing of the self-narrative that describes her as anti-religious pulls her towards continuing to disvalue all religious things. This example agrees with Strawson's claims about the sometimes problematic nature of narrativity.

Narrativity *can* be bad for some people.⁶ Sometimes, one can engage in self-narrating in a way that involves rationalizations and focusing on what supports rather than undermines a long-standing self-narrative. Hence, self-narrative is problematic when a self-narrative makes one resistant to making transitions to new VND attitudes,

⁶ Strawson has slightly different reasons than I do for holding that narrativity *can* be problematic. See Strawson's discussion on revision within self-narration, Strawson, "Against Narrativity," 443-446. He concludes that rationalization is not a necessary feature of narrativity by Diachronics but suggests that it may be a problematic part of some self-narrating.

Even with these concerns about narrativity, I do not believe that Strawson is trying to suggest that Episodics always avoid self-deception, internal inconsistency, immorality, and failures to progress. Instead, as he notes on page 437, he thinks that those who are *true* Diachronics can do well with self-narrative, while *true* Episodics are likely to make a mess of themselves through engaging in self-narrative.

Finally, Strawson seems to suggest that the best moral life is Episodic rather than Diachronic. See Strawson, "Against Narrativity," 437. In response to those that may be skeptical that Episodics can be particularly moral at all, he explains how the Episodic can be moral in Strawson, Galen. "Episodic Ethics," *Philosophy: The Journal of the Royal Institute of Philosophy* (Vol. 60, Supplement (January 2007), 85-115).

even though those transitions are arguably more justifiable, from one's way of valuing taken as a whole, than continuing on as before without those new VND attitudes.

While I do think self-narrative can cause such stubbornness or inflexibility in some, one could be just as stubborn without a self-narrative. The source of a problematic resistance to change is not a self-narrative itself. Instead, the problem is *the valuing of* the self-narrative, or, more specifically, the valuing of some of the particular attributes ascribed to one's self within one's self-narrative. Kim's valuing the part of her self-narrative about how she is anti-religious explains why she would experience ongoing emotional turmoil between her long-standing narrative and valuing one religion.

However, Sally, who is not engaged in narrativity, could have held onto her anti-religious position, simply because she was too invested in all religious things being something to disvalue and would be terribly resistant to understanding what she read in the article in a way that would lead her to change her anti-religious stance. I hold that both those with or without narrative have avenues of problematic stubbornness. These avenues of stubbornness stem from investment in one's current understanding of "the truth," whether that truth is about one's self (in a self-narrative) or about the value of other things, such as religion.

Moreover, while Kim's story makes narrativity look potentially problematic, narrativity can aid rather than hinder self-improvement. I can also imagine Tim, someone like Kim and Sally, similar in disvaluing of all religious things based on similar ideals, but his ongoing narrative greatly emphasizes him as someone who despises religion *because* he upholds certain ideals. When Tim reads the article, unlike Kim, he

will be more comfortable moving to valuing one religion because doing so agrees with his self-narrative which says that he is someone who stands for certain ideals.

While self-narrative may aid or hinder self-improvement, even for someone who has an ongoing self-narrative, something like the ordinary valuing of one's computer and then later disvaluing of one's broken computer, seems unlikely to be a part of anyone's self-narrative. The relative unimportance of the computer makes it highly unlikely that Kim or Tim would have as part of their respective self-narratives something like, "I am someone who valued and now disvalues the computer." Nevertheless, the earlier valuing and later disvaluing of the computer is a part of who these individuals are, and the transitions to different ways of valuing explains the different emotions that each would have towards a computer before it broke compared to after it broke. Just as the non-narrating Sally, their VND attitude towards the computer is part of each, yet each could easily value and later disvalue a computer without self-narrating about it.

Self-narrative might occur, but is not required, on the valuing theory of personal identity over time. An Episodic who lacks a self-narrative to create one kind of connection between his further past and further future, could still have a connection to his further past and further future, through transitions from one way of valuing to the next, and thereby persist over time as the same person.

Transitioning on a Consciousness Spectrum

One changes one's way of valuing consciously, but to various degrees, with the highest possible degree of focus on transitioning at one end of the spectrum and with minimal attention on transitioning at the other end of that spectrum. Transitioning with minimal attention on transitioning is what I will call transitioning fairly sub-consciously

and is similar to seeing what is in one's peripheral vision. What is in one's central vision is the focus of one's attention. However, one can easily take into account the presence and location of that object without making that object the focus of one's visual field. At the same time, that object in one's peripheral vision *can* become one's visual focus. In a similar way transitioning one's way of valuing can be conscious in a low key way, even if it is not the focus of one's attention.

Any transition falls on a conscious spectrum, just as anything in one's field of vision falls on a spectrum from central to peripheral. First, there is range of focus that one can have on one's reasoning about transitioning: while one is engaged in reasoning, one can be more or less attentive to that reasoning. Second, there can be a range of focus on how a transition to a new way of valuing would be a transition; one can be highly focused on how a transition, as such, is occurring or has occurred or fairly unfocused on how a transition, as such, is occurring or has occurred.

When Sally comes to disvalue the computer, she is not focused on her reasoning about the event of the computer becoming broken irreparably. Instead, she is simply engaged in that reasoning. Moreover, Sally does not focus on how her disvaluing of the computer is a transition; she simply transitions. In contrast, when Sally is coming to value that particular religion, she is very focused on her reasoning regarding why she should now value this one religion and very focused on how valuing one religion would be a transition.

Someone could go from valuing to disvaluing a computer that she valued extrinsically in a highly focused manner, by focusing on her reasoning regarding the computer, by focusing on how disvaluing the computer constituted a transition for her, or

by focusing on both. One or both of these higher levels of consciousness might occur for such a transition if someone greatly valued her computer and she and it were inseparable, such as a techie in love with her computer as the latest and greatest technological advance. Likewise, someone could go from strongly disvaluing religion to valuing one particular religion with low levels of consciousness, by not focusing on the reasoning for the transition and simply engaging in that reasoning, or by not focusing on how this would be a transition and simply transitioning, or both.

Sally's transition from valuing to disvaluing the computer was still a part of her consciousness. It was she who reflected on the computer and found reasons to disvalue it. Her new disvaluing of the computer did not happen to her like a sensation of pain in her leg can just happen to her. In seeing or finding new reasons, she brought about the transition. In contrast, she may do nothing to bring about the sensation of pain in her leg.⁷

In general, consider a human being who changes from one way to another, Way_n to Way_{n+1} . If the person defined by Way_n finds reasons based on Way_n to adopt Way_{n+1} , then even if that change is fairly subconscious, that change is in fact a transition and means that the persistence of the person who had Way_n to then having Way_{n+1} .

Indirect to Direct or Vice Versa

We can see that Sally's disvaluing of the computer occurred subconsciously, but we can see that she affected coming to disvalue the computer in part because she could

⁷ Moreover, even in the case of disvaluing her computer, we can see that it involved her reasoning because if someone were to ask her why she discarded her computer, she could easily provide her reasons. Providing such reasons for something as simple as coming to disvalue a broken computer might be easy, but it can be more difficult other times, e.g., if one both values and disvalues something. Hence, in general, I do not want to suggest that one must be able to recount one's reasons in order to have those reasons.

easily justify discarding the computer. Other times, one's way of valuing changes, but one is unable to explain how a new attitude towards something is connected to reasons that one sees or finds. Instead, one may recognize this change sometime after the change occurred and then one does not see how this change occurred through seeing or finding reasons. Rather, it seems to one that this change simply happened to one without one's reasoning. However, the valuing theory holds that ways of valuing cannot change within one person without exercise of the capacity to value. How can we explain these seeming shifts in a way of valuing that seem to occur to a person without any reasoning?

For example, consider Christina and her relationship to *The Bachelor*, a reality television show, in which, extremely attractive and successful women appear on the show to compete for the love of one wealthy bachelor. In the past year, Christina went from disvaluing to taking a slightly positive attitude towards this show. Christina thinks that, despite this mellowing, she certainly does not value the show; she knows that she certainly still would not watch it. At the same time, she finds that when seeing commercials for or references to the show where she once frowned and rolled her eyes, she now shrugs her shoulders to emote I-don't-get-why-people-like-it-but-maybe-it's-alright. Moreover, she would no longer bother to criticize her twenty-five-year-old daughter for watching it. She recognizes that her attitude towards the show has mellowed. However, when she considers that new attitude, it does not appear to her to amount to valuing the show, nor even to disvaluing it less. In fact, Christina finds any redeeming quality that this show may possess to entice a person to watch the show to be elusive. In this mellowing, did her way of valuing somehow shift from disvaluing without her changing her own way of valuing?

When I asked Christina why she thought that her view of the show had mellowed, she mentioned that over the last year, she learned that several of her friends enjoy the show. This response suggests to me that it is because she values her friends and values not only their experiencing pleasure, but also their opinions that her disvaluing of the show changed from disvaluing to slightly valuing it. This explains her mellow, slightly positive attitude.⁸ This attitude of somewhat valuing the show came about through her transitioning to a new way of valuing based on her previous way of valuing: she previously valued her friends, and that continued valuing supported an adjustment to her valuing attitude regarding the show.

I claim that this shift in Christina partly depends on the synchronic unity of any way of valuing. As discussed in Chapter 4, some of a person's VND attitudes are directly connected while others are only indirectly connected. Christina's shift involves going from indirect to direct valuing connections between the friends and the show, towards both of which she held specific VND attitudes in her old way of valuing. Generally, when going from an indirect to a direct valuing connection, one has particular VND attitudes to two or more things, but there is no direct valuing connection between these things in the previous way of valuing. Then, one sees or finds reasons which change not only one's emotion-stance regarding one or more of these things, but which constitute a new direct valuing connection between these things. Christina valued her friends and so

⁸ It is not clear to me that the real life example, on which Christina is based, came to just disvalue the show less, value it slightly or take a neutral stance regarding it. However, it is not relevant to my point that Christina changed to one of these particular new VND attitudes. Rather, the point is that Christina, from her way of valuing prior to the change, reacted to learning that her valued friend valued the show by shifting her VND attitude in a positive direction towards the show indirectly though subconsciously.

once she learned of a supportive connection between her friends and the show, she valued the show herself at least partly.

Transitions from direct to indirect connections are similar to transitions from indirect to direct connections. Suppose that Molly was like Sally in that she needed her own computer to help her do things she values, which is partly why she valued it, and recently her computer broke. Suppose Molly's broken computer, a laptop, had lasted many years, which is partly why she valued it, and was not taking up needed space even after it broke, and she found no reason to be frustrated with it once it was broken. Even though Molly takes the event of having a broken computer in stride, her VND attitude towards the computer becomes more negative than before it broke, due to the new absence of a previous direct connection between the computer and other things she valued. She would value it less because it cannot help her with other valued projects, though still valuing it to some degree because of all the years supporting valued projects that it gave her.

New direct connections between things previously connected only indirectly, or vice versa, can occur fairly subconsciously. Sally's and Molly's respective new VND attitudes towards their computers, respectively, as well as Christina's new VND attitude towards the TV show, illustrate one way that a change in one's way of valuing may not involve attention focused on the change itself. However, any of these changes could have occurred with greater attention to that change instead, e.g., Christina could have recognized that her changed attitude towards the show was a change in VND attitudes and that she was coming to value the show slightly after strongly devaluing it because she greatly valued her friends.

These shifts from direct to indirect valuing connections and vice versa often reflect how persons have a tendency towards rational coherence. Christina's example clearly shows this tendency towards rational coherence. For example, Christina's respect for the taste and judgments of her friends is in tension with the contempt for *The Bachelor*, and so she is motivated to find reasons to either change her views of her friends or of the television show.⁹ Though striving for overall rational unity is not all that unifies persons, I agree with Rovane's emphasis on rationality. When a person changes herself by transitioning from a previous way of valuing to the current way of valuing, she makes this change on the basis of reasons that cohere with at least some portion of her previous way of valuing.

Meticulously v "Quick & Loose"

One's change can be more or less meticulous and methodical. That is, as one transitions to a new way of valuing, one may consider the details of the event which precipitates a transition and its relation to one's current way of valuing *more or less carefully*. When a transition is meticulous, that transition involves a methodical consideration of possible reasons to change one's way of valuing and often involves more detailed rational analysis, though such analysis may still be flawed due to human error or moments of irrationality. Sally's change in disvaluing to valuing many religious things was not only fairly conscious, it was methodical. She was very careful to consider why she should change her view on one religion, and what she expected to see at the service that would make her more secure in her new valuing of this one religion.

⁹ Molly's example is complex because even after the computer breaks, for her, it has some direct connection to valuing projects which it helped to complete in the past.

In contrast to the methodical nature of her coming to value many religious things, Sally's coming to disvalue of the computer was quick and loose. A quick and loose change is not brought about through any sort of consideration of what might qualify as a cogent argument in favor of changing one or more VND attitudes. Instead, such changes in VND attitudes occur through quick use of rules of thumb, whose use, I claimed in Chapter 3, was an exercise of reason. As I noted there, use of rules of thumb involves reasoning that is generally flawed in that it is more likely to have formal and informal fallacies. The more methodical one's reasoning the more likely one can avoid such errors. However, the strength of reasoning with rules of thumb lies in how this reasoning can occur more quickly and it often leads one to the same conclusions as a more meticulous analysis would.

The more methodical one's reasoning the more likely one is to do that reasoning with one's full attention.

With or Without Surprise

Some transitions to a new way of valuing involve the emotion of surprise. Someone might object to the valuing theory on the grounds that a human being that we think is the same person before and after a change to a new way of valuing experiences surprise about this change. This surprise suggests that the person persisted but without changing himself. In response to this possible objection, I will argue that the fact that we could be surprised that our way of valuing has changed or is changing does not imply that we are not in fact transitioning from one way of valuing to another way of valuing.

I would like to suggest that the emotion of surprise occurs when one forms a judgment not anticipated given prior judgments. The bodily sensations of surprise

include an urge to freeze, which I suspect enables one to take a moment to decide what to do next in the light of this unexpected judgment. Sally has changes which involve surprise and others which involve no surprise.

Sally may have been surprised that her computer broke, but there is no emotion of surprise around her change from valuing the computer to disvaluing it. It is no surprise, and so there is no emotion of surprise for Sally in disvaluing her computer once it was no longer useful, since she only valued it for its usefulness.

In contrast to coming to disvalue her computer, there were moments when Sally experienced the emotion of surprise in regards to some religious things because she had originally strongly disvalued all religious things, and found and saw no redeeming qualities in anything religious. Hence, she was surprised when she learned of a religion that might have qualities that she would take as reasons to admire it. She was surprised when she learned that the members were people she already admired and surprised that some of their admirable qualities were partially due to upbringings in religions that Sally continued to think were nevertheless too dogmatic and authoritative. She was surprised to see reasons to at least partially admire many religious things, and surprised to come to believe that many religions are partially admirable, though she continues to feel many were still problematic as well. Sometimes she felt surprise when she simply thought about how much she valued many religious things; now it feels surprising given how strongly she disvalued all religious things in the past.¹⁰

¹⁰ In a similar way, some third parties who knew Sally as someone who disvalued all religious things would also experience surprise if they learned that she is religious today. Moreover, they would not experience surprise in learning that Sally came to disvalue her computer.

There is not only surprise as one is transitioning, but if one does think about one's further past, one might experience surprise when one compares some aspect of one's past ways of valuing to some aspect of one's current way of valuing. Sally could experience the emotion of surprise if she compares her VND attitude on religion one year ago and her VND attitude on religion today. The difference between those two attitudes is stark. However, if Sally can recall the various events that lead to various transitions, she will understand how she changed herself. While Sally might be able to recall many of the experiences that support her very different VND attitude towards religion, others could have difficulty making sense of how one transitioned from a VND attitude long ago to one of today. If those transitions were complex, less meticulously, and occurred more subconsciously, such self-knowledge is more elusive.

Pleasures

In college, Ryan loves "to party" – going bar-hopping and going to loud drunken house parties – most nights of the week. After college, Ryan gets a job and puts partying aside. Right after college Ryan still valued partying but also valued gainful employment and its many benefits, so finds that he needs to stop partying. He is saddened initially, but, given its costs, he does not return to partying. Ten years later, a college friend, Victor, is in town and Ryan is looking forward to a night of bar-hopping like the old days. However, he goes out bar hopping and he no longer values partying.

Before the outing, Ryan would say that he still values partying. Afterwards, Ryan concludes he was wrong: he did not value partying anymore. Ryan might wonder, "When did I change?" Ryan's case might make someone question whether Ryan changed his own way of valuing. This challenge claims that rather than changing

through seeing or finding reasons, Ryan simply no longer enjoyed the attributes of bar-hopping.

I will explain how Ryan's new disvaluing of partying is a transition from the way of valuing that he had when he entered the bar (call it Way₁₀₀₀). Way₁₀₀₀ developed from many transitions between his way of valuing after college when he gave up partying and Way₁₀₀₀. Several of Ryan's earlier transitions related to him coming to value attributes absent in partying and disvalue attributes present in partying, e.g., coming to highly value small gatherings with lucid conversation on interesting topics with light music in the background. By the time Ryan went bar-hopping ten years later, his way of valuing during that outing was such that the attributes of partying were already disvalued and Ryan simply needed to realize that bar-hopping had those disvalued attributes.

Even though this change did involve seeing and finding reasons as dictated by Way₁₀₀₀, it is also true that the attributes that Ryan disvalued in partying, were disvalued because he stopped enjoying those attributes. Ten years ago he enjoyed many attributes of the bars: the smokiness, the crowdedness, the pulse from the loud music, the silliness of friends and folks when intoxicated. We might say this change to not enjoying these things is the result of getting older. Getting older can change what is enjoyable. Before the outing, Ryan still valued partying but only because he thought it would still be enjoyable to him. However, once he experienced partying again, he realized how he disliked its attributes and so he quickly found reasons to no longer value it. Ryan saw many aspects of the bar experience as reasons to disvalue it – e.g., smoky rooms, silly acts of intoxication, conversations that one cannot hear, ear thumping music.

Ryan's story reflects how some things are valued because those things are pleasant, rather than those things being pleasant because one values those things.¹¹ VND attitudes include emotion-stances, but this does not entail that VND attitudes account for *all* bodily sensations surrounding something for which one has that VND attitude. Some emotions towards v , something VaNDed, are completely independent of one's VND attitude towards v . Moreover, one can find that the emotions that are independent of VaNDing v constitute reasons to VaND v . If one valued something solely or mainly because it was pleasant and it is no longer pleasant, then one has reason to change one's VND attitude towards it.

Conclusion

This chapter describes several ways that a person can transition from one way to another way of valuing. Sometimes we have the intuition that a human being persists as the same person while also thinking that there was a non-transitional change from one way of valuing to the next. Besides introducing the final version of the valuing theory, this chapter focused on resolving the conflict between the valuing theory and several of these possible intuitions.

It is highly implausible that any human person has full self-knowledge and so, due to this lack of self-knowledge, someone could transitions but it seems like a non-transitional change. The (1) subconscious level of a change, (2) surprise about a change, (3) engaging in very quick and loose change, can help to *disguise* transitional changes that involve (1), (2), or (3) as non-transitional changes.

¹¹ Of course some things are valued despite the fact that they are unpleasant.

Finally, some transitions are disguised as non-transitional ones because the valuing of v was based on the enjoyment of v . In such cases, the new lack of pleasure may not be a psychological change in which one had a say but it is also not a change that determines whether one persists. Instead, the person persists through seeing or finding reasons to disvalue v in the fact that v is no longer pleasurable.

Hopefully, this chapter has shown that because persons can transition without full self-knowledge and often (1) subconsciously, (2) surprisingly, (3) quickly and loosely, and (4) regarding something VaNDed because of the bodily sensation(s) it produced, that any of these phenomena alone or in combination does not entail that the person has a non-transitional change. Not surprisingly, there are times when a human being intuitively appears to persist as the same person yet not through transitions but those times fall outside of the topics discussed within this chapter. For example, someone could have the intuition that a change in mood or appetite caused a non-transitional change in a person's way of valuing but that person persists. Section III, the next section, which consists of the final chapter of the project, focuses on early applications and addresses how to understand human beings when they experience changes in mood and appetites.

SECTION III: EARLY APPLICATIONS

CHAPTER 6: VALUING THEORY OF PERSONAL IDENTITY OVER TIME - EARLY APPLICATIONS

Introduction

The valuing theory was developed to help identify persons when it is important to know whether to treat a human being as the same person. Is Margo the same person as before her illness or during earlier stages of her illness? If not, is that because she is no longer a person or because she is someone else?

An Objection

Someone might object to the valuing theory being applied to analyzing the identities of patients with Alzheimer's disease – a rather hard case – because, this objection goes, the valuing theory appears to be far from our pre-theoretical understanding of clear cases of when human beings remain the same persons. The valuing theory requires that persons be efficacious valuers. Suppose q is a person within a given human being before t_1 . The valuing theory holds that if after t_1 , that human being VaNDs or changes his way of valuing differently than q 's way of valuing dictates, then that human being is either temporarily or permanently not q . Don't persons fail to be efficacious valuers, when they act akratically and when they act erratically while in an atypical mood – in a mood not typical for that person? Periods of akrasia and periods in an atypical mood are both periods of time when that human being appears to fail to act (or experience emotions) as q 's way of valuing dictates. However, the objection

concludes, when human beings go through periods of akrasia and atypical moods, these are clear cases when a human being remains the same person, not when that human being becomes someone else, even temporarily.

Plan for a Response

In this chapter, I will explain how, on the valuing theory, persons *do* persist through akrasia and atypical moods. Persons can engage in akrasia, when they choose actions contrary to their actual or potential better judgments. Persons can experience atypical moods and perform different actions or have different emotions than they would have had if they were not in that mood. In both the cases of akrasia and moods, I will explain how the valuing theory agrees with the pre-theoretical intuition that persons persist – the human being does not temporarily become someone else. I will provide an application of the valuing theory during akrasia and atypical moods which includes considering exactly what akrasia and atypical moods are. After explaining how the valuing theory can accommodate akrasia, I will argue that erratic actions and emotions under moods can be understood as phenomena similar to akrasia – in fact, some actions under moods are akratic. I will explain how akratic actions as well as erratic actions and emotions during atypical moods actually stem from one's way of valuing, rather than fall outside of it.

Akrasia and Moods Involve Emotions

The two phenomena akratic actions and mood-related erratic actions have a common feature between them. Both involve emotions. In the case of akrasia, against one's better judgment, a person acts in agreement with his or her passions, which is to say that one acts in a way that avoids or ends unwanted emotions or begins or continues

welcomed emotions. When Donna is in a happy atypical mood, she has emotional reactions to various things, which appear contrary to the reactions of the person she was – contrary to the reactions dictated by her recent way of valuing. Daniel is typically happy and carefree but, due to a headache, falls into a bad mood with many moments of anger.

In addition to sporadic atypical moods, a person can have a typical mood that is interrupted intermittently by an atypical mood (e.g., Monday blues, evening commuter road rage, drug-induced depression), and during each interruption, the person who has one set of typical actions and emotions during her typical mood, has a different set of actions and emotions during her recurrent mood. Whenever, Holly takes a certain drug to help improve her cholesterol, after a few days on the drug, she falls into a depressed mood, with regular moments of atypical sadness and hopelessness and avoids activities that, during her typical mood, she engages in often.

Because both akratic actions and erratic actions during atypical moods involve emotions, overlap exists. While in an atypical bad mood, uncharacteristically enraged with his children, Daniel yells at them, but he regrets his yelling and recognizes that yelling goes against his better judgment after the mood has passed.

That moods can lead to atypical emotions only fans the flames to the objection that the valuing theory is forced to say a human being is sometimes someone else while in an atypical mood. Daniel and Holly each have a typical mood that is accompanied by typical actions and emotions, but if they were in an atypical mood, they would perform atypical actions *and* have atypical emotions. There appears to be one way of valuing while in an atypical mood, whether rare or recurrent, and another way of valuing while in that human being's typical mood. Moreover, a human being appears to bounce between

two ways of valuing if that human being has an intermittently atypical mood. Given that the mood drives these shifts in ways of valuing with no seeming transitions between those ways of valuing, isn't the valuing theory forced to say that human beings often fail to remain the same persons while in an atypical mood?

I will argue that the valuing theory can explain how human beings do not become different persons simply because they are in an atypical mood. Given that moods relate to some akratic actions, first I will discuss how the valuing theory can accommodate akratic action and then return to considering the more complex phenomena of moods.

Akasia: Internal Conflicts

I will borrow the description of akasia from Aristotle. He discussed two types of akratic persons, the impetuous akratic and the weak-willed akratic.¹ The **impetuous akratic** is a person who often regrets acting in agreement with his passions and failing to deliberate. The impetuous akratic realizes too late that had he deliberated, he would have come to a judgment which disagrees with those passions and perhaps he would have acted according to that judgment. The other type of akratic is a person who deliberated and formed a judgment but then acts in agreement with his passions contrary to that judgment. This is the **weak-willed akratic**. Aristotle considered these forms of akasia as types of persons: the impetuous akratic *regularly* – has a pattern of – doing actions

¹ Aristotle, *Nicomachean Ethics*, trans. David Ross, rev. J. L. Ackrill and J. O. Urmon, Oxford, UK (Oxford University Press, 1925). My summary of Aristotle on akasia is based on Kraut, Richard, "Aristotle's Ethics", *The Stanford Encyclopedia of Philosophy (Winter 2011 Edition)*, Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/win2011/entries/aristotle-ethics/>

Kraut discusses alternative analyses of Aristotle on akasia in: Kraut, Richard, "Aristotle's Ethics", *The Stanford Encyclopedia of Philosophy (Winter 2011 Edition)*, Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/win2011/entries/aristotle-ethics/>

that he regrets later because those actions go against later formed judgments, while the weak-willed akratic *regularly* does actions contrary to his already formed judgment.

While Aristotle considered these forms of akrasia as types of persons, we could define akratic actions independently. An **impetuous akratic action** is one that agrees with the person's passions but, had the person deliberated at the time, he would have found that his judgment goes against taking the action in agreement with those passions. Instead of forming this judgment before performing the action, he performs the action and then realizes that he should have done otherwise; he forms the judgment too late. If the person actually deliberated and found that his judgment was contrary to his passions, then the person might have failed to act according to his judgment and thereby perform a weak-willed action. A **weak-willed akratic action** is one that agrees with the person's passions but not with the person's actual judgment. An impetuous akratic person regularly does impetuous akratic actions, while a weak-willed akratic person regularly does weak-willed akratic actions.

My discussion of akrasia will focus on the two types of akratic action, in part, to avoid defining who would count as an akratic person: individuals will disagree about whether someone is an akratic person or is simply a person who sometimes acts akratically. Specifically, there might be disagreement about this regarding the fictional persons that I describe throughout this chapter. I would like to avoid making controversial claims about whether someone is or is not an akratic person. At the same time, since akratic persons can be defined by akratic action, my discussion also applies to akratic persons. Given the details of my argument, if I succeed in showing how the valuing theory does not imply that akratic actions are done by someone else, then, that

worry no longer plagues the valuing theory regarding akratic persons, those who regularly engage in akratic actions.

For some if not all akratic actions, the passions involved are connected to appetites. There are innate appetites, which are shared by most human beings for things such as food, drink, and sex, as well as what we might call acquired appetites for such things as nicotine, alcohol, drugs, and for particular foods and drinks. Appetites produce cravings² for the object(s) of a particular appetite. Those cravings, in turn, bring welcomed bodily sensations when they are satisfied and unwelcomed bodily sensations when they are not satisfied. These are welcomed bodily sensations in that one has positive emotions regarding having and negative emotions regarding not having such bodily sensations and the reverse is true of unwelcomed bodily sensations. The appetite for food can bring a sense of physical satisfaction when satisfied (e.g., full stomach) and stomach pains and headaches when not satisfied. The appetite for alcohol for an alcoholic can bring lower anxiety when intoxicated and higher anxiety and headaches when not satisfied.

One incorrect way to think of akrasia is this akrasia is when a person has his rational judgment on one side and appetites, with their various bodily sensations, on the other and the latter wins *by force*. If akratic action is understood as appetites winning

² I am using the term ‘craving’ more broadly than in everyday language, to include desire for things that fulfill one’s basic biological needs, e.g., thirst is a craving in my use of the term though it is the desire to drink something, though nothing in particular. I recognize that ‘craving’ is a term that tends to be reserved for desires that go beyond or are even contrary to basic biological needs. In everyday language, it would be odd to say of someone who had been without food for several days, “He has a craving for food.” rather than simply saying, “He is hungry.” However, it would be in keeping with everyday language to say of that starving person, “He craves St Louis ribs,” since there are other ways to satisfy his hunger, so satisfying it with St. Louis ribs *per se* is unnecessary. In my use of the term ‘craving’ though, I would say of the starving man that he has a craving for food.

over reason by forcing a human being's movements, then there are no akratic actions. Instead, so-called akratic actions would not be actions at all and would merely be movements similar to the ongoing functioning or malfunctioning of internal organs (e.g., circulation of one's blood) and not something a person chooses to do. Akrasia is a potential problem for the valuing theory only if it is understood as an occasion when a human being is *willing or choosing* to go against judgments.

The objection to the valuing theory worries that in going against judgments of a person, akratic actions seem to reflect something other than the person's way of valuing. However, these akratic actions are the action of *some* person. If those actions do not agree with the original person's way of valuing, then it appears that the valuing theory is forced to say that those akratic actions were performed by a *different* person. Intuitively, though, when a human being acts akratically, it is the same person who had (or will have) a judgment *and* failed to act upon (or in agreement with) that judgment. In my analysis of akrasia, both impetuous and weak-willed akrasia involve internal conflicts within one's way of valuing. What one does (or would) conclude is an insufficient reason is acted upon over reasons that one does (or would) find to be decisive generally or in a particular situation. Simply put, akrasia involves acting against one's actual or potential better judgments in favor of more myopic judgments. I will argue that **better judgments** differ from **myopic judgments** in that the former are based on a larger portion of one's way of valuing than the latter.

Suppose Joslyn, a vegan, has a long-standing judgment that she should not eat meat because she is disturbed by the idea of eating a fellow animal. For many years, she has disvalued eating meat; in particular, she is disgusted by eating meat. However,

currently, Joslyn is pregnant and wants to create a healthy baby, and believes that she needs three servings of meat protein daily to create a healthy baby.³ She forms the better judgment that her disgust in eating meat is an insufficient reason to not eat meat three times daily during the pregnancy; her valuing of a healthy baby overrides her disgust decisively. Her myopic judgment is that she should continue to not eat meat given that she disvalues it. Both her better judgment and her myopic judgment stem from her way of valuing.⁴

On the valuing theory, akrasia reveals internal conflicts within one's way of valuing. Except for a way of valuing of the elusive perfectly virtuous person, each person's way of valuing is not perfectly rational and coherent. Two or more things can be VaNDed such that the VaNDing of those two or more things makes for a formally incoherent way of valuing, e.g., one both values and disvalues liberty as a whole. Moreover, even if one's way of valuing is formally coherent, it is likely, in a sense, incomplete. One's way of valuing is formed from various events that one either experiences or imagines, but it is not formed considering *all* possible events. Hence, various circumstances can arise which are event-based clashes between things VaNDed.

³ Joslyn may be wrong in this, and be able to create a healthy baby without eating meat. What and how one VaNDs depends on what one takes to be true, not what is actually true. Moreover, akrasia involves acting against one's better judgment, but not necessarily one's best judgment, or all-things-considered judgment. See following footnote.

⁴ A better judgment is one that is at least closer to an all-things-considered judgment but not necessarily an actual all-things-considered judgment. By an **all-things-considered judgment**, I mean one that considers one's entire way of valuing or at least all of those things VaNDed which are somehow affected by the decision at hand.

For Joslyn, the decision to eat meat is a better judgment but not necessarily an all-things-considered judgment: perhaps, all-things-considered, though she is troubled by eating meat, Joslyn should eat the meat generally (not just during pregnancy). For several years Joslyn has known that she is someone who has a mild but nevertheless harmful allergy to plant proteins, and so eating meat proteins is her only option for being even somewhat healthy.

During such clashes sometimes one will act according to a myopic judgment, which reflects only a smaller subset of what one VaNDs rather than act according to a better judgment, which reflects a larger subset of what one VaNDs. One's better judgment is (or would be) one that reflects a larger subset of what one VaNDs. But, again, only the elusive perfectly virtuous person will *always* form and follow his or her better judgment when faced with such clashes.

The valuing theory agrees with the intuition that a new person does not emerge when a human being performs akratic actions. Akasia involves the exercise of the capacity to value and VaNDing as a way of valuing dictates, but akasia simply reveals one or more places that that way of valuing is either incoherent or incomplete. This is true of both impetuous akratic actions and weak-willed akratic actions. Consider two examples of weak-willed akasia, first a continuation of Joslyn's story and then the story of the unwilling addict.

Joslyn concludes that she should eat meat three times daily for the health of the baby though she disvalues eating meat. Suppose she fails to eat meat more than twice a week during her pregnancy. Joslyn values creating a healthy child, but her disgust of eating meat does not cease. Instead, event-based clashes occur whenever the question of eating meat arises during the pregnancy, and she often chooses to forgo meat favoring her disvaluing of eating meat. If Joslyn's better judgment is acted upon over her myopic judgment, then she may be pleased with her choice, but mostly, she experiences revulsion while eating meat. When Joslyn's myopic judgment is acted upon over her better judgment, then she may feel horribly guilty for not eating meat, especially if, at her next

obstetrics appointment, her doctor tells her the fetus is not developing well due to a lack of meat protein.

Another example of weak-willed akrasia would appear to be unwilling addicts who continue to use that to which they are addicted.⁵ Physical addictions appear to produce cravings that promote myopic judgments and akrasia (rather than forced movements or non-actions). My example will be of an unwilling cigarette addict who continues to smoke. When other physical addictions lead to myopic judgments, then a similar analysis applies to unwilling addicts who are addicted to things other than cigarettes.

Consider an unwilling cigarette addict who continues to smoke – suppose it is Sally many years ago – who disvalues cigarettes and, though she is addicted to them, she does not value them to any degree.⁶ The better judgment common to most unwilling cigarette addicts is, “I should not smoke for long-term health reasons.”⁷ Not only does

⁵ This discussion of physical addiction assumes when one gives into one’s physical addiction, it is an action, a choice. If physical addiction sometimes or always does not involve agency, then, when it does not involve agency, it is not a case of akrasia: when it does not involve agency, it is not an example of an action at all, let alone an akratic one.

Even if acting on physical addiction is sometimes or always a choice, an action, I am not assuming that unwilling using addicts are somehow blameworthy compared to unwilling non-using addicts, the latter who are sometimes called recovering addicts. Even if unwilling addicts who use are always making a choice, for some of the unwilling addicts who use the physical cravings could be so strong that almost no one could rationally choose to not use under such pressures. While it is called weakness of will, I am *not* assuming that using unwilling addicts are weak: instead, the physical addiction may be so intense, that only a few extreme events would lead one to act against it.

⁶ An unwilling cigarette addict could both disvalue and value cigarettes, but her better judgment is that she should act on her disvaluing and not smoke. If she actually values cigarettes, then her akratic action is even easier to explain: irrationally, her valuing of cigarettes is acted upon instead of her disvaluing of it, though the latter is derived from a better judgment. So, in the main body of the text, I consider the slightly harder case of an unwilling addict who disvalues cigarettes and does not value them.

⁷ Each individual unwilling cigarette addict values health extrinsically given unique things in his or her possible future, e.g., retirement, grandchildren, world travel, more book reading, more sports watching.

Sally have this better judgment, but this judgment leads her to disvalue cigarettes, even though she usually smokes when nicotine cravings strike and she enjoys that smoking. While her action of smoking does not reflect her disvaluing of smoking, often her thoughts, emotions, and actions before and after occasions of smoking *do* reflect her disvaluing of smoking. Sometimes before taking her next smoke, she reminds herself of the health and other costs of smoking. After smoking, she feels the emotion of hopelessness (rather than indifference or happiness) about being unable to quit, she feels some self-hatred and berates herself for smoking, and sometimes she throws her cigarettes away (though, only to buy some more as soon as possible during a subsequent wave of cravings). Each occasion of smoking is an action and a choice on the part of Sally.

The objection to the valuing theory wonders why the valuing theory does not imply that every time Sally is taking steps to smoke and then smokes, she is a different person than she is after smoking and when occasionally she resists her cravings to smoke. She does not value cigarettes, yet she smokes. Two very different ways of valuing seem to be at play when Sally acts on cravings and when Sally either has no cravings or resists them.

In response to the objection, this is how I explain Sally's actions. While Sally disvalues cigarettes, she values other things that she gains from smoking. She experiences an event-based clash between the things that she values, which she gets from cigarette smoking, and the cigarettes themselves, which she disvalues. With a physical addiction such as nicotine addiction, during cravings, one cannot help but focus on the possibility of having many welcomed pleasant bodily sensations through smoking as well

as the actual unwelcomed unpleasant bodily sensations of continuing to not smoke. However, various bodily sensations may not be the only thing or the main thing valued that smoking provides. One thing likely to be foremost on any addict's mind during cravings is her lack of peace of mind and inability to focus well on anything other than wanting to use. Not only addicts, but most persons, value peace of mind and an ability to stay focused. This is because peace of mind and the ability to focus are extremely important for enabling one to do any action related to things VaNDed. Sally smokes cigarettes partly because she, like most persons, values peace of mind but her unsatisfied cravings deny peace of mind. On her better judgment, Sally judges that she should not smoke given the health costs. This better judgment is partly based on the hope that the outcome of a few weeks of non-smoking is that the cravings and their level of distraction will decrease, after which she would have greater peace of mind, a stronger ability to focus, and better health. Most often, though, Sally chooses to smoke. Even when she does not smoke during cravings, her disvaluing of lack of both peace of mind and focus remains high and she is, thereby, fairly miserable. Hence, Sally disvalues smoking but also values the peace of mind that it brings, experiences event-based clashes, and often irrationally chooses to continue to smoke.

Not only can the unwilling cigarette addict's smoking be explained by her way of valuing, she may be able to stop using by turning to her way of valuing, and transitioning to a new way of valuing that puts more passion on the side of not smoking. One means to stop using an addictive substance involves creating new event-based synergies. One set of new event-based synergies would be between the disvalued addictive substance and other things disvalued, together, which could more strongly discourage usage. The other

set of new event-based synergies would be between abstaining and other things valued, together which could make abstinence more appealing. Abstaining through strong cravings is more likely to be successful through constantly reminding one's self of the costs of smoking,⁸ and creating more immediate costs for smoking and immediate rewards for not smoking. Now there is added emotion against smoking and added emotion in favor of not smoking. When the unwilling cigarette addict must face times of cigarette cravings, there is now a heightened disvaluing of smoking. This heightened disvaluing will elicit more unwelcome bodily sensations than previously during and after smoking, such that smoking would bring a lower peace of mind than abstaining would. With this new way of valuing, then, one has reasons backed by greater emotions – one disvalues the smoking more deeply – and one is able to resist cravings for smoking until the cravings and the level of distraction they produce decreases over time.^{9 10 11}

⁸ Sometimes reminding oneself of the costs of smoking alone can be enough, as when one – or a friend – has recently suffered severely from those costs, e.g., one was recently released from the hospital after suffering a heart attack or suffering a near-fatal inability to breathe due to a flare up of one's severe emphysema. Now one's peace of mind is already lost by considering one's impending mortality and this can be a greater loss of peace of mind than the cravings of the addiction itself can bring.

⁹ The physical addiction may never subside completely, which is why one is wise to keep some costs and rewards at play to keep from returning to smoking later. Of course, there may be some physical addictions which never decrease their level of distraction.

¹⁰ I am not trying to claim that all addictions can be so conquered; the distraction and lack of peace of mind may simply be too strong that there is nothing that one can stack against it to make the choice of not using stronger.

¹¹ Another form of weak-willed akrasia would occur if someone forms a better judgment but it does not connect to any VND attitude – it is a judgment without an accompanying emotion-stance or action-stance. Such better judgments are idle. On the valuing theory, the problem here is that such a better judgment is not a part of one's way of valuing and so it cannot lead to action and it is not surprising that it does not.

Consider a student who has taken an introduction to philosophy class, and who concludes based on the arguments discussed that eating animals raised in corporate farm conditions is wrong, but continues to eat meat from extremely cruel farms without any level of guilt or resolve to do differently in the future even when he is reminded of the cruelty on such farms. The irrationality of this form of weak-willed

I heard once of one impressive example of successfully quitting smoking through creating new direct valuing connections. A woman of the Women's Movement told herself that, if she smoked, she would have to give money to anti-abortion groups that she vehemently disvalued because she saw them as anti-women. She never smoked again.

On the valuing theory, an impetuous akratic action is similar to a weak-willed akratic action. An impetuous akratic action involves the absence of a particular better judgment; however, one's actions are explained by more myopic judgments that stem from one's way of valuing. The potential better judgment that would take a greater portion of one's way of valuing into account is unformed. However, one's myopic judgment focuses on a narrower portion of one's way of valuing such that one's reasons behind VaNDing some narrower group of things are seen as providing reasons – e.g., physical pleasure is seen as a reason, end of distraction is seen as a reason – to act. As with weak-willed actions, when someone acts impetuously, an area of internal incoherence or incompleteness is revealed within the person's way of valuing. For the person performing an impetuous akratic action, though, one or more imperfections of a person's way of valuing are revealed to that person *after* the impetuous action is

akrasia has no hope of resolution until, rather than simply thinking eating meat from such farms is of disvalue, the student actually comes to disvalue cruelty to animals.

In all honesty, this possibility of idle better judgments does not seem realistic to me. If someone concludes that some action is wrong, then she likely has corresponding actions and emotions against such actions (especially when done by others), but she finds reasons for why she can somehow participate in (e.g., through helping to create a demand for commercial farm meat) or directly perform such actions herself. Hence, more realistically than idle better judgments, the following happens: After learning of the cruelty on commercial farms, the student transitions to disvaluing such farms, then sees the irrationality of continuing to eat massed produced meats (an transitional event), then she transitions again to a new way of valuing that includes newly seen or found reasons that make her eating of such meat rational (e.g., she cannot afford organic meat but she needs some meat for her basic survival.), even though, if she were to make an even better judgment (as in taking a larger portion of what she VaNDs), she would disvalue eating such meat more deeply again as she did when first learning of the cruelty on these commercial farms.

performed. While the person might not realize *that* or *how* his way of valuing is imperfect, that imperfection is still revealed in how the chosen act causes an event-based clash that might have been avoided. The better judgment comes too late and the person regrets the impetuous action.

A quick example of impetuous akratic action is when Katie unnecessarily hurts herself when trying to stop a small fire caused by a faulty cell phone battery. When Katie sees the fire, Katie's fear of that fire leads her to use her own unprotected hands. She greatly disvalues the fire and its potential to destroy various things valued, which is why she feared it, but had she taken a moment to think, she would have realized that using some object to squelch the fire instead of her hands would have saved her from second-degree burns *and* eliminated the disvalued fire. Here, a small incompleteness in or incoherence of her way of valuing is revealed: either she had not thought through dealing with such a fire, or she disvalued fires too strongly given other things she values, e.g., her hands.

Another example of impetuous akrasia might be the unwilling addict before he or she tries an addictive substance for the first time. Consider a marijuana addict, Bob. In high school, before Bob tried marijuana, he was an awkward teenager and he strongly disvalued his lack of popularity and friendships. In light of many days and nights of loneliness and sadness about being unpopular, when offered marijuana today for the first time by some popular students, he decided then and there to try some of it. Afterwards, though, he regretted it. Suddenly addicted to the substance, his emotional well-being was connected to marijuana use, and that usage cost him money and made him unreliable.

Bob's decision to try marijuana reveals either incompleteness or incoherence.^{12 13} Like the fire victim, perhaps he did not think about how to better handle peer pressure or perhaps he valued his popularity too much given the other things that he VaNDed.

To summarize, while there is certainly a flaw in persons who regularly, or even just rarely, do actions that are weak-willed or impetuous, on the valuing theory, that flaw is the lack of a complete and rationally coherent way of valuing. Akratic actions come from a person's way of valuing, not outside of that way of valuing.

Moods: Akrasia and More

Atypical moods sometimes lead to erratic actions and erratic emotional responses that usually only last as long as the mood. Consider Lisa and her different possible

¹² Sally could have started smoking impetuously but it seems less likely to me that she would regret it soon afterwards, rather than a few years down the road when the health costs were experienced. Instead, I suspect that most people start smoking as teenagers. Most teenagers know of the dangers of smoking but judge that the dangers somehow don't apply to them, thinking: "I am different than those weak-willed people out there who cannot stop smoking whenever they please. I will try this and enjoy it, but if I start to dislike some of its effects, I will simply quit." I suppose that might be an example of an impetuous action, if impetuous actions include those actions that are thought through but just not well enough. See next footnote for more on understanding impetuous action as including those that are thought through but not well enough.

¹³ Someone might think that the willing addict, someone who wants to be an addict, provides an example of an impetuous person. Often, even willing addicts have moments of regret when their indulgence leads to physical pains. However, willing addicts seems to me to be more complex than the self-inflicted burn victim or the marijuana addict who fairly soon after comes to regret it. Not all willing addicts are impetuous, or at least there is some debate about it.

Consider the willing addict who has deliberated and forms a judgment that he should continue to use that to which he is addicted. He would describe it this way: his better judgment is that having and satisfying this addiction supports his way of valuing, though, he occasionally experiences some costs to satisfying the addiction. Of course, therapists, family, and friends might reply that even this willing addict has failed to think through the issue well enough, and that his addiction and continued use is costing him too much, though the addict himself might say that he does not value those things lost to the addiction, that, really, he does still value these things: he is in denial. Those concerned for the willing addict claim that an *even better judgment, based on the addict's own way of valuing*, would show that he should stop using, stop being impetuous, and at least become a weak-willed *unwilling* addict who hopefully later becomes a sober addict.

responses to her daughter's playfulness one Monday morning when, if this playfulness continues at its current pace, both the mother and daughter will be fifteen minutes late.

Lisa would have very different emotional responses to that playfulness if she were in a bad or good mood, but a minimal emotional response if she were in her typical okay mood, as in "nothing-is-bad-or-great-it's-just-okay" mood. In a good mood, Lisa would be laughing at and enjoying her daughter's playfulness and would not be terribly concerned about being fifteen minutes late, and so they would arrive fifteen minutes late. In a bad mood, she would be very agitated by her daughter's playfulness. She would think of how that playfulness will make them late. She would yell at her daughter to hurry up, her daughter would begin to cry, and they would be an extra ten minutes late due to Lisa losing her temper. Finally, in her typical okay mood, Lisa would smile at her daughter's playfulness, mildly enjoying it, but gently move her daughter towards being ready to leave. In her okay mood they would be just on time.

Someone might think wrongly that the mood alters who Lisa is temporarily. Someone could have the intuition that moods *do* change who human beings are; during some atypical moods, a human being is someone else. Sometimes human beings certainly act and have emotions that are very different from their usual selves when they are in an atypical mood. However, I will argue that this intuition, call it the moods-change-us intuition is wrong, though understandably so.

Alternatively, someone else may not share the moods-change-us intuition, and perhaps even have the opposite intuition, the moods-do-not-change-us intuition, but believe that the valuing theory is forced to agree with the moods-change-us intuition. According to this problematic position, which would be an objection to the valuing

theory, some atypical moods alter Lisa's way of valuing in a way that the valuing theory would say is not person preserving. This account might be explained like this: during some atypical moods, Lisa has erratic actions and erratic emotional responses, which reflect that she acquired one or more temporary new VND attitudes. In the good mood, she appears to strongly value the playfulness. In a bad mood, she appears to strongly disvalue it. In an okay mood, Lisa appears to value playfulness only mildly. The moods seem to create different VND attitudes – as evidenced by the erratic actions and emotional responses – and if moods create VND attitudes, that would be a non-transitional change from Lisa's old way of valuing to a new way of valuing which would, of course, fail to satisfy the valuing theory's requirement for person-preserving changes. Rather than a transitional event and the very same person finding new reasons, it appears that an atypical mood produces a new way of valuing. Isn't the valuing theory forced to say that Lisa is not the same person during a mood?

This account of moods is incorrect. It is true that something that elicits no or one set of emotions with one's typical mood can elicit different emotions during an atypical mood. However, these emotions stem from one's way of valuing. In her okay mood, Lisa values being on time, values her daughter's playfulness, values her daughter, and so forth. However, in her typical okay mood, she finds it easier to take all of these VND attitudes into consideration – she makes better judgments and acts upon those. During some atypical moods, one is temporarily inclined to make myopic judgments which focus on some aspects of a state of affairs, aspects that a mood leads one to emphasize in one's thinking. Often moods lead one to take akratic actions, often impetuous ones, which, in

retrospect, one realizes were contrary to one's better judgment, though one may or may not realize that this akratic action was performed in the context of a mood.

If a particular mood is recurrent (e.g., someone gets the Monday blues or rush-hour-traffic-road-rage, but the mood is still atypical because another mood is more common), then that atypical mood might lead to weak-willed actions. Thus, for example, imagine that one tends to fall into a bad mood driving home in the evening back-to-back traffic. Once one might unproductively yell at the kids for being noisy impetuously, which goes against one's unformed better judgment. However, the next evening when they are noisy again and one performs the same action, it is a weak-willed one: one formed the better judgment yesterday and knows that yelling at the kids goes against one's better judgment.¹⁴

A mood could lead to a temporary VND attitude. One makes a myopic judgment and then concludes from that that one should VaND differently. In a bad mood that morning, Lisa might conclude that her daughter's playfulness is not so great, and decide to value it less. However, another time, when she is in an okay mood again, her daughter will do something else playful and Lisa will conclude that she should value her daughter's playfulness more highly. Lisa could vacillate in this way because of ongoing

¹⁴ One could probably act impetuously around the same better judgment more than once. There is a gray area between impetuous and weak-willed actions: someone can forget her better judgment and then act against that judgment. When someone forgets her better judgment that seems closer to impetuosity rather than a weakness of will. Yet, after performing the act, it also seems appropriate to say to oneself, "Oh why did you do that again!?! You know better!" I suspect that the more opportunities that one has had to make or more recently one has made a particular better judgment but one repeatedly forgets that better judgment, the more weak-willed the action appears. In contrast, I suspect that the less opportunities or less recently that one has had to make a particular better judgment – one has forgotten a few times but rarely has an opportunity to act on that better judgment at all, the more impetuous the action appears.¹⁵ Generally, the fact that a human being is not someone else, though, does not entail that we ought to hold these persons using such drugs responsible for their behavior. Instead, the intensity of the cravings that those drugs produce might be grounds to excuse them for wrong actions. Perhaps only a virtuous person could consistently resist cravings at that level of intensity.

moodiness, in which she is in a bad mood off and on, and she may or may not even realize that her attitude regarding her daughter's playfulness is vacillating over time.

When an atypical mood leads to erratic actions and emotional responses, those actions may not be akratic. Occasionally, an atypical mood leads to an action that conforms with one's better judgment. Suppose during any of the three moods, if Lisa would have considered a greater portion of her way of valuing, she would have made the better judgment to act as she would have acted in a good mood: fully appreciating her daughter's playfulness and being fifteen minutes late that day. However, Lisa would not form this better judgment while in her typical okay mood or in an atypical bad mood. In contrast, a good mood that morning would lead her to see that her way of valuing supported her valuing of her daughter's playfulness more and her devaluing of tardiness less. If Lisa were in fact lucky enough to be in a good mood, then, enjoying the playfulness could qualify as an event for Lisa. Her good mood could help her realize that, taking more of her way of valuing into account she should worry less and appreciate the joys in life more.

Even bad moods can sometimes lead to judgments that agree with our better judgments. Consider Ken, who is typically melancholy, and generally too giving and trusting and values himself relatively little. Suppose Ken is in an atypical bad mood, which leads him to angrily reject an arrangement that treats him unfairly. During this bad mood, Ken values himself more, but the mood does not directly cause him to value himself more. Instead, while still in that bad mood, Ken realizes that he is normally unfair to himself; thereby, he comes to value himself more, and so transitions to a new way of valuing. When the bad mood fades, he may still value himself more than before

the bad mood (e.g., he would still not go for a similarly unfair arrangement) though, without a bad mood, disrespect towards him may not elicit the strong emotional reactions that it did while he was in the bad mood. Instead, now that he values himself more he has a new emotion-stance or action-stance that reflects this increased self-valuing. However, back in his typical melancholy mood, he may still only have mild emotional reactions to others disrespecting him.

As some of my examples suggest, a person may not realize that he or she is in a mood or that it has led to erratic actions or emotional responses. Instead, one might only notice that one's erratic actions or emotional responses are erratic, or one may even fail to notice that they are erratic.

To summarize, moods can be frustrating. However, this entire discussion of moods is meant to explain how moods do not lead to the displacement of one person for another. During an atypical mood, on the valuing theory the correct analysis of moods explains that when a person has erratic actions or emotional responses during an atypical mood, the person transitions from or expresses her way of valuing, given the myopic judgments that the mood encourages.

Moods – a tentative analysis

One might wonder, 'How do moods encourage myopic judgments?' I do not know that I can give a full answer, and I will need to speculate to some degree. In my speculation of how moods encourage myopic judgments, I will note the importance of one's way of valuing in providing the content of those myopic judgments.

It seems to me that a mood is itself a cluster of bodily sensations which are similar to an emotion or are the actual lingering bodily sensations of an emotion. Therefore, if

one is in a bad mood, one has the cluster of bodily sensations that occur when one is angry. However, a mood can have many different sources, some non-emotional sources. A bad mood – a cluster of bodily sensations which also partly constitute the emotion of anger – can be caused by a number of sources. A few sources of a bad mood are an enraging encounter, a pain in one's leg, hormonal shifts, side effects of pharmaceuticals, severe hunger, or severe dehydration.

I propose that a cluster of bodily sensations that is similar to the cluster of bodily sensations of an emotion is called a mood (even when that cluster is in fact the lingering bodily sensations of an emotion) when one does not remain constantly focused on the source of those bodily sensations. Failing to focus constantly on the source of these bodily sensations (or actually not knowing the source of these bodily sensations), one searches, sometimes fairly subconsciously, for things which could elicit those bodily sensations, thereby, searching for a justification for why one feels as one does. Thus, in a bad mood, one subconsciously searches for a justification for why one “feels angry”: the cluster of bodily sensations which come with anger feel like anger, even though their source might be a pain in one's leg. One might realize or be reminded that the source of the bad mood is an enraging encounter. Alternatively, one might realize or be reminded that the source of the bad mood is some non-emotional cause, such as hormones or pains. Finally, one may never figure out why one is in a bad mood. Whether one does or does not realize the source of the mood, if one does not get rid of the bodily sensations that comprise that mood, then when one turns one's attention elsewhere, one is still in that mood, and one is still searching for an answer to the question, “Why am I feeling as I do?” As a result, if one remains in a bad mood, one keeps asking, again perhaps

subconsciously, “Why am I so angry?” In trying to answer this question, one looks for angering things. In looking for angering things, one sees what is angering in various states of affairs but fails to give equal attention to what is not angering, but endearing or somehow positive. One is engaged in forming myopic judgments.

At this point, one’s way of valuing plays a central role. One’s way of valuing defines what can be angering to a person; what is enraging to one person might be amusing to another. Suddenly, in a bad mood, one is engaged in myopic judgments which stem from one’s own way of valuing and focus on things that indicate that the world is unfairly undermining things valued or unfairly promoting things disvalued. These judgments overlook how many other things in the world are going well according to one’s entire way of valuing, so that, on balance, a better judgment is that the world is no better or worse than usual. Similarly, a sad mood would be one that promoted myopic judgments that focus on how the world is not going well according to some portion of one’s way of valuing and how one is limited in what one can do about this unfortunate state of affairs. A fearful or anxious mood would be one that promoted myopic judgments that focus on how some of what one values may be in harm’s way (though part of the myopathy here is not realizing how even the object of one’s worry is not in terrible danger on balance, e.g., that one is treating horrible, but remote, possibilities as likely outcomes). A happy or good mood would be one that promoted myopic judgments that focus on how what one values is being promoted and, perhaps, how what one disvalues is being undermined. As I have already said, while one’s judgments tend to be myopic while in a mood, those myopic judgments might coincide with a better judgment, and ultimately lead one to form a better judgment. If one’s judgments tend to be myopic in

one direction already (recall Ken with low self-valuing who thought ‘I’m not worth much’; ‘I must care for others over myself’), and an atypical mood promotes a myopic judgment in a different direction (‘I deserve MUCH better’), then one has an opportunity to consider all of these judgments, and find reasons to VaND differently (Ken values himself more). **Two further points on moods**

Finally, I would like to discuss two further features of moods. First, one’s typical mood is, in part, due to one’s way of valuing, though a typical mood can (and perhaps must) also have additional non-emotional sources. One’s typical mood depends on (1) one’s (likely transitioning) way(s) of valuing and (2) likely some stable degree of flexibility in the range of bodily sensations that one can readily experience. One’s typical mood might be a bad mood because one values many things that one regularly sees as being unfairly undermined. At the same time, that typical bad mood cannot easily be sustained through hormones surges that promote a state of relaxation, which is why there are also atypical moods. Moreover, if one’s typical mood is a bad mood, then one is unlikely to be able to continue to have that as one’s typical mood if there is a shift in hormonal balance (which is not reversed for decades) that encourages a state of relaxation in otherwise healthy human beings. One’s typical mood, then, would seem to depend on one’s way of valuing but also the ability of that human being (of which one is a part) to have various bodily sensations.

Second, I would like to clarify that I do not mean to suggest that moods *necessitate* myopic judgments; if one is very aware of one’s thinking and mood, one can watch and correct myopic judgments and sometimes simply avoid them all together. However, one may need to keep *constant* focus on one’s mood as well as how well one is

thinking to be aware enough to avoid myopic judgments, as opposed to simply thinking about various things. In other words, in order to avoid the myopic thinking that moods encourage, one may need to keep thinking about how one is thinking, as opposed to simply going about one's thinking. This level of attention to the quality of one's thinking is difficult to maintain. When thinking occurs unaccompanied by critically evaluating about how well one is thinking, one begins or returns to searching for a justification for how one "feels," and so the atypical mood promotes myopic judgments.

To summarize, moods do *not* temporarily make a human being another person, and furthermore moods must work with a way of valuing as the source of particular myopic judgments. Those myopic judgments come from one's way of valuing and they might be behind akratic actions, though occasionally a mood gives a person the opportunity to realize that her better judgment is supportive of a greater number of current VND attitudes or of forming new VND attitudes, the latter which are at least closer to the emotional reactions that an atypical mood encourages.

Conclusion & Depression, Mania, Psychosis, and Shifting Appetites

This analysis of moods shows that sources of moods, such as drug side-effects, if they are only mood-altering, do not make a human being someone else.¹⁵ My discussion

¹⁵ Generally, the fact that a human being is not someone else, though, does not entail that we ought to hold these persons using such drugs responsible for their behavior. Instead, the intensity of the cravings that those drugs produce might be grounds to excuse them for wrong actions. Perhaps only a virtuous person could consistently resist cravings at that level of intensity.

Moreover, it is not clear to me that the virtuous person could consistently resist extreme cravings because the virtuous person is someone that has trained himself to have passions that agree with his better judgment, but that self-training only involved modification of or elimination of a sub-set of possible passions, passions that he had and perhaps those about which he sees others regularly grappling. He has not trained himself to withstand any and all cravings that could possibly be made to arise in him, e.g., cravings for a drug never taken. The virtuous person's better judgment would likely be to avoid mood-altering drugs, but it is not clear that he could simply have virtuous passions instead takes a drug with such

of moods also suggests that periods of longer lasting mood-states such as depression and mania, do not make a human being another person. Instead, those periods promote more myopic judgments ('There is nothing worth living for.') rather than better judgments, and may even lead the person to make a number of transitions which leave the person with a greater number of false judgments ('There is nothing worth living for.' 'No one loves me.' 'The people I value will die suddenly without warning and the pain will be too great to bear.' 'No one will miss me if I'm gone anyway.') These false judgments are false in that they do not reflect inter-subjective reality.

While these judgments may be false, it may be incorrect to call such judgments myopic. Suppose that because these mood-states were long standing, the person has made a number of transitions. Suppose further that his most recent way(s) of valuing are such that, when one takes *those* entire ways of valuing into account, *those* ways of valuing would wholly support the judgments encouraged by the mood-state, then those judgments, while false, would not be myopic.

The possibility of false judgments moves us into the possibility of disconnection from inter-subjective reality. On the valuing theory, becoming severely disconnected from inter-subjective reality (i.e., psychosis) does not temporarily make a human being a different person. It does not matter whether the source of that psychosis is genetic, pharmacological, accidental, or some other source. Instead, during a psychotic episode, the same person responds as his way of valuing dictates based on the false reality that he

side-effects without realizing that it has such side-effects. My point here is similar to the point that some make about how the virtuous person would not necessarily perform the right action or be happy after being horribly tortured. (E.g., Korsgaard, *Self-Constitution*, 178-180 and Lawrence Becker, *A New Stoicism*, (Princeton, NJ: Princeton University Press, 1999): 146-148.)

experiences. A severely false reality gives the person grounds to transition dramatically from the way of valuing he had when last connected to reality.¹⁶ My general position on psychosis is this: no matter how surreal and impossible this false reality is given our current understanding of physical law, suppose this false reality could actually occur. Consider how this human being, as a particular person, would react in two scenarios: (1) a particular surreal state of affairs is made real and (2) that particular surreal state of affairs is only “real” in that person’s own imagination. If the person would respond to these two scenarios in the same way, then that human being is the same person during psychosis as he or she was before psychosis. Finally, the discussion of akratic action due to appetites points to the fact that an altered appetite – any innate or acquired appetite that one gains or loses or that increases or decreases – does not create another person: when those changes in appetite begin, the differences in cravings provide reasons for the same person to take erratic actions or have erratic emotional responses. Take any increased or decreased interest in food, drink, sex, drugs, nicotine, and so forth and whatever that change’s cause, speed, duration, and so forth. A change in appetite, in and of itself, will not displace one person for another. Instead, initially at least, that change may lead a person to form myopic judgments that stem from that person’s way of valuing (Way_0) or make a transition to a new ways of valuing that stems from Way_0 .

If one has the moods-change-us intuition or the parallels – psychosis-changes-us intuition or altered-appetites-change-us intuition – these intuitions are completely understandable. We certainly speak this way. Regarding an altered appetite plus atypical

¹⁶ This seems true to me even if the source of psychosis also creates moods. I do not see how if moods and psychosis separately do not create new persons, that both together create new persons.

mood-state, we might say, “On cocaine, he was another person,” or regarding a human being with psychosis, we might say, “Off of his schizophrenia medication, he is another person.” Given that typical moods as well as some acquired appetites partly owe their existence to one’s way of valuing, then one appears to be someone else when a typical mood or typical appetite is displaced. Atypical mood-states, disconnected realities, and altered appetites only metaphorically create new persons because each of those lead a person to see or find reasons to take erratic actions or have erratic emotional responses as well as see or find reasons to make transitions to new ways of valuing.

Radically different experiences for a person that *are* connected to inter-subjective reality do the same thing; unusual experiences lead a person to see or find reasons to take erratic actions or have erratic emotional responses as well as see or find reasons to make transitions to new ways of valuing. Transitions that are based on radically different experiences than the person has experienced previously can appear to displace one person for another. We might say of a widower, “Since his wife died, he is no longer the same person.” This is only metaphorical: the widower, a particular person, sees or finds grounds to transition from one way of valuing to another, seeing and finding those grounds in the way of valuing that he had upon his wife’s death and the radically different (and disturbing) experiences of life without his wife. Thereby, the widower *changes* himself, but *remains* himself.

As with transitions based on radically different experiences, such talk of human beings becoming someone else due to atypical mood-states, disconnected realities, and altered appetites is only metaphorical. Just as with radically different experiences, with atypical mood-states, disconnected realities, and altered appetites, the person sees and

finds reasons based on his way of valuing (as that person's way of valuing dictates) and, thus, the person persists.

Even though extreme atypical mood-states, disconnected realities, and extreme altered appetites are not person-destroying, these things often pose problems. They often encourage a person to move away from her better judgments as well as give her appetites or a false reality that are often unwelcome.¹⁷ Perhaps most importantly, they make it difficult or impossible for such persons to seamlessly interact with and fully engage in various types of relationships with other persons. Barring other problems, each human being who has psychosis, drug addiction, or severe depression or mania, remains the same person. However, we may be wrong to hold some of them morally responsible for a significant portion of their actions. Some may be individuals who cannot be relied upon as those within a community, a family, or a friendship are relied upon to agree to and live up to various obligations intertwined with privileges. Hence, due to extreme atypical mood-states, disconnected realities, or extreme altered appetites, some persons on my analysis may no longer be full moral agents.

While some may identify personhood with full moral agency, this project offers a different analysis of personhood. According to the valuing theory, the human being is a person, if he or she has the capacity to value. Using that capacity to value, the human being will be a particular person with a way of valuing and that particular person will persist if he or she continues with that way of valuing or transitions to a new way of valuing through the exercise of her capacity to value. I assert that even though those with psychosis, drug addiction, severe depression or mania may not be full moral agents, as

¹⁷ An example of an unwelcome appetite would be the addiction of an unwilling addict.

valuers, they deserve our respect. As with any VND attitude, what respect entails varies. This respect will not always be in the form of honoring their wishes or doing as they ask, since we will conclude that those wishes, though the person himself may not realize it, will lead to unjustifiable harm to other persons or the person himself.¹⁸

¹⁸ These last statements about respect for persons, even those persons who are not full moral agents, are in need of a defense, which I will not provide here. I believe that the capacity that I have described in this project is one that makes another deserving of respect such that the wishes of such individuals should be taken into account, though, again, not necessarily honored. Fully defending the position that valuers deserve our respect is something that I may take up in future work.

CONCLUSION - PROJECT SUMMARY AND CONCLUDING REMARKS

Summary

In the preceding chapters, I have presented a new analysis of personal identity over time and an accompanying definition of a person. This analysis of personal identity is meant to help address practical questions which partly turn on whether we have the same person, given that we have the same human being. On this theory, persons are valuers: they have the capacity to value which involves seeing and finding reasons to have valuing or disvaluing attitudes (as well as possibly some neutral attitudes) – VND attitudes – towards various things. Each VND attitude has an emotion-stance and a likely action-stance and each attitude is partly shaped by the other things that one VaNDs.

Some things are VaNDed in virtue of their connections to other things VaNDed, and the reasons for each are connected, e.g., because yoga promotes one's valued health, one values yoga for the reasons that one values health. These connections reflect our tendency towards rational unity. Other things VaNDed are connected in virtue of the fact that some possible event would lead to an emotional clash or synergy between them, e.g., Mary values her completion of two incompatible projects, which is why she is regularly miserable. Often when she works on one project, she is upset about undercutting the other project, and vice versa.

The reason-based and event-based connections form an interconnected set of VND attitudes – a way of valuing – which constitutes synchronic unity at a time. Ways

of valuing do not exist for only an instant. Diachronic unity is constituted partly by the persistence of a particular way of valuing. Then, diachronic unity continues, if an event occurs in virtue of which the person with a particular way of valuing finds reasons, based on his/her way of valuing, to modify old or form new VND attitudes and thereby transition to a new way of valuing. I called these changes to new ways of valuing based on the older way of valuing ‘transitions’ to note that they are the type of changes in ways of valuing that are person-preserving rather than person-destroying.

Strengths over relevant views discussed in Section I

I hope to have presented an analysis that appears stronger than those of Parfit, Schechtman, Rovane and Korsgaard. Parfit argued that the amount of similarity in psychological features from day-to-day is the fundamental building block of personal identity. However, it is not the amount of similarity that matters but the source of that similarity or lack thereof. The valuing theory focuses on how changes occur, not how many occur. Sometimes one will see grounds that lead to one very large transition or many small transitions but which together account for a large change over a short period of time.

Korsgaard, Schechtman, and Parfit assume that human being defines the borders of persons. Parfit assumed that we are able to pick out one person at one time to compare to another person at another time. Schechtman claimed that a self-narrative is self-constituting if it is accurate, when accuracy is satisfied if it includes any part of a human being’s life. Korsgaard assumed that having one human body and life – if it has personhood – creates a practical necessity for that human being to unify as one person. However, each of these positions assumes that there is at most one person per human

being. The valuing theory does not exclude the possibility of multiple persons within one human being, such as Boris the human being who becomes another person after a brain injury.

Schechtman, Rovane and Korsgaard realize that it is a psychological *capacity* that is definitive of persons and their identities. Schechtman's chosen capacity is the capacity to self-narrate, but this fails to include some individuals those who, pre-theoretically, are persons: Episodics do not narrate and Diachronics do not always narrate. If Episodics or all Diachronics are included, then narration is optional, and this suggests that self-narrative may be sufficient but not necessary for personal identity.

In contrast to Schechtman, the valuing theory picks out the capacity to value, a capacity which appears to be active whenever a human being is making choices and perhaps whenever one is conscious. This capacity appears to match what Schechtman was looking for when she sometimes distanced herself from self-narrative, while still wanting to call her view a self-narrative view, acknowledging that often self-narrating does not occur and claiming that even subconscious aspects of a person can qualify as parts of that person's self-narrative:

I realize that this is a somewhat unusual way to think about "self-narrative," and that I may not have satisfied everyone that it is a legitimate one. In the end, however, very little really turns on this choice of words. It does not matter much whether we say that identity is determined by a person's self-narrative or by his psychical organization, so long as it is understood that the psychological forces constituting identity are dynamic and active – things a person *does* – rather than static and passive features she *has*. I use the term "self-narrative," even though it is somewhat controversial here, to underscore these features of a person's psychological life.¹

The valuing theory would seem to satisfy Schechtman's hopes for the self-narrative view. Seeing and finding reasons for VND attitudes, expressing those VND

¹ Schechtman, *Constitution of Selves*, 117.

attitudes with emotions and actions, changing those VND attitudes based on one's current way of valuing are all active and dynamic activities -- something persons *do* rather than being passive features they *have* – and, unlike self-narrating they do these things all of the time when they are conscious. On the valuing theory exercising the capacity to value is the way that persons create their own identities. Persons do not always have control over what appetites or moods they have, but, given appetites or moods, they see or find reasons to value, disvalue, or take a neutral stance regarding various things and thereby express a current way of valuing or transition to a new way of valuing. The interconnected set of VND attitudes that result defines a person's identity for a stretch of time and then a stream of valuing which the person creates himself defines the person's identity over time.

Both Rovane and Korsgaard pick reason and rationality as the capacity which unifies persons. While Korsgaard's ideas were promising, her view was preliminary and then she took future work in a different direction, focusing on what makes for good unification of persons, rather than what qualifies as unification at all. Rovane's view is a fully developed view on personal identity which focuses on rationality. I argued that the capacity to reason or even striving for perfect rationality is insufficient to provide synchronic or diachronic unity. Sometimes a person keeps knowingly incompatible projects over modifying those projects and does not criticize herself for this failure to achieve overall rational unity. Sometimes a person prefers irrationality to giving up one of one's projects. Mary, the woman who wanted to be the kind of CEO and the kind of mother that would require 36 hours a day, realizes that her projects are incompatible, but she refuses to drop either one. On Rovane's analysis of persons this human being Mary

thereby dissolves into multiple persons within one body (at least if and until those projects are modified to take one another into better account).

Narrativity is likely sufficient but not necessary for personhood, and reason and rationality are *one* source of unity for persons, but our emotional investment, even in projects that do not rationally cohere, provides another source of the unity of persons. Mary's misery, though she is doing one project well because the other is undermined simultaneously, is due to the fact that she is one person, though, admittedly a fractured one on the rational level.

Lingering Objections and Replies

Rationality is sufficient, Part i

At this point, though, one might think that the intuitions of Rovane and Korsgaard are correct, though their analyses may be somehow flawed; the capacity to reason alone is sufficient to be a person.

A purely empirical basis for skepticism that the capacity to reason alone is sufficient for personhood comes in the form of cases in which patients suffer brain damage to their emotional centers or to the connection between their emotional and reasoning centers of the brain, while the capacity to reason remains intact.

Elliot, one patient described by Antonio Damasio,² had a benign tumor in the area of the brain responsible for emotions. Its removal left him able to reason well about many things, but unable to carry out any plans in any coherent way. Elliot went from someone who was reliable at work to someone who would fail to stay on task, focusing

² Antonio Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Penguin Books, 1994. 'Elliot' is an alias.

on relatively unimportant details for hours. He would read articles in detail that his job required that he only skim in order to do some larger more important task, or focus for hours on how to best alphabetize papers, when that decision would have previously taken him a few seconds to make and a few minutes to implement.

Mr. Marvin Bateman, another patient that Demasio discusses,³ suffered a stroke that left his emotional and rational centers of the brain intact but killed the neurons that allowed for communication between them. His body would have bodily sensations of emotions in response to exposure to things he once valued, e.g., he would smile at old photos of loved ones and special events. However, while Mr. Bateman remembered these individuals and occasions, he reported no emotion in response to these photos. In terms of decision making and carrying out plans, he, like Elliot, was unable to make decisions, prioritize, and carry out even simple plans.

The cases of Elliot and Mr. Bateman suggest that loss of emotion to accompany reason appears to transform a human being into something similar to a Humean bundle, in which there are competing suggestions of reason but an inability to prioritize them in order to formulate projects and plans. The preliminary empirical evidence suggests that emotion appears to be necessary for the personhood of human beings. Moreover, Mr. Bateman's case suggests that emotion needs to be integrated with reason. Elliot and Mr. Bateman cannot be said to have projects or plans. Instead they can merely reason about plans or projects and the pros and cons of various projects or plans. They cannot make decisions and others must decide for them.

³ Demasio discussed Mr. Peter Bateman in "Emotions and the Brain" of "Episode Four: The Adult Brain: To Think by Feeling" in *The Secret Life of the Brain*. DVD. Produced by David Grubin. 2001. New York, NY: PBS Video, 2010.

Consider Kant's two formal principles of reason.

Hypothetical Imperative: Take the means to one's end or give up that end.

The Categorical Imperative: Act only on that maxim that can be willed as universal law.

Elliot and Mr. Bateman have no ends, so they cannot take any means. They can consider maxims but they *have no maxims* to consider because they can never decide which maxims are potential candidates for actions that should then be run through the universalizability test. They can run the universalizability test but neither has a maxim that he has decided is a maxim on which he might act. Lacking plans and projects, it is difficult for me to see how these individuals qualify as agents and so as persons.

Rationality is sufficient, Part ii

An objection related to the intuition that rationality alone is sufficient for personhood leads to skepticism regarding two claims in Chapter 3. In Chapter 3 I claimed that there cannot be persons who had *only* neutral attitudes towards various things or who do not evaluate anything – no VND attitude was had about anything. Someone could think that, intuitively, persons can be completely neutral to everything – neither valuing nor disvaluing attitudes towards various things – or have not VND attitudes towards anything: emotions are only an optional dimension of our personhood.

My response is that I could imagine that a person had a *better* or *all-things-considered* judgment that nothing was of value or nothing should be evaluated. However, without having more myopic judgments that support actually valuing and disvaluing various things, it is difficult to see how or why such so-called persons would *choose* to do anything. Perhaps animal instincts would take over in getting basic needs

met, but then reason is not in charge of behavior or emotions. I am left wondering how a human being would choose to do anything without valuing and disvaluing attitudes. On what maxims would they act? These hypothetical persons – one with only neutral attitudes and the other with no VND attitudes – appear to be similar to Elliot and Mr. Bateman. Elliot and Mr. Bateman are certainly neutral or non-valuing towards everything, but to me this stance is at the cost of their agencies.

Emotions are sufficient for valuing

The idea of a human being running purely on instincts connects to a different objection that asks, “Aren’t non-human animals valuers?” This objection suggests that this project has hijacked and misapplied the use of verbs such as “to value”. Non-human animals value and disvalue too. This is seen, for instance, in how they protect their offspring.

My response is that there is no denying that non-human animals have attitudes with emotion-stances and action-stances; they are happy to receive food but frustrated to be denied food – say if that food is just out of reach, and attempt to get that food out of reach. They care for their young, seem to enjoy caring for them, and protect them. While there is no denying non-human animals having attitudes, to call those valuing, disvaluing, or neutral attitudes is, in my mind, to go too far, (except perhaps in the case of other great apes). Few or perhaps no non-human animals value or disvalue. The verb ‘to value’ does make me think of emotions, but also an evaluation or the use of reason as well.

This, though, is likely a semantics debate. I respectfully will stay on this side of the semantic debate about the use of verbs “to value” and ask that this project be judged

on the following: whether it picks out a capacity, which involves integrated capacities to reason and have emotions, which is definitive of persons and enables them to persist. If one takes issue with my use of the verb “to value,” this project can still be judged on whether I have properly picked out a capacity that separates healthy adult humans from most or all healthy adult non-human animals. Isn’t this capacity that I have described something that separates non-human animals from healthy human beings?⁴ **Memory**

Someone might object that the valuing theory ignores the importance of memory. Aren’t some intuitions about when a human being is not the same person and other intuitions about when a human being is the same person turn on memory loss or retention, respectively?

This is an area of future work. Memory deserves an entire chapter. On the valuing theory, a loss of memory is one possible indication that a person fails to persist. In future work, I would defend the claim that when we think a person persists despite memory loss, the valuing theory can explain those intuitions. Often a person forgets something in the relevant moment because he is focused on some other events that occurred recently. Moreover, forgetting *one* VND attitude is likely unproblematic given that the rest of the way of valuing can restore easily that VND attitude to its rightful place – this is similar to how a building wall might lose a cinder block at its foundation and still stay up. But how many more VND attitude losses would bring down the whole way of valuing will depend on the particular things VaNDed and the connections between those things. It is not a matter of more or less but a matter of that particular way of valuing and

⁴ Korsgaard and Rovane make the mistake of thinking that some psychological capacity which makes human beings distinct – rationality – is *all* that is necessary for personhood, while I am claiming that emotions – something we share with non-human animals – is necessary as well. Schechtman is making a similar mistake with narrative, though, as Galen Strawson shows, narrativity does not even include all healthy adult human beings.

its interconnections. These are all conclusions to arguments not fully formed and not defended here.

In a chapter on memory, I would also argue that when we think a person does not persist due to memory loss, the valuing theory can explain those intuitions as well. Complete memory loss is a problem because the way of valuing is gone. Dementia remains perplexing for me, in part because there may be less agreement between the intuitions of different people regarding the personal identities of human beings during stages of dementia.

In Closing – The original practical questions

The motivation for this project was to help answer some moral questions whose answers partly turn on whether a given human being is the same person. Hence, this project provides a practical, non-metaphysical analysis of personal identity. Now that I have described and defended the valuing theory, what can I say about Margo?

Margo was the Alzheimer's disease patient who Ferlik observed during the late stages of her disease.⁵ After formulating the valuing theory, I realize that Margo's case is too complex to easily apply my view. I find myself extremely uncomfortable saying that Margo is not a person. There are simply too few details in Ferlik's account to enable an assessment of whether she is a person and a valuer. Ferlik probably made many observations that he does not summarize in his *one-page* description of Margo. All the things that Margo enjoys could be valued by a person: repainting the same thing every day, eating peanut butter and jelly sandwiches, skimming around a novel (especially an

⁵ Margo is discussed in this project's Introduction as later discussed in Chapter 2's section on the question of this project.

old favorite), or re-listening to favorite music. Many Alzheimer's disease patients are repetitive in this way, but are all repetitive about the exact same things? And if her repetitiveness is unique to some extent, could this suggest the persistence of a person? I am not saying that I know Margo *is* a person. I simply do not have enough information from Ferlik's brief account to know whether she is a person and, further, whether she might be the same person as she was years ago. Ferlik has no knowledge of what she was like before her illness and during the earlier stages of her illness.

In order to properly apply the valuing theory of personal identity to a progressive disease such as dementia, future work needs to be done. That work will involve me speaking to professional caregivers of dementia patients and even observing such patients myself. On a personal note, my 90+ year old grandmother had mid-stage to late-stage dementia. She could not keep track of a conversation. About every 90 seconds, she would ask again who I was and then who my two-year old daughter was, but she seemed to fit my account of a person, VaNDing various things, e.g., valuing our company, disvaluing living in a nursing home. Since I only saw this grandmother a few times in my life, I certainly cannot tell whether she was the same person as before the illness, a different person, a non-person, or perhaps, a fading person, whatever that would mean. My knowledge of Margo is even less.

My grandmother's inability to keep track of the conversation is certainly not decisive. Individuals who suffer from transient global amnesia go through a 90 second loop over a one or two day period, where they cannot remember what has happened 90 seconds before, reacting to the news of their situation the same way every 90 seconds. Moreover, each patient has their own unique 90 second reaction. The 24 hours or so that

this lasts is not a time when the personhood of the human being is dormant or another person emerges. Instead, the very same person as before the amnesia is stuck in a 90 second loop. An emergency room doctor⁶ describing transient global amnesia found this 90 second endless loop disconcerting; he thought it suggested that we were machines who lacked free will. I find the 90 second loop comforting. I do not want any form of amnesia, but, to me, the loop shows that persons have an interrelated set of VND attitudes, from which they see and find reasons. As a person with a particular way of valuing, this means I will react as that way of valuing dictates, as I see fit. If a human being has the same way of valuing each time, that human being, according to my theory, will react the very same way to the very same set of circumstances.

I wonder if dementia is similar to this form of amnesia and involves 90 second segments of persons, but it may be a different person each time, with one or a few VND attitudes which are similar to but not exactly the same as the person who existed before the illness began.⁷ At some point in the illness, there is not enough mental capacity to

⁶ “Loops,” *RadioLab*. Season 10, Episode 3, Podcast, <http://www.radiolab.org/2011/oct/04/> (Most recently assessed March 17, 2012.)

⁷ The object and charge of the VND attitudes could be the same but those attitudes would be different because those VND attitudes would not be informed by the original set of VND attitudes. For example, consider Gladys,* a real Alzheimer’s patient who, in the middle stages of the disease after she no longer could recognize that she had problems with her memory, took up smoking again, often became very cross with her four-year-old grandson, and was angry at her husband for putting things in the wrong place. In one respect, she might seem the same person: the person before the illness likely (1) valued the benefits of smoking even after she stopped smoking, (2) disvalued disrespectful behavior of a young child, and (3) disvalued her husband placing things out of order. On the other hand, the VND attitudes that she has in the middle stages of the disease are mere shadows of similar VND attitudes in the larger way of valuing had by the person who existed before the illness. Before the illness, that person found reasons to conclude that the benefits of cigarette smoking provided insufficient grounds to continue to smoke. In addition, the person before the illness would have recognized that her own self-centered behavior reasonably frustrated her grandson such that it would be reasonable for him to respond to *her* behavior with otherwise disrespectful behavior. Finally, the person before the illness would have realized that her own problems with memory were the source of her inability to find things where she expected them and her husband was neither willfully nor neglectfully putting things out of order.

enable them to form better judgments. When they turn their mind to another consideration, it is not possible to hold the original consideration in their mind in order to come to further conclusions, better judgments. The point at which we might say that a middle or late stage dementia patient is no longer a person may be when that patient is unable to form better judgments, and that patient's mind lacks a complexity found in the ways of valuing of healthy human adults. Is there some kind of complexity that is necessary within a set of VND attitudes such that it forms a person's way of valuing? This is something that I need to consider, and learning more about the lives of dementia patients might aid me in defining whether there is some kind or level of complexity that is essential. Currently, I am quite torn here.

It seems worth mentioning again that while narrativity is not necessary for personhood, it seems to me sufficient. I question Schechtman's requirement that narratives must not be terribly inaccurate. One social scientist found a dementia patient, whom she called Ms. Fine, with a wholly fabricated self-narrative.⁸ In this narrative, Ms. Fine forms a coherent but false narrative explaining why she was in this tiny room of her

This description suggests that Alzheimer's disease, once the patient no longer realizes he or she has problems with memory, brings about an ever-changing psychotic state where the patient can only recognize a few aspects of inter-subjective reality but not enough aspects of inter-subjective reality to have a sufficient inter-subjective understanding of reality for moral agency. In other words, she could not take into account enough facts in order to be held responsible for her response to the facts that she did recognize. To the extent that Alzheimer's disease leads to something like psychosis, I would lean towards saying that the same person persists. Alzheimer's disease is not like psychosis though in that the misunderstanding of reality keeps shifting. It is not like a patient who continues to think himself Napoléon. Instead each time the patient's attention shifts, the delusion is very different, based on only some, not all aspects of reality.

*Gladys discussed in *The Forgetting*, directed by Elizabeth Arledge, DVD, PBS: 2004 also at <http://www.pbs.org/theforgetting/watch/index.html> with transcript at http://www.pbs.org/theforgetting/watch/transcript_english.html. Last accessed April 5, 2012.

⁸ Athena Helen McLean, "Coherence without Facticity in Dementia: The Case of Mrs. Fine" in *Thinking about Dementia: Culture, Loss, and the Anthropology of Senility*, 157 – 179, ed. Annette Leibing and Lawrence Cohen, (New Brunswick, NJ: Rutgers University Press, 2006)

nursing home. This self-narrative was completely confabulated from the various pictures she found of herself in that room. Her narrative: her husband thought she was cheating on him due to a suggestive photo that someone fabricated and he kicked her out, and her cousins had stolen her money. The truth was that her husband had died twenty years prior and then, ten years after his passing, she had fallen in love with another man (there is a portrait of them smiling warmly together) who passed away eight years later, which was two years prior to her entering the nursing home. When she was no longer able to care for herself in more recent years, she was placed in a nursing home in a tiny, but private, room, which was extremely expensive and her remaining wealth was paying for that relatively nice room. Schechtman would claim that this individual was not a person because due to the narrative's inaccuracy, that narrative was not self-constituting. Just as with Margo, I could not say whether this patient is now the *same* person as before the illness. However, being able to form such a narrative at all, narrating at all, suggests to me that she *is* a person, who valued her husband, valued various comforts, disvalued her relatives, valued her reputation – the interviewer believed that Ms. Fine wanted others to know that she did not do something to deserve this drab living situation.

While dementia still leaves me perplexed, in the project, I have said a few things about applications. Transient and longer standing moods (e.g., depression) and shifting appetites do not make a human being a different person. Disconnection from intersubjective reality does not make a human being a different person. Permanent retrograde amnesia – when a human being can retain the ability to understand and speak language, but cannot recall anything prior to a brain injury or defect – results in a different person. (Boris, in the introduction to Section II, can be understood as a case of permanent

retrograde amnesia.) In contrast, transient global amnesia with its 90 loop does not make a human being someone else.

Dementia is still perplexing for me. With dementia, when is it ever the case that a new person or non-person emerges or with dementia does person simply fade (and what would the last option mean under the valuing theory)? At the end of this project, I am still left with Ferlik's question, whose answer I intend to continue to pursue: Who *is* Margo?

APPENDIX A: THE CAPACITY TO HAVE EMOTIONS

Introduction

I am defining the capacity to have emotions using a composite analysis of emotions. A composite analysis holds that emotions are a combination of judgments and feelings.¹

Feelings: Bodily Sensations

On a composite analysis, emotions are partly feelings. By ‘feelings’, I mean bodily sensations.² By ‘bodily sensations’, I mean experience of anything as occurring in

¹ The work of Robert Solomon, who was considered a cognitive theorist of the emotions, is from where I draw some of the ideas for this composite analysis of emotions. I would have simply called it a Solomonian analysis of emotions, but hesitated to do so for two reasons. First, Solomon had a large body of work, some of which I am unfamiliar. So I do not wish to claim that I am providing an accurate account of his view, without reviewing all of his work. Second, some of what I say here draws from somatic accounts of emotions, namely, Jesse Prinz, “Embodied Emotions,” in *Thinking about Feeling: Contemporary Philosophers on Emotion*, ed. by Robert Solomon, 44 – 58, (New York, NY: Oxford University Press, 2004). Still other ideas come from other philosophers. A composite view combines the insights of somatic theorists and cognitive theorists but the composite view I describe does not necessarily reflect Solomon’s view, which is sometimes described as a cognitive view, though Solomon seemed to take issue with this label for complex reasons (Robert Solomon, “Emotions, Thoughts, and Feelings: Emotions as Engagements,” in *Thinking about Feeling: Contemporary Philosophers on Emotion*, ed. by Robert Solomon, 76 – 88, (New York, NY: Oxford University Press, 2004). However, I here acknowledge that my sketch of a composite analysis of emotions is based on the work of Solomon and others and do not wish to suggest in any way that these various ideas on the emotions are original to me.

² It turns out that parts of the brain can cause bodily sensations, so that one can have emotions, without the body itself being the cause of the bodily sensations. (See Demasio, *Descartes’ Error*) But for simplicity, here I say that emotions are partly constituted by bodily sensations, rather than saying more accurately that emotions are partly constituted by bodily sensations, though the experience of these bodily sensations sometimes may not have as their source actual processes in the body, but rather the brain’s conveying the inaccurate message that a number of things are occurring in or to the body. This seems to open up a can of worms, though, in that one may wonder if emotions often partly consist of the brain’s inaccurate perception of what is occurring in one’s body.

The brain’s ability to have bodily sensations even when those does not match up with what is actually occurring in the body certainly can go wrong, e.g., pain in a phantom limb, which do not convey accurate information that this limb needs attention because said limb was amputated long ago. Overall, though, the brain’s ability to have bodily sensations which do not match up with what is actually occurring in the body can be understood to be an asset because this ability allows animals to react more quickly when needed. E.g., having the bodily sensations of fear, before the body has had time to go through the changes which otherwise would cause the bodily sensations of fear, allows the animal to take action regarding that danger much more quickly. (See Prinz, “Embodied Emotions,” 48.

or to the body. Bodily sensations are internal, but there are bodily sensations that are about the internal (one's body) and the external (the world beyond one's body). Bodily sensations focused on the internal include experiencing pain, heartbeat, and inflation and deflation of lungs due to breathing. Bodily sensations focused on the external include what we experience through our five senses of seeing, hearing, tasting, smelling, and touching. Our sense of kinesthetic balance may count as a bodily sensation about both the internal and external, as our sense of balance provides information about our body's position in the world. Each particular emotion (e.g., anger, sadness) has a specific cluster³ of bodily sensations associated with it. For example, the emotion of fear consists of bodily sensations, such as experiencing heightened awareness, rapid heartbeat, and muscle tension.

On a composite analysis of emotions, the relationship between feelings and emotions is complex. First, not all feelings are emotions, e.g., physical pains brought on by some external object harming the body are not part of any emotion. However, some physical pains are a *part* of emotions, e.g., fear can include a sour stomach, which is a stomach in some degree of physical pain. Second, even those feelings that are not a part of emotions can lead to emotions. For example, someone's sadness could be partially constituted by her judgments regarding some physical pain; namely, that that physical pain is horrible, yet the physical pain itself is not a part of this sadness.⁴

³ Use of the word 'cluster' in this context is found in *Ibid*.

⁴ Third, the cluster of bodily sensations which partly constitute an emotion, sometimes instead constitute a mood. In fact, after an emotion, the bodily sensations of that emotion can linger as a mood. The bodily sensations of anger can linger even after one realizes the inaccuracy of the judgment part of the anger – one is angry at a colleague but now one realizes the colleague did one no harm – but the bodily sensations of anger go on to constitute a bad (i.e., aggressive) mood. Before a full definition of emotions, it is not possible to clearly distinguish them from moods. Discussion on the nature of moods and their role in the exercise of the capacity to value occurs in Chapter 6.

Emotions partly Bodily Sensations & Partly Judgments

In contrast to a composite analysis of emotions, there is a traditional cognitive analysis of emotions holds that all emotions have propositional content: e.g., someone fears a lion because he sees the truth of some proposition like, “The lion has the physical strength plus the desire to kill me.” This traditional cognitive analysis of emotions runs against pre-theoretical intuitions regarding emotions. Some emotions have propositional content, but, surely, all do not have propositional content. Non-human animals and infants have emotions but lack the linguistic capacity necessary to form propositions: e.g., gazelles cannot form propositions regarding lions or anything else, yet, pre-theoretically, it seems clear to me that gazelles fear lions.

A composite analysis of emotions is a modified version of the traditional cognitive analysis. This composite analysis defines emotions as judgments, as Robert Solomon did, and can avoid the traditional cognitive theory’s problem of excluding non-human animals. A gazelle is hardwired to judge that large fast-moving objects, including lions, are a danger to it and, thereby, fears the lion.⁵ Solomon’s justification for claiming that emotions are judgments rather than something else, includes an explanation of why emotions as judgments is inclusive of non-human animals as emotional beings:

Judgments, unlike thoughts, are geared to perception and may apply directly to the situation we are in, but we can also make all sorts of judgments in the utter absence of any object of perception. Thus while I find the language of “thought” just too intellectual, too sophisticated, and too demanding in terms of linguistic ability, articulation, and reflection to apply to all emotions, “judgment” seems to me to have the range and flexibility to apply to everything from animal and infant emotions to the most sophisticated and complex adult human emotions such as jealousy, resentment, and moral indignation. In other words, I argue the following to be essential features of emotion and

⁵ The gazelle might be hard-wired to judge other aspects of the lion as fearful, such as: its smell, its teeth and claws that are large and sharp, its maintaining constant eyes on the gazelle while pacing back and forth in close proximity to the gazelle, its sometimes attempting to pounce on the gazelle or other gazelles, and its presence causing other gazelles run away. Exactly what the gazelle is hard-wired to judge as fearful is a question for biologists.

judgment: they are about the world (including oneself in the world). They are episodic but possibly long-term processes as well. They must span conscious and nonconscious awareness. . . . Emotions as judgments must accept as their “objects” both propositions and ordinary objects of perception (imagination, memory, etc.). They must be appropriate both in the presence of their objects and in their absence. And ... they must artfully bridge the categories of the voluntary and the involuntary.⁶

Judgments alone are not enough

All judgments certainly are not emotions. A human being might judge that a lion is dangerous and yet fail to have any bodily sensations related to such a judgment. Yet, intuitively, without the bodily sensations that we associate with fear, it seems false to say that this judging human being *fears* the lion. On the composite analysis, the emotion of fear consists of bodily sensations, such as experiencing heightened awareness, rapid heartbeat, and muscle tension, which all contribute to a general physical urge to fight, flee, or freeze to escape the danger.⁷ If someone simply judges that something is dangerous, yet fails to have any of the bodily sensations of fear, we cannot accurately call this the emotion of fear. In instances when bodily sensations of fear do not follow from the judgment that something is frightful, we do not find the emotion of fear but rather a judgment regarding what is frightful without the emotion of fear itself.

Judgments account for Intentionality & Appropriateness of Emotions

Some emotions are intentional – they are directed at some object, action, or state of affairs: e.g., if Sam is angry at Jane, Jane is the object of his anger. Moreover, we have standards of appropriateness which are connected to this intentionality: If Jane is a

⁶ Solomon, “Emotions, Thoughts, and Feelings,” 82-3.

⁷ Elster talks about how emotions can bring about urges for action in Jon Elster, “Emotions and Action,” in *Thinking about Feeling: Contemporary Philosophers on Emotion*, ed. by Robert Solomon, 151 – 162, (New York, NY: Oxford University Press, 2004), 151-5.

nice person and has never done Sam any wrong, it is inappropriate for Sam to be angry at Jane.

On a composite analysis, since emotions are judgments, it is clear how emotions can be intentional and be appropriate or inappropriate. Emotions are partly bodily sensations combined with judgments which elicit those bodily sensations. The judgments bring with them the intentionality and appropriateness dimensions of emotions. Sam's bodily sensations of anger were elicited by judgments regarding Jane, and these judgments are what make those bodily sensations, such as rapid heartbeat and the urge to strike, ones that are connected to Jane.

We can also turn to the judgment to consider the appropriateness of the emotions. Sam's bodily sensations of anger, such as rapid heartbeat and the urge to strike, would be understandable but unjustified, if those bodily sensations are part of an instance of anger which is also partly constituted by his judgment that she told a vicious lie about him to their supervisor, when in fact she did no such thing. Sam's anger makes sense given what he judges to be the truth, but he happens to be wrong about the truth. On the composite analysis, a judgment triggered Sam to have certain bodily sensations, and his anger consists of these judgments and bodily sensations. Moreover, Sam experiences anger, and not, say, fear, which might have the same feelings of rapid heartbeat and urge to strike, given the judgment: The type of judgment that partly constitutes anger, e.g., a judgment that one has been treated unfairly, is different from the type of judgment that partly constitutes fear, e.g., a judgment that one is in danger.

Partly Judgments and Partly Bodily Sensations together account for Families of Emotions

The heterogeneity of judgments⁸ helps to define different types of emotions within a family of emotions, e.g., anger vs. indignation, as well as helps to define different types of emotions with clusters of bodily sensations in common. Thus, for example, anger and cases of fear with fight responses have similar bodily sensation but different eliciting judgments, which define and constitute them as different emotions. At the same time, the cluster of bodily sensations help to define emotions such that a cluster of bodily sensations which partially define one emotion can be a part of a different emotion when combined with different kinds of judgments, e.g., fear of public speaking stemming from abstract thoughts to fear of lions stemming from observing the lion's attributes. On the composite analysis, one cannot look solely to either the judgment or the subsequent bodily sensations to help distinguish emotions because emotions are a composite of judgments and bodily changes, rather than one or the other.

How Non-Human Animals are Judging via Emotions

Why think that non-human animals are engaged in anything that we should properly call judgments?⁹ This analysis anthropomorphizes animals. In the case of the gazelle, isn't it more that the animal's fear is a hard-wired response to observing the largeness, speed, and other attributes of the lion? What it sees is that the lion has one or more attributes, and, after observing these attributes, it has all of the bodily sensations of fear towards the lion. It does not *judge* that the lion is dangerous or to be feared. In

⁸ Prinz, "Embodied Emotions," 53.

⁹ Somatic theorists do question a composite analysis on this front. E.g., Prinz, "Embodied Emotions," 56 – 57.

general, emotions in non-human animals are not elicited by judgments; rather, feelings are elicited in response to observing one or more attributes or combinations of attributes, which, innately or due to experience, elicit those feelings. Solomon's reply to this challenge is that the animal's emotions as well as many of our emotions are judgments, "*judgments of the body*":

Does it make sense to call these judgments? I am sure the answer is yes, and I would defend this in two steps. First, [as noted in an earlier quote from Solomon,]¹⁰ judgments are not necessarily articulate or conscious and so the sorts of discriminations we make and the construals that we perform are sometimes (often) made without our awareness of, much less reflection on, our doing so. Second, a relatively small store of [even] human knowledge is of the form "knowing that. ... [I]t is well known that "knowing how" cannot be reduced to any number of "knowing that" – type propositions. But it is a distortion of cognition and consciousness to suggest that "knowing that" – type propositional knowledge is in any way primary or independent of "knowing how." ... [E]motional judgments are not necessarily propositional but rather ways of engaging the world. Since they are ways of engaging with the world], the way is open to make the further claim that they are not necessarily "knowing that" – type cognitions either.¹¹

Solomon's explanation suggests that the gazelle is not making a judgment in the form of knowing that the lion is to be feared. It has no propositional knowing-that type knowledge. Instead, the gazelle is making a judgment in the form of knowing-how, knowing how to respond given what it observes. It knows how to respond via feelings of fear, which include rapid heartbeat, heightened awareness, and muscle tension which all contribute to an urge to freeze, flee, or fight in order to reduce or eliminate those feelings of fear.

Optional Acceptance of Composite View

While I find the composite view and Solomon's comments which help to offer a defense of that view response promising, others may find the composite view

¹⁰ See quote associated with Footnote 6.

¹¹ Solomon, "Emotions, Thoughts, and Feelings," 87.

unsatisfactory and would prefer a somatic analysis of emotions. On a somatic analysis of emotions, emotions are defined solely as the cluster of bodily sensations, which would sometimes be elicited by judgments, which are prior to but not partly constitutive of, emotions. However, even if one prefers to define emotions as clusters of bodily sensations, one need not reject my theory of personal identity over time. This is because valuing, the capacity at the center of my theory of personal identity, is a capacity to have reason-based emotions, which always involve judgments of reason and bodily sensations that are elicited or supported by judgments of reason.¹² In order to accept my theory, one only needs to agree that valuing involves judgments of reasons interacting with and often helping to elicit bodily sensations, one does not need to agree about whether ‘emotion’ should be properly defined as merely the bodily sensations or as the bodily sensations and the judgments which elicit them. Somatic theorist, as such, need not take issue with the idea that valuing involves feelings elicited by judgments, just as they do not take issue with the fact that some clusters of bodily sensations (which they say are solely constitutive of emotions) are elicited by judgments of reason, e.g., indignation.¹³ Moving forward, when I say “emotion”, I am denoting the bodily sensations of emotions together with the judgment(s)¹⁴ which elicit those feelings. Within the context of the emotions of valuing, those with somatic leanings can simply translate my use of the term “emotion”

¹² Belief is another capacity in which one will have feelings triggered by judgments. As Scanlon notes, “[b]elief is not merely a matter of judgment but of the connections, over time, between this judgment and dispositions to feel conviction”. (Scanlon 1998, p. 35) See Footnote **Error! Bookmark not defined.** for a discussion of the connection but distinction between beliefs and valuing. This discussion can only occur after the nature of valuing is fully described.

¹³ For example, see Prinz, “Embodied Emotions,” 53 and 57. “Indignation is something like anger at an injustice.” Prinz suggests that seeing something as an injustice involves judgments of reason.

¹⁴ Those judgments may be either non-reason based “knowing-how” type judgments or reason-based judgments of both “knowing-how” and “knowing-that” type judgments.

to “the bodily sensations and the judgment(s) of reason which elicits those bodily sensations”.

BIBLIOGRAPHY

- Aristotle, *Nichomachan Ethics*, trans. David Ross, rev. J. L. Ackrill and J. O. Urmon, Oxford, UK: Oxford University Press, 1925
- Baker, Lynn Rudder. *Persons and Bodies: A Constitution View*. Cambridge, UK: Cambridge University Press, 2000.
- _____. "Reply to Olson," *Book Symposia*, Spring 2001(E-Symposium on Persons and Bodies: A Constitution View,")
http://host.uniroma3.it/progetti/kant/field/bakersymp_replytoolson.htm
- Becker, Lawrence. *A New Stoicism*, Princeton, NJ: Princeton University Press, 1999.
- Buchanan, Allen. "Advance Directives and the Personal Identity Problem," *Philosophy and Public Affairs*. Vol. 17 (Fall 1998), 277 – 302.
- Crisp, Jan "Disorderly Ramblings: Alzheimer's and Narratives of Subjectivity" in *Deciphering Culture: Ordinary Curiosities and Subjective Narratives*. Crisp, Jane, Ferres, Kay, and Swanson, Gillian, eds. New York, NY: Routledge, 2000.
- Demasio, Antonio. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Penguin Books, 1994.
- DeGrazia, David. "Are We Essentially Persons? Olson, Baker, and a Reply," *The Philosophical Forum*, Vol. 33, No. 1 (Spring 2002), 101-120.
- _____. *Human Identity and Bioethics*. New York, NY: Cambridge University Press, 2005.
- Doris, John. *Lack of Character : Personality and Moral Behavior*. New York: Cambridge University Press, 2002
- Dresser, Rebecca "Dworkin on Dementia: Elegant Theory, Questionable Policy," *Hastings Center Report*. Vol. 25, No. 6, (1995); pp. 32-38
- Dworkin, Ronald. *Life's Dominion: An Argument about Abortion, Euthanasia, and Individual Freedom*. New York, NY: Knopf, 1993.
- Elster, Jon. "Emotions and Freedom." in *Thinking about Feeling: Contemporary Philosophers on Emotion*, Edited by Robert Solomon, 151 – 162. New York, NY: Oxford University Press, 2004.
- Ferlik, Andrew. "Margo's Logo," *JAMA - Journal of the America Medical Association*, Vol. 265, No. 2 (January 9, 1991), 201
- The Forgetting*, directed by Elizabeth Arledge, DVD, PBS: 2004 also at <http://www.pbs.org/theforgetting/watch/index.html> with transcript at

- http://www.pbs.org/theforgetting/watch/transcript_english.html. Last accessed April 5, 2012.
- Frankfurt, Harry G. *The Importance of What We Care About: Philosophical Essays*. Cambridge, UK: Cambridge University Press, 1988
- _____. *Necessity, Volition, and Love*. Cambridge, UK: Cambridge University Press, 1999
- Free Will*. 2nd Edited by Gary Watson New York, NY: Oxford University Press, 2003.
- Gosseries, O. *et al.* “Disorders of consciousness: what’s in a name?” *NeuroRehabilitation*. Vol. 28;No. 1 (2011); 3-14
- Hall, J K, “Caring for corpses or killing patients?” *Nursing Management*. Vol. 25, Issue 10 (Nov. 1994); 81-89
- Jasmin, Luc, MD, PhD, VeriMed Healthcare Network, David Zieve, MD, MHA, “What is Alzheimer’s Disease?” A.D.A.M., Inc.
http://www.alz.org/alzheimers_disease_what_is_alzheimers.asp
- Jaworska, Agnieszka “Respecting the Margins of Agency: Alzheimer's Patients and the Capacity to Value” *Philosophy and Public Affairs*. Vol. 28, No. 2 (Spr 99); pp. 105-138
- Johnston, Mark “Human Beings” *Journal of Philosophy* 84 1987, pp. 59-83
- Korsgaard, Christine. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, Oxford, UK: Oxford University Press, 2004
- _____. “Personal Identity and the Unity of Agency: A Kantian Response to Parfit.” *Philosophy and Public Affairs*. Vol. 28, No. 2 (1989), 101 – 132.
- _____. *Self-Constitution: Agency, Identity, and Integrity*. Oxford, UK: Oxford University, 2009.
- _____. *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996
- Kraut, Richard, “Aristotle's Ethics”, *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), Edward N. Zalta (Edited), forthcoming URL = <http://plato.stanford.edu/archives/win2011/entries/aristotle-ethics/>
- Kraut, Richard, “Supplementary Document: Alternate Readings of Aristotle on Akrasia”, *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), Edward N. Zalta (Edited), forthcoming URL = <http://plato.stanford.edu/archives/win2011/entries/aristotle-ethics/supplement1.html>

- Lewis, David "Survival and Identity," in A. Rorty, Edited, *The Identities of Persons*. Berkeley: University of California Press, 1976
- Locke, John. *An Essay concerning Human Understanding*, Edited Nidditch. Oxford, UK: Carendon Press, 1979.
- "Loops," *RadioLab*. Season 10, Episode 3, Podcast, <http://www.radiolab.org/2011/oct/04/> (Most recently assessed March 17, 2012.)
- MacIntyre, Alasdair *After Virtue: A study in moral theory*. Nortre Dame, IN: University of Nortre Dame Press, 1981
- Marzen, Thomas J. & Avila, Daniel "Will the Real Michael Martin Please Speak Up! Medical decisionmaking for questionably competent persons." *University of Detroit Mercy Law School*, Vol. 72, (Summer 1995); pp. 833-871
- Matthews, Debra, Bok, Hilary and Rabins, Peter. Edited *Personal Identity and Fractured Selves*. Baltimore, eds. MD: John Hopkins University Press, 2009.
- McLean, Athena Helen, "Coherence without Facticity in Dementia: The Case of Mrs. Fine." in *Thinking about Dementia: Culture, Loss, and the Anthropology of Senility*. 157 – 179, ed. Annette Leibing and Lawrence Cohen. New Brunswick, NJ: Rutgers University Press, 2006
- Nassbaum, Martha. "Emotions as Judgments of Value and Importance" in *Thinking about Feeling: Contemporary Philosophers on Emotion*, Edited by Robert Solomon, 183 – 199. New York, NY: Oxford University Press, 2004.
- Olson, Eric T. *The Human Animal: Personal Identity Without Psychology*. Oxford, UK: Oxford University Press, 1997.
- _____. "Personal Identity", *The Stanford Encyclopedia of Philosophy* (Winter 2010 Edition), Edward N. Zalta (Edited), URL = <<http://plato.stanford.edu/archives/win2010/entries/identity-personal/>>
- _____. *What are We? A Study in Personal Ontology*. Oxford, UK: Oxford University Press, 2007.
- Parfit, Derek. *Reasons and Persons*. New York: Oxford University Press, 1984.
- Perry, John. "Diminished and Fractured Selves." *Personal Identity and Fractured Selves*, Edited Debra Matthews, Hilary Bok, and Peter Rabins. 129 – 162. Baltimore, MD: John Hopkins University Press, 2009.
- Prinz, Jesse J. "Embodied Emotions" in *Thinking about Feeling: Contemporary Philosophers on Emotion*, Edited by Robert Solomon, 44 – 58. New York, NY: Oxford University Press, 2004.

- Quine, W. V. Review of Milton K. Munitz, ed., *Identity and Individualization*, *The Journal of Philosophy*, Vol. 69 (1972), 488 – 491, 490
- Rawls, John. “Two Concepts of Rules,” *The Philosophical Review* 64 (1955): 3- 32.
- Rorty, Amelie Oksenberg “Enough Already with “Theories of Emotions” ” in *Thinking about Feeling: Contemporary Philosophers on Emotion*, Edited by Robert Solomon, 269 – 278. New York, NY: Oxford University Press, 2004.
- Rosati, Connie S. “Agency and the Open Question Argument,” *Ethics* 113 (Apr 2003): 490-527.
- Rosnack, Rebecca L. “Termination of Life Sustaining Treatment: The decision-making process after Cruzan v. Missouri Department of Health.” *Baylor Law Review*. Vol. 43, (Fall 1991); pp. 841-74
- Rovane, Carol. *The Bounds of Agency: An essay in revisionary metaphysics*. Princeton: Princeton University Press, 1998
- _____. “Personal Identity and Choice.” *Personal Identity and Fractured Selves.*, Edited Debra Matthews, Hilary Bok, and Peter Rabins. 93 – 128. Baltimore, MD: John Hopkins University Press, 2009.
- Sachs, Oliver. *The Man who Mistook his Wife for a Hat and other clinical tales*. New York: Summit Books, 1985
- Sayre-McCord, Geoffrey. “ ‘Good’ on Twin Earth,” *Philosophical Issues* 8, Truth. (1997): 267-292.
- Scanlon, Thomas M. “Contractualism and Utilitarianism” in *Utilitarianism and Beyond*. Amartya Sen and Bernard Williams, eds. Cambridge: Cambridge University Press, 1982; 103-128
- _____. *What We Owe To Each Other*. Cambridge, MA: Belknap Press of Harvard University Press, 1998
- Schechtman, Mayra. *The Constitution of Selves*. Ithaca, NY: Cornell University Press, 1996.
- _____. “Empathic Access: The Missing Ingredient in Personal Identity” *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, Vol. 4, No. 2 (May 2001), 95-111
- _____. “Getting Our Stories Straight: Self-Narrative and Personal Identity.” In *Personal Identity and Fractured Selves.*, Edited Debra Matthews, Hilary Bok, and Peter Rabins., 65 – 92. Baltimore, MD: John Hopkins University Press, 2009.

- _____. "Personhood and Personal Identity," *Journal of Philosophy*, Vol. 85, No. 2, (1990); pp. 71-92
- "Emotions and the Brain" of "Episode Four: The Adult Brain: To Think by Feeling" in *The Secret Life of the Brain*. DVD. Produced by David Grubin. 2001. New York, NY: PBS Video, 2010.
- Shoemaker, David. "The Insignificance of Personal Identity for Bioethics," *Bioethics*. Vol. 24, No. 9, 481 – 489.
- _____. *Personal Identity and Ethics: A Brief Introduction*. (Buffalo, NY: Broadview Press, 2008).
- _____. "Personal Identity and Ethics", *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), Edward N. Zalta (Edited), forthcoming URL = [<http://plato.stanford.edu/archives/win2011/entries/identity-ethics/>](http://plato.stanford.edu/archives/win2011/entries/identity-ethics/).
- Singer, Peter. "Is the Sanctity of Life Ethic Terminally Ill?", *Bioethics*, Vol. 9, Nos. 3/4 (1995), 327-343.
- Solomon, Robert C. "Emotions, Thoughts, and Feelings: Emotions as Engagements with the World" in *Thinking about Feeling: Contemporary Philosophers on Emotion*, Edited by Robert Solomon, 76 – 89. New York, NY: Oxford University Press, 2004.
- Strawson, Galen. "Against Narrativity," *Ratio*. Vol. 17, No. 4 (December 2004), 428 – 452.
- _____. "Episodic Ethics," *Philosophy: The Journal of the Royal Institute of Philosophy*, Vol. 60 , Supplement (January 2007), 85-115
- Watson, Gary. "Free Action and Free Will." *Mind* 96 (Apr 1987): 145-172
- Wilkes, Kathleen. *Real People. Personal Identity with Thought Experiments*. Oxford: Oxford University Press, 1988.
- Williams, Bernard. *Problems of the Self*. Cambridge: Cambridge University Press, 1973
- Wittgenstein, Ludwig, *Zettel*, ed. By G. Anscombe and G. von Wright, and translated by G. Anscombe. Berkeley, CA: Blackwell, 1967.
- Wolf, Susan, "Moral Saints," *Journal of Philosophy*. Vol. 79, No. 8, (August 1982), 419-439.
- _____. "Self-Interest and Interest in Selves," *Ethics*. July 1986, Vol. 96, No. 45: 704-720
- Wollheim, Richard. *The Thread of Life*. Cambridge: Harvard University Press, 1984.

Unger, Peter. *Identity, Consciousness & Value*. Oxford, UK: Oxford University Press, 1990.

Unger, Peter. *All The Power In The World*. Oxford, UK: Oxford University Press, 2007.